Numerical Linear Algebra with Applications

WILEY

RESEARCH ARTICLE OPEN ACCESS

Structured Nuclear Norm Matrix Completion: Guaranteeing Exact Recovery via Block-Column Scaling

Konstantin Usevich[1] | Jonathan Gillard[2] | Philippe Dreesen[3] | Ivan Markovsky[4]

[1]Université de Lorraine, CNRS, CRAN, Nancy, France | [2]School of Mathematics, Cardiff University, Cardiff, UK | [3]Department of Advanced Computing Sciences (DACS), Maastricht University, Maastricht, the Netherlands | [4]Catalan Institution for Research and Advanced Studies and Centre Internacional de Mètodes Numèrics en Enginyeria, Barcelona, Spain

Correspondence: Konstantin Usevich (konstantin.usevich@cnrs.fr)

Received: 20 September 2023 | **Revised:** 15 April 2025 | **Accepted:** 9 June 2025

Funding: This work was supported by the Catalan Institution for Research and Advanced Studies (ICREA), Fond for Scientific Research Vlaanderen (FWO) project (G033822N), and Spanish Ministry of Science State Research Agency (MCIU/AEI/10.13039/501100011033,grantPID2023-148952OB-I00).

Keywords: exact recovery | matrix completion | nuclear norm | structured matrices

ABSTRACT

The goal of low-rank matrix completion is to minimize the rank of a matrix while adhering to the constraint that known (non-missing) elements are fixed in the approximation. Minimizing rank is a difficult, non-convex, NP-hard problem, often addressed by substituting rank with the nuclear norm to achieve a convex relaxation. We focus on structured matrices for completion, where, in addition to the constraints described earlier, matrices also adhere to a predefined structure. We propose a technique that ensures the exact recovery of missing entries by minimizing the nuclear norm of a matrix where the non-missing entries are first subject to block-column scaling. We provide the proofs for exact recovery and propose a way for choosing the scaling parameter to ensure exact recovery. The method is demonstrated in several numerical examples, showing the usefulness of the proposed technique.

1 | Introduction

Low-rank matrix completion aims to reconstruct matrices with missing entries by exploiting an inherent low-rank structure. In the general case, low-rank matrix completion is an optimization problem with the objective of minimizing the rank of a matrix subject to constraints. Rank minimization problems are generally NP-hard and non-convex (see, e.g., [1]), so several authors propose to instead solve a convex relaxation based on the nuclear norm, which has well-grounded theory for the case of unstructured matrices [2–4]. While there are other approaches to relax the nonconvexity—see, for example, [5]—we focus on the nuclear norm approach since it is pervasive. A natural question emerges—when does the solution of the convex relaxation coincide with the rank minimization problem? It is known that the nuclear norm performs well when the missing values are sampled in a random fashion, where perfect (or exact) recovery is ensured with high probability, see the seminal work [2, 3] and numerous follow-ups.

In this paper, we consider the case of structured matrix completion. Informally speaking, structured matrices have dependencies between their elements, and we wish to preserve the structure whilst making the completion. Examples of well-known structured matrices include Hankel, block-Hankel, Toeplitz, Sylvester, and circulant; each of these structures is associated with particular problem domains, and we offer some examples shortly. The completion of structured matrices can be viewed as a special case of structured low-rank approximation [6, 7]. In the applications we consider, and for structured matrices generally,

missing data often appear in blocks as opposed to being randomly located throughout the matrix.

Specifically, we consider structured matrix completion based on minimizing the nuclear norm in the deterministic scenario (structure and missing-data-not-at-random). In this case, the use of the nuclear norm breaks down, as shown by several studies [8]. For Hankel matrices References [9, 10] suggest that an exponential rescaling (dampening) of the problem is possible, but it is difficult to provide good bounds for such a rescaling. In this paper, we propose a simple block-column scaling of the observed entries, which guarantees the exact recovery of missing entries. Our contribution in this paper is applicable to any affinely-structured matrix and provides theoretical guarantees for exact recovery as well as a method to choose the scaling parameter.

There are other methods available to solve some of the example problems described in this paper. Note, however, that our work applies to general affine structures (not just Hankel-type); for general structures, parametric model equivalents may not be available. In this paper, we propose an alternative to estimating parameters of a parametric model which is not only demonstrably useful for practical application, but also interesting mathematically, and complements the results in seminal work such as [2, 3, 11]. Hence, our contributions are beyond just a method to solve a particular Hankel matrix completion problem.

This paper has the following structure. In Section 2, we introduce the matrix completion problem and provide some examples of matrix structures and typical completions needed. In Section 3, we describe the means to scale the matrix structure so that the nuclear norm relaxation achieves the correct recovery for the scaled matrix structure. Our result on scaling columns to achieve exact recovery of missing values is also given in Section 3. Before giving the proof of the main result in Section 5, we provide a roadmap in Section 4. Numerical examples are given in Section 6 before the paper is concluded in Section 7.

2 | Affine Structured Matrix Completion

2.1 | Structured Matrices

An affine matrix structure $\mathcal{S}(\cdot)$ is an affine mapping $\mathcal{S} : \mathbb{F}^N \to \mathbb{F}^{L\times K}$ where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$, thus parameterizing (an affine) set of structured matrices. An example of such matrices are given below and a formal construction is offered in Section 4.1.

Example 2.1 (running example, Hankel matrices). The Hankel matrix structure $\mathcal{H}_L : \mathbb{F}^N \to \mathbb{F}^{L\times K}$, $K = N - L + 1$, maps a vector $\boldsymbol{p} \in \mathbb{F}^N$ to an $L \times K$ matrix with

$$(\mathcal{H}_L(\boldsymbol{p}))_{i,j} = \boldsymbol{p}_{i+j-1}$$

that is, the values are constant on the antidiagonals and parameterized by the elements of $\boldsymbol{p}$ (see also (1)).

Note that the notation $\mathcal{H}_L(\boldsymbol{p})$ is used for a vector $\boldsymbol{p}$ of any size $N \geq L$, and the number of columns of the matrix is determined by the size of the input vector.

2.2 | Matrix Completion and Rank Minimization

The goal of low-rank matrix completion is to fill in the missing elements of the matrix $\mathcal{S}(\boldsymbol{p})$ based on the low-rank assumption. From now on, we write $N = n + m$, and we assume that the first n values of $\boldsymbol{p}$ (denoted $\boldsymbol{p}_{1:n}$) are k**n**own and the goal is to recover the last m **m**issing values[1], $\boldsymbol{p}_{n+1:n+m}$ (so that $N = n + m$).

Formally, the exact structured low-rank matrix completion is formulated as the rank minimization

$$\min_{\boldsymbol{p}\in\mathbb{F}^{m+n}} \operatorname{rank}\{\mathcal{S}(\boldsymbol{p})\} \quad \text{subject to} \quad \boldsymbol{p}_{1:n} = \boldsymbol{p}_0 \tag{RMIN}$$

where $\boldsymbol{p}_0 = (p_{0,1}, p_{0,2}, \ldots, p_{0,n})$ is a given vector of known values. We denote an optimal solution to (RMIN) as $\widehat{\boldsymbol{p}}^{(\text{RMIN})}$.

Example 2.2 (Example 2.1, continued). Our running example stems from time series analysis, where we would like to fill in the last values of a scalar Hankel matrix. In Equation (1), the gray-shaded elements are known (fixed) and the remaining values are missing. The Hankel matrix in Equation (1) has size $L \times K$ so that $K + L - 1 = n + m$.

$$\mathcal{H}_L(\boldsymbol{p}) \overset{\text{def}}{=} \begin{bmatrix} p_1 & p_2 & \cdots & \cdots & \cdots & p_K \\ p_2 & p_3 & \cdots & \iddots & \cdots & \vdots \\ \vdots & \cdots & \iddots & \iddots & \cdots & p_n \\ \vdots & \iddots & \iddots & \iddots & \cdots & p_{n+1} \\ \vdots & \iddots & \iddots & \iddots & \iddots & \vdots \\ p_L & \cdots & p_n & p_{n+1} & \cdots & p_{n+m} \end{bmatrix} \tag{1}$$

The Hankel low-rank completion problem is shown to be useful in the context of forecasting [12–14], where $\boldsymbol{p}$ corresponds to a scalar time series. The first n values of the time series $\boldsymbol{p}$ are known (observed), and we would like to forecast m steps forward.

This approach is motivated by the fact that time series, which can be written as a sum of products of polynomial, exponential, and sinusoidal functions, have low-rank Hankel matrices (see, e.g., [14]). Such time series constitute an exceptionally rich class able to model complex trends and multiple modulated periodic components (see also [15]). Here, forming the forecasting problem as Hankel structured low-rank matrix completion is very attractive because it avoids parameter estimation necessary for classic model-based statistical forecasting methods [16]. A more general class of problems with multivariate time series is described in Section 6.4. In practice, an approximate version of (RMIN) is often considered when we allow for the error in the known values.

2.3 | Convex Relaxation and Nuclear Norm

In this manuscript, we investigate the performance of the nuclear norm relaxation, which for the rank minimization problem (RMIN) is replaced with:

$$\min_{\boldsymbol{p}\in\mathbb{F}^{(n+m)}} \|\mathcal{S}(\boldsymbol{p})\|_* \quad \text{subject to} \quad \boldsymbol{p}_{1:n} = \boldsymbol{p}_0 \tag{NNMIN}$$

where $\|\boldsymbol{X}\|_*$ is the nuclear norm (sum of all singular values of $\boldsymbol{X}$). The intuition behind this relaxation is the same as for using the ℓ_1-norm in compressed sensing. The nuclear norm is expected to force all but a few singular values to be zero (a low-rank solution). In what follows, we will denote an optimal solution to (NNMIN) as $\widehat{\boldsymbol{p}}^{(\mathrm{NNMIN})}$.

In our applications, the pattern of missing data is fixed, as opposed to the random sampling in Reference [2, 3], and the conditions for perfect recovery are much more restrictive in such cases. In particular, it was shown in References [9] and [10] that in Hankel matrix completion, the time series need to be sufficiently damped to allow for perfect recovery. In this paper, we show that for a wide range of structures, a simple block-wise scaling guarantees perfect recovery of the scaled matrix structure.

2.4 | Alternatives to the Nuclear Norm

Nuclear norm minimization has been celebrated for its convex formulation, which offers strong theoretical recovery guarantees and robustness in the presence of noise. Nonetheless, several alternative methods for matrix completion exist that avoid its convex relaxation while offering complementary advantages. For instance, low-rank factorization-based approaches model the given matrix as a product $\boldsymbol{A}\boldsymbol{B}$ and apply alternating minimization to optimize for $\boldsymbol{A}$ and $\boldsymbol{B}$. Although these methods tend to be more computationally efficient, they often require careful initialization and might converge to local minima rather than the globally optimal solution [17]. In addition, Riemannian optimization techniques directly optimize on the manifold of fixed-rank matrices to benefit from the intrinsic geometric structure of the solution space, typically resulting in rapid convergence; however, their performance can be sensitive to noise and depends on the quality of the starting point [18]. Iterative hard thresholding methods enforce a strict rank constraint via singular value truncation at each iteration, which can be highly efficient in practice, yet they often lack the universal convergence guarantees that come with convex nuclear norm methods [19]. Overall, while these alternatives provide efficient and scalable solutions in various settings, nuclear norm minimization remains a cornerstone approach due to its balanced trade-off of robust global optimality, provable error bounds, and practical effectiveness in a wide range of applications, as described in this paper.

3 | Block-Column Scaling for Nuclear Norm Completion

3.1 | Block-Column Scaling

In this section, we describe the proposed scaling approach ensuring that nuclear norm minimization for a fixed structure obtains exact recovery of missing values. Our proposal is to replace $\mathcal{S}(\boldsymbol{p})$ with the scaled matrix structure

$$\mathcal{S}_\varepsilon(\boldsymbol{p}) := \mathcal{S}(\boldsymbol{p}) \begin{bmatrix} \boldsymbol{I}_{K-\kappa} & \\ & \varepsilon \boldsymbol{I}_\kappa \end{bmatrix} \tag{2}$$

where $\varepsilon \geq 0$ is a small parameter, that is, we scale the last κ columns of $\mathcal{S}(\boldsymbol{p})$, with $0 < \kappa < K$, by a small number ε. The following remark is of prime importance:

Remark 3.1. For $\varepsilon > 0$, we have $\mathrm{rank}\{\mathcal{S}_\varepsilon(\boldsymbol{p})\} = \mathrm{rank}\{\mathcal{S}(\boldsymbol{p})\}$ therefore all the solutions of the exact matrix completion (RMIN) using either the original matrix $\mathcal{S}(\boldsymbol{p})$ or its scaled version $\mathcal{S}_\varepsilon(\boldsymbol{p})$ coincide.

Our proposal is to replace $\mathcal{S}(\boldsymbol{p})$ with $\mathcal{S}_\varepsilon(\boldsymbol{p})$, in the nuclear norm minimization problem (NNMIN), and we seek to solve the modified problem

$$\min_{\boldsymbol{p}\in\mathbb{R}^{(n+m)}} \|\mathcal{S}_\varepsilon(\boldsymbol{p})\|_* \quad \text{subject to} \quad \boldsymbol{p}_{(1:n)} = \boldsymbol{p}_0 \qquad \text{(NNMIN-}\varepsilon\text{)}$$

Under several natural assumptions, replacing $\mathcal{S}(\boldsymbol{p})$ with the scaled $\mathcal{S}_\varepsilon(\boldsymbol{p})$ leads to guaranteed exact recovery of the missing values $\boldsymbol{p}_{n+1:n+m}$, for suitably chosen ε. The goal is to choose such a scaling so that the solution $\widehat{\boldsymbol{p}}^{(\text{NNMIN-}\varepsilon)}$ of (NNMIN-ε) leads to exact recovery of $\boldsymbol{p}_{n+1:n+m}$ and is therefore better than $\widehat{\boldsymbol{p}}^{(\mathrm{NNMIN})}$.

3.1.1 | Related Work

Scaling of rows and columns has already been proposed for the case of unstructured matrix completion. Such a scaling, equivalent to a weighted nuclear norm minimization [20], was shown to be effective and provides better recovery guarantees than ordinary nuclear norm minimization [21]. There has been an increasing activity in analyzing non-uniform deterministic patterns of missing values [22]. However, the authors are unaware of any such works that address the case of structured matrix completion, except for the recent work [23] that mainly reveals connections between weighting and structured completion (rather than using scalings to improve the performance of the nuclear norm in the structured case). Our work, therefore, offers a novel contribution to ensuring the exact recovery of missing entries of structured matrices.

3.2 | Assumptions and the Main Result

We now describe the necessary assumptions before discussing them in more detail in the subsequent section.

Assumption 3.1. The solution $\widehat{\boldsymbol{p}}^{(\mathrm{RMIN})}$ of the rank minimization problem (RMIN) is unique.

Assumption 3.2. For any optimal solution $\widehat{\boldsymbol{p}}^{(\mathrm{RMIN})}$ of (RMIN), we have

$$\mathrm{rank}\{\mathcal{S}(\widehat{\boldsymbol{p}}^{(\mathrm{RMIN})})\} = \mathrm{rank}\{\mathcal{S}_0(\widehat{\boldsymbol{p}}^{(\mathrm{RMIN})})\}$$

where $\mathcal{S}_0(\boldsymbol{p})$ is the scaled structure (2) with $\varepsilon = 0$.

Assumption 3.3. The matrix structure $\mathcal{S}_0(\boldsymbol{p})$ depends only on $\boldsymbol{p}_{1:n}$ (i.e., the first n elements of the vector $\boldsymbol{p}$); in such a case we write $\mathcal{S}_0(\boldsymbol{p}) = \mathcal{S}_0(\boldsymbol{p}_{1:n})$.

Before formulating the main results, we make some remarks about these technical assumptions.

Remark 3.2. Assumptions 3.2 and 3.3 concern the matrix $S_0(\boldsymbol{p})$

$$S_0(\boldsymbol{p}) := S(\boldsymbol{p}) \begin{bmatrix} \boldsymbol{I}_{K-\kappa} & \\ & \boldsymbol{0} \end{bmatrix} \tag{3}$$

This is a limiting ($\varepsilon = 0$) case of $S_\varepsilon(\boldsymbol{p})$, which is special for the following reasons:

- In this case, Remark 3.1 (on the equivalence of rank minimization problems for the scaled matrix $S_\varepsilon(\boldsymbol{p})$ and $S(\boldsymbol{p})$) does not necessarily hold. Thus Assumption 3.2 ensures that an analogue of Remark 3.1 holds true for $\varepsilon = 0$;
- As seen in Equation (3), $S_0(\boldsymbol{p})$ contains the first $K - \kappa$ columns of $S(\boldsymbol{p})$ (which are left unscaled in $S_\varepsilon(\boldsymbol{p})$ for any ε); thus Assumption 3.3 is equivalent to assuming that the parameters $\boldsymbol{p}_{n+1:n+m}$ do not appear in the first $K - \kappa$ columns (i.e., all the missing elements $\boldsymbol{p}_{n+1:n+m}$ are scaled by ε);
- From a practical point of view, $S_0(\boldsymbol{p})$ is not relevant as it does not depend on the missing elements (so it does not provide a way to complete them); however, it is important for the theoretical analysis of the scaled problem (NNMIN-ε) which appears later.

We are also going to provide examples for the assumptions in the following subsections. Under Assumptions 3.1–3.3, the following result holds.

Theorem 3.3. *Let $S_\varepsilon(\boldsymbol{p})$ be the scaled matrix structure (2) satisfying Assumptions 3.1–3.3, and $\widehat{\boldsymbol{p}}^{(\mathrm{RMIN})}$ be the unique solution of (RMIN). Then there exists ε_0 such that for any $\varepsilon \in (0, \varepsilon_0)$, the minimizer $\widehat{\boldsymbol{p}}^{(\mathrm{NNMIN}\text{-}\varepsilon)}$ of the scaled problem* (NNMIN-ε) *is unique and is equal to $\widehat{\boldsymbol{p}}^{(\mathrm{RMIN})}$.*

For ease of reading, the proof of the theorem will be split into several stages and into dedicated subsections. In the rest of the section, we provide a discussion on the plausibility of the assumptions, show the key ideas behind the proof, and also indicate how we can find an estimate for ε_0.

3.3 | Discussion on the Assumptions: The Case of Hankel Matrices

In this subsection, we give examples for the assumptions using our running example, that of completing Hankel matrices, which are linked to some well-known conditions in the literature.

Lemma 3.4. *For completing the Hankel matrix $\mathcal{H}_L(\boldsymbol{p})$ given in Equation* (1)*, Assumption 3.3 is equivalent to $\kappa \geq m$.*

Proof. The last elements $\boldsymbol{p}_{n+1:n+m}$ appear only in the m last columns of $\mathcal{H}_L(\boldsymbol{p})$ and so the first $K - \kappa$ columns do not depend on $\boldsymbol{p}_{n+1:n+m}$ if and only if $\kappa \geq m$. □

To illustrate Lemma 3.4, consider a special case of a Hankel matrix (1) with $n = 7$, $m = 2$, and $\kappa = 3$ (i.e., the right-hand side block is scaled):

$$\mathcal{H}_3(\boldsymbol{p}) = \begin{bmatrix} p_1 & p_2 & p_3 & p_4 & p_5 & p_6 & p_7 \\ p_2 & p_3 & p_4 & p_5 & p_6 & p_7 & p_8 \\ p_3 & p_4 & p_5 & p_6 & p_7 & p_8 & p_9 \end{bmatrix}$$

and it is seen that Assumption 3.3 is satisfied. We also note that a necessary condition for Assumption 3.1 is $r = \operatorname{rank}\{\mathcal{H}_L(\boldsymbol{p}_0)\} < L$ (otherwise any completion of $\mathcal{H}_L(\boldsymbol{p}_0)$ has rank L, and is thus nonunique). Under a stronger, well-known condition, we can guarantee that both Assumptions 3.1 and 3.2 are also satisfied.

Lemma 3.5. *Let Assumption 3.3 be satisfied for the Hankel matrix structure in Equation (1). If, in addition,*

$$\boldsymbol{e}_L = (0, \ldots, 0, 1) \notin \operatorname{span}\{\mathcal{H}_L(\boldsymbol{p}_0)\}$$

then both Assumptions 3.1 and 3.2 are also satisfied.

Lemma 3.5 is well-known in Hankel low-rank approximation literature (see [24, Def. 5.9, p. 99], [15, §5.3], but also [6, 14]), and we summarize a proof in Appendix A. In particular, Lemma 3.5 holds true for time series that are sums of complex exponentials and polynomials, and this is the example that follows.

Example 3.6 (see, e.g., [14, Thm 5.1]). Let p_k, $k = 1, \ldots, n$ be given as

$$p_k = \sum_{j=1}^{s} P_j(k) \lambda_j^k, \quad k = 1, 2, \ldots \tag{4}$$

where $P_j(k)$ are complex polynomials of degrees at most $\nu_j - 1$ and $\lambda_j \in \mathbb{C}$. Then if $r = \nu_1 + \cdots + \nu_r \leq \min(K\text{-}\kappa, L - 1)$, then the rank minimization of $\mathcal{H}_L(\boldsymbol{p})$ is unique (is of rank r), and the unique completion is given by the same formula ($\widehat{p}_k = p_k$, $k = n + 1, \ldots, n + m$ in Equation (4)). Therefore, both Assumptions 3.1 and 3.2 are satisfied.

We offer further commentary on the applicability of Assumptions 3.1–3.3 for some other problem settings.

Remark 3.7. The problem treated in Lemma 3.5 and Example 3.6 can be viewed as a special case in a more general context behavioral systems theory. In particular, Assumptions 3.1–3.3 are similar to Assumptions A1–A3 in Reference [25], which considers the same problem (for the special case of a so-called mosaic-Hankel structure) from a systems theoretic perspective. The main difference is the method of solution, which is not based on the nuclear norm relaxation.

Additionally, Assumptions 3.1–3.3 are intuitive for the problem of data-driven simulation, which is presented in depth in Section 6.4. For example, Assumption 3.2 corresponds to the condition of persistency of excitation [26]. We also note that a similar scaling was heuristically proposed for the special case of data-driven simulation in Reference [27], where it was proposed to scale the matrix as

$$\begin{bmatrix} \gamma \boldsymbol{I}_{K-\kappa} & \\ & \boldsymbol{I}_\kappa \end{bmatrix} S(\boldsymbol{p})$$

where γ is a large number. Such a scaling is, in fact, equivalent to (2) if we take $\gamma = \varepsilon^{-1}$.

Finally, we discuss the applicability of Assumptions 3.1–3.3 when the missing values occur elsewhere in the parameter vector $\boldsymbol{p}$.

Remark 3.8. Our results also apply when:

- the missing data are located in the middle of the vector $\boldsymbol{p}$;
- the columns in the middle of the matrix are scaled by ε.

Indeed, this case can be treated by rearranging the columns of the matrix and the elements of the parameter vector (this operation affects neither the rank nor the nuclear norm).

Consider the following example

$$\mathcal{H}_3(\boldsymbol{p}) = \begin{bmatrix} p_1 & p_2 & p_3 & p_4 & p_5 & p_6 & p_7 \\ p_2 & p_3 & p_4 & p_5 & p_6 & p_7 & p_8 \\ p_3 & p_4 & p_5 & p_6 & p_7 & p_8 & p_9 \end{bmatrix}$$

with missing elements p_4 and p_5. We can rearrange the columns to get an equivalent matrix

$$\mathcal{S}(\boldsymbol{p}) = \begin{bmatrix} p_1 & p_6 & p_7 & p_2 & p_3 & p_4 & p_5 \\ p_2 & p_7 & p_8 & p_3 & p_4 & p_5 & p_6 \\ p_3 & p_8 & p_9 & p_4 & p_5 & p_6 & p_7 \end{bmatrix}$$

and by permuting the arguments of the vector, we obtain an equivalent matrix structure that satisfies Assumption 3.3.

4 | Roadmap for the Proof

4.1 | Nuclear Norm Minimization: Optimality Conditions

To give the main ideas behind the proof of Theorem 3.3, we recall the optimality conditions for (NNMIN) from References [10, 13], and we cover both real- and complex-valued matrices.

Let $\widehat{\boldsymbol{p}}$ be a parameter vector for which we wish to test its optimality, i.e., whether it is a minimizer of (NNMIN), and let $r = \operatorname{rank}\{\mathcal{S}(\widehat{\boldsymbol{p}})\}$. Let us introduce the following notation for the compact SVD:

$$\mathcal{S}(\widehat{\boldsymbol{p}}) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}},\ \boldsymbol{U} \in \mathbb{C}^{L\times r},\ \boldsymbol{\Sigma} \in \mathbb{R}^{r\times r},\ \boldsymbol{V} \in \mathbb{C}^{K\times r}$$

The key quantities needed for formulating the optimality conditions are:

- the polar factor $\boldsymbol{B} \overset{\text{def}}{=} \boldsymbol{U}\boldsymbol{V}^{\mathrm{H}}$ (semi-unitary matrix in polar decomposition)
- and projectors on the left and right nullspace of $\mathcal{S}(\widehat{\boldsymbol{p}})$:

$$\boldsymbol{Q}_1 = \boldsymbol{I}_L - \boldsymbol{U}\boldsymbol{U}^{\mathrm{H}}, \quad \boldsymbol{Q}_2 = \boldsymbol{I}_K - \boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}$$

Note that the polar factor $\boldsymbol{B}$ and the projectors $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ depend on $\widehat{\boldsymbol{p}}$ (which is fixed in the proofs), therefore we omit this dependence in the notation for the ease of reading.

Next, we recall notation for basis matrices, which are needed to consider structured matrix completion problems. An affine matrix structure [7] can be parameterized as:

$$\mathcal{S}(\boldsymbol{p}) = \boldsymbol{S}_0 + \sum_{k=1}^{m+n} p_k \boldsymbol{S}_k \tag{5}$$

For example, for the Hankel structure (1), the basis matrices are

$$\boldsymbol{S}_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \\ \vdots & \ldots & \ldots & \ldots & \vdots \\ 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \boldsymbol{S}_2 = \begin{pmatrix} 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \ldots & \ldots & \ldots & \vdots \\ 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

$$\ldots, \boldsymbol{S}_{m+n} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \\ \vdots & \ldots & \ldots & \ldots & \vdots \\ 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

with $\boldsymbol{S}_0$ being the zero matrix.

Using the notation introduced above, the necessary (and sufficient) optimality conditions for a vector to be an optimal solution of (NNMIN) can be formulated as follows.

Lemma 4.1 (First-order optimality conditions [10, Proposition 14] [13]). *The point $\widehat{\boldsymbol{p}}$ is a minimizer of (NNMIN) if and only if there exists a matrix $\boldsymbol{M} \in \mathbb{C}^{L\times K}$ with spectral norm $\|\boldsymbol{M}\|_2 \le 1$, satisfying*

$$\langle \boldsymbol{Q}_1 \boldsymbol{M} \boldsymbol{Q}_2 + \boldsymbol{B}, \boldsymbol{S}_k \rangle_F = 0, \quad \text{for all } k \in \{n+1, \ldots, n+m\} \tag{6}$$

If, in addition, the norm of matrix $\boldsymbol{M}$ satisfies $\|\boldsymbol{M}\|_2 < 1$, and the set of matrices $\{\boldsymbol{Q}_1 \boldsymbol{S}_k \boldsymbol{Q}_2\}_{k=n+1}^{n+m}$ is linearly independent, then $\widehat{\boldsymbol{p}}$ is the unique minimizer of (NNMIN).

A matrix $\boldsymbol{M}$ satisfying the conditions of Lemma 4.1 (with $\|\boldsymbol{M}\|_2 < 1$) is called a *dual certificate*. It is a standard term in the nuclear norm minimization literature [2, 3] as it refers to duality in convex optimization [28]. Our proof strategy relies on constructing a particular dual certificate for the given solution of (NNMIN), that is, the matrix $\boldsymbol{M}$ satisfying linear constraints (6), for which we can guarantee $\|\boldsymbol{M}\|_2 < 1$.

4.2 | Candidate Dual Certificate and Its Norm

Instead of finding $\boldsymbol{M}$ with small spectral norm (as suggested by Lemma 4.1), we relax the problem and find $\boldsymbol{M}$ with small Frobenius norm (subject to constraints (6)), which has explicit solution. A matrix with minimal Frobenius norm will be called a *candidate dual certificate* (see also [3]), this is defined next.

Definition 4.2. A candidate dual certificate for $\widehat{\boldsymbol{p}}$ in Equation (NNMIN) is a minimizer of

$$\boldsymbol{M}^* \overset{\text{def}}{=} \operatorname{argmin} \|\boldsymbol{M}\|_F \quad \text{subject to} \quad \text{constraints in (6)} \tag{7}$$

Due to the standard inequality $\|\boldsymbol{M}\|_2 \le \|\boldsymbol{M}\|_F$, a candidate dual certificate becomes a dual certificate if $\|\boldsymbol{M}_*\|_F < 1$ (and thus will ensure optimality in Equation (NNMIN)). Therefore, the key point of the proof would be to guarantee that such an inequality holds for a sufficiently scaled problem, as shown by the following proposition.

Proposition 4.3. *Under Assumptions 3.1–3.3 let $\widehat{\boldsymbol{p}} = \widehat{\boldsymbol{p}}^{(\mathrm{RMIN})}$, and let $\boldsymbol{M}^*(\varepsilon)$ denote the candidate dual certificate for the scaled problem (NNMIN-ε). Then, for small $\varepsilon > 0$, the candidate dual certificate exists, is unique, and its squared Frobenius norm has the expansion*

$$\|\boldsymbol{M}^*(\varepsilon)\|_F^2 = m_0\varepsilon^2 + o(\varepsilon^2)$$

where m_0 is a constant depending only on $\mathcal{S}(\widehat{\boldsymbol{p}}^{(\mathrm{RMIN})})$.

The proof of the proposition, as well as precise form of the constant m_0 will be given in Section 5.3. Note that, in particular, Proposition 4.3 ensures that $\lim_{\varepsilon\to 0}\|\boldsymbol{M}^*(\varepsilon)\|_F^2 = 0$, which shows that for sufficiently small ε we have $\|\boldsymbol{M}_*(\varepsilon)\|_F < 1$.

We conclude this section with a remark on the usefulness of Proposition 4.3; it will enable us to provide a practical estimate for the value of the scaling needed to obtain exact recovery.

Remark 4.4. For small ε, Proposition 4.3suggests that we can replace $\|\boldsymbol{M}^*(\varepsilon)\|_F^2$ with its leading term approximation. Therefore, the value of ε, chosen such that $\|\boldsymbol{M}^*(\varepsilon)\|_F^2 \approx \varepsilon^2 m_0 = 1$ is a good estimate to enable exact recovery. It is given by

$$\tilde{\varepsilon} = \frac{1}{\sqrt{m_0}} \tag{8}$$

4.3 | Sketch of the Proof

To give a gist of the proof, we first note that the constraints (6) are linear in elements of $\boldsymbol{M}$ and thus can be equivalently vectorized as

$$\boldsymbol{A}\mathrm{vec}(\boldsymbol{M}) = \boldsymbol{b} \tag{9}$$

where $\boldsymbol{A} \in \mathbb{C}^{m\times LK}$ and $\boldsymbol{b} \in \mathbb{C}^m$ are the matrices obtained by vectorizing (6):

$$\boldsymbol{A} = \begin{bmatrix} \mathrm{vec}^{\mathrm{T}}(\boldsymbol{S}_{n+1})(\boldsymbol{Q}_2^{\mathrm{T}} \otimes \boldsymbol{Q}_1) \\ \vdots \\ \mathrm{vec}^{\mathrm{T}}(\boldsymbol{S}_{n+m})(\boldsymbol{Q}_2^{\mathrm{T}} \otimes \boldsymbol{Q}_1) \end{bmatrix} \tag{10}$$

$$\boldsymbol{b} = \begin{bmatrix} -\langle \boldsymbol{B}, \boldsymbol{S}_{n+1}\rangle \\ \vdots \\ -\langle \boldsymbol{B}, \boldsymbol{S}_{n+m}\rangle \end{bmatrix} \tag{11}$$

and $\mathrm{vec}(\cdot)$ denotes the vectorization operation which stacks the columns of a matrix into a single column vector, $\otimes$ is the Kronecker product, and the polar factor $\boldsymbol{B}$, and the projectors $\boldsymbol{Q}_1, \boldsymbol{Q}_2$ are as in Lemma 4.1.

With such notation, the candidate dual certificate is more easily constructed. Indeed, in terms of vectorized constraints (9), the minimization (7) is equivalent to minimizing the 2-norm of the vector $\mathrm{vec}(\boldsymbol{M})$ subject to linear constraints $\boldsymbol{A}\mathrm{vec}(\boldsymbol{M}) = \boldsymbol{b}$. Therefore (see, e.g., [28, Ch.6]), when $\boldsymbol{A}$ is full row-rank (i.e., $\mathrm{rank}\{\boldsymbol{A}\} = m$), the candidate dual certificate $\boldsymbol{M}^*$ is unique and has an explicit expression in terms of $\boldsymbol{A}$ and $\boldsymbol{b}$ in Equation (9):

$$\mathrm{vec}(\boldsymbol{M}^*) = \boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{-1}\boldsymbol{b} \tag{12}$$

in addition, the Frobenius norm of the optimal solution is given by

$$\|\boldsymbol{M}^*\|_F^2 = \|\mathrm{vec}(\boldsymbol{M}^*)\|_2^2 = \boldsymbol{b}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{-1}\boldsymbol{b} \tag{13}$$

The proof of Theorem 3.3 can be summarized as follows:

1. For a fixed $\widehat{\boldsymbol{p}} = \widehat{\boldsymbol{p}}^{(\mathrm{RMIN})}$, we test its optimality for (NNMIN-ε) using Lemma 4.1 and the vectorized form of constraints (9).
2. For the case of the scaled structure $\mathcal{S}_\varepsilon(\boldsymbol{p})$, the matrices in Equation (9) also depend on ε (i.e., they become $\boldsymbol{A}(\varepsilon), \boldsymbol{b}(\varepsilon)$).
3. Under Assumptions 3.1–3.3, we show $\boldsymbol{A}(\varepsilon)$ is full row rank (hence (12) and (13) hold for the candidate dual certificate $\boldsymbol{M}^*(\varepsilon)$).
4. We then show that the norm $\|\boldsymbol{M}^*(\varepsilon)\|_F < 1$ for small ε, and hence $\|\boldsymbol{M}^*(\varepsilon)\|_2 \le \|\boldsymbol{M}^*(\varepsilon)\|_F < 1$, which implies that $\boldsymbol{M}^*(\varepsilon)$ is a valid dual certificate (in the sense of Lemma 4.1) for (NNMIN-ε).

5 | Proof of the Main Result

This section is organized as follows. We first discuss in Section 5.1 the implications of the assumptions of Theorem 3.3 and derive the form of the perturbed projectors and basis matrices. Based on lemmas in Section 5.1, we provide in Section 5.2 expressions for $\boldsymbol{A}(\varepsilon)$ and $\boldsymbol{b}(\varepsilon)$ from Equations (10) and (11). Finally, in Section 5.3, we give the proof of Theorem 3.3 and Proposition 4.3.

5.1 | Scaled Matrix Structure: Basis Matrices and Orthogonal Projectors

To get analogous expressions for (12) and (13) for the scaled matrix structure $\mathcal{S}_\varepsilon(\boldsymbol{p})$, we adapt (10) and (11) to the scaled problem (NNMIN-ε). Specifically, in this subsection, we provide details on the matrices $\boldsymbol{B}, \boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{S}_k$ for the scaled matrix structure $\mathcal{S}_\varepsilon(\boldsymbol{p})$. Note that the scaled matrix structure $\mathcal{S}_\varepsilon(\boldsymbol{p})$ is expressed as

$$\mathcal{S}_\varepsilon(\boldsymbol{p}) = \boldsymbol{S}_0(\varepsilon) + \sum_{k=1}^{N} p_k \boldsymbol{S}_k(\varepsilon)$$

Follow where $\boldsymbol{S}_k(\varepsilon)$ depend linearly on ε, which follows from the linearity of the scaling. We have the following lemma relating $\boldsymbol{S}_k(\varepsilon)$ to $\boldsymbol{S}_k$ for the missing values.

Lemma 5.1. *Under Assumption 3.3, the scaled basis matrices for the missing values can be expressed as*:

$$\boldsymbol{S}_k(\varepsilon) = \varepsilon \boldsymbol{S}_k, \quad \text{for all } k \in \{n+1, \ldots, n+m\} \tag{14}$$

Proof. Since the scaling (2) amounts just to multiplying by the matrix $\mathcal{S}(\boldsymbol{p})$ on the right and by linearity of (5) we have that, for all k,

$$\boldsymbol{S}_k(\varepsilon) = \boldsymbol{S}_k \begin{bmatrix} \boldsymbol{I}_{K-\kappa} & \boldsymbol{0} \\ \boldsymbol{0} & \varepsilon \boldsymbol{I}_\kappa \end{bmatrix}$$

Since, by Assumption 3.3, the limiting structure $\mathcal{S}_0(\boldsymbol{p})$ (and thus the first K-κ columns of $\mathcal{S}(\boldsymbol{p})$) does not depend on $p_{n+1}, \ldots, p_{n+m}$, we obtain (14). □

In what follows, we fix $\widehat{\boldsymbol{p}} = \widehat{\boldsymbol{p}}^{(\mathrm{RMIN})}$ and assume that one of Assumptions 3.2 and 3.3 holds. For the scaled matrix structure $\mathcal{S}_\varepsilon(\widehat{\boldsymbol{p}})$, we look at the corresponding projectors $\boldsymbol{Q}_1(\varepsilon), \boldsymbol{Q}_2(\varepsilon)$ (as defined in Section 4.1), and show that these projectors have the following form.

Lemma 5.2. *Let $\widehat{\boldsymbol{p}} = \hat{\boldsymbol{p}}^{(RMIN)}$ be fixed. Then*:

1. *Under Assumption 3.2, the projector $\boldsymbol{Q}_1(\varepsilon)$ on the left nullspace of $\mathcal{S}_\varepsilon(\widehat{\boldsymbol{p}})$ is constant for any ε (even, for $\varepsilon = 0$)*

   $$\boldsymbol{Q}_1(\varepsilon) = \boldsymbol{Q}_1$$

   where $\boldsymbol{Q}_1 = \boldsymbol{Q}_1(0)$ is the projector on the left nullspace of $\mathcal{S}_0(\widehat{\boldsymbol{p}})$.

2. *Under Assumption 3.3, the matrix $\boldsymbol{Q}_2(0)$ (the projector on the right nullspace of $\mathcal{S}_0(\widehat{\boldsymbol{p}})$) has the form*

   $$\boldsymbol{Q}_2(0) = \begin{bmatrix} * & \boldsymbol{0}_{(K-\kappa)\times\kappa} \\ \boldsymbol{0}_{\kappa\times(K-\kappa)} & \boldsymbol{I}_\kappa \end{bmatrix}$$

   in particular, $\boldsymbol{S}_k \boldsymbol{Q}_2(0) = \boldsymbol{S}_k$ for $k \in \{n+1, \ldots, n+m\}$

3. *Under Assumptions 3.2 and 3.3, the right projector $\boldsymbol{Q}_2(\varepsilon)$ has the expansion*

   $$\boldsymbol{Q}_2(\varepsilon) = \boldsymbol{Q}_2(0) + o(\varepsilon)$$

   and thus is continuous in a neighborhood of $\varepsilon = 0$.

Proof.

1. Due to Assumption 3.2, the range of $\mathcal{S}_\varepsilon(\boldsymbol{p})$ does not change with ε and is equal to the range of $\mathcal{S}_0(\widehat{\boldsymbol{p}})$. Therefore, the left nullspace also does not change with ε and $\boldsymbol{Q}_1(\varepsilon) = \boldsymbol{Q}_1(0)$.

2. Note that we have

   $$\mathcal{S}_0(\widehat{\boldsymbol{p}}) = \begin{bmatrix} * & \boldsymbol{0}_{L\times\kappa} \end{bmatrix}$$

   and therefore the right projector has the form

   $$\boldsymbol{Q}_2(0) = \boldsymbol{I}_K - \left(\mathcal{S}_0(\widehat{\boldsymbol{p}})^\dagger \mathcal{S}_0(\widehat{\boldsymbol{p}})\right) = \boldsymbol{I}_K - \begin{bmatrix} * & \boldsymbol{0}_{(K-\kappa)\times\kappa} \\ \boldsymbol{0}_{\kappa\times(K-\kappa)} & \boldsymbol{I}_\kappa \end{bmatrix} \tag{15}$$

   In particular, this implies that $\boldsymbol{S}_k \boldsymbol{Q}_2(0) = \boldsymbol{S}_k$ for $k \in \{n+1, \ldots, n+m\}$ since by Assumption 3.3 only the last κ columns of such $\boldsymbol{S}_k$ are nonzero.

3. By Assumption 3.2, the rank of $\mathcal{S}_\varepsilon(\widehat{\boldsymbol{p}})$ is constant for all $\varepsilon \in \mathbb{R}$. Therefore, the right nullspace projector $\boldsymbol{Q}_2(\varepsilon)$ is continuous and analytic in a neighborhood of 0 (for example, due to the analyticity of the projectors [29, Ch. II, Theorem 1.10]). □

Finally, we return to the unscaled case to demonstrate the implications of the assumptions on the solution of the rank minimization problem.

Lemma 5.3. *Under Assumptions 3.2 and 3.3, a vector $\widehat{\boldsymbol{p}} = \widehat{\boldsymbol{p}}^{(\mathrm{RMIN})}$ is a minimizer of (RMIN) if and only if it is the solution of the following linear system of equations*:

$$\begin{cases} \boldsymbol{Q}_1 \mathcal{S}(\widehat{\boldsymbol{p}}) = 0, \\ \widehat{\boldsymbol{p}}_{1:n} = \boldsymbol{p}_0 \end{cases} \tag{16}$$

where $\boldsymbol{Q}_1$ is the projector on the left nullspace of $\mathcal{S}_0(\boldsymbol{p}_0)$.

Proof. By Assumption 3.2, $\widehat{\boldsymbol{p}} = \widehat{\boldsymbol{p}}^{(\mathrm{RMIN})}$ is an optimal solution to (RMIN) if and only if

$$\operatorname{rank}\{\mathcal{S}(\widehat{\boldsymbol{p}})\} = \operatorname{rank}\{\mathcal{S}_0(\boldsymbol{p}_0)\} \tag{17}$$

Next, we notice that the nonzero part of $\mathcal{S}_0(\boldsymbol{p}_0)$ is a submatrix of $\mathcal{S}(\widehat{\boldsymbol{p}})$, and therefore (17) holds if and only if the column space (resp. left nullspace) of $\mathcal{S}(\widehat{\boldsymbol{p}})$ coincides with the column space (resp. the left nullspace) of $\mathcal{S}_0(\boldsymbol{p}_0)$, which happens if and only if (16) is satisfied. □

5.2 | Candidate Dual Certificate for the Scaled Case

To get closed-form expressions of the matrices in Equation (9), we require additional notation. Let $\boldsymbol{S}_k$ be the basis matrices in Equation (5). We define the following matrix $\boldsymbol{S}'$:

$$\boldsymbol{S}' = \begin{bmatrix} \operatorname{vec}(\boldsymbol{S}_{n+1}) & \cdots & \operatorname{vec}(\boldsymbol{S}_{n+m}) \end{bmatrix} \in \mathbb{R}^{LK\times m} \tag{18}$$

Using such notation, the following lemma holds true.

Lemma 5.4. *Under both Assumptions 3.3 and 3.2, we have that, for the scaled sturcture $\mathcal{S}_\varepsilon(\boldsymbol{p})$ the matrices $\boldsymbol{A}(\varepsilon)$ and $\boldsymbol{b}(\varepsilon)$ defined in Equation (9) have the form*

$$\boldsymbol{A}(\varepsilon) = \varepsilon(\boldsymbol{S}')^{\mathrm{T}}(\boldsymbol{Q}_2(\varepsilon) \otimes \boldsymbol{Q}_1^{\mathrm{T}}) = \varepsilon\left((\boldsymbol{S}')^{\mathrm{T}}(\boldsymbol{I} \otimes \boldsymbol{Q}_1^{\mathrm{T}}) + \varepsilon\right) \tag{19}$$

$$\boldsymbol{b}(\varepsilon) = -\varepsilon(\boldsymbol{S}')^{\mathrm{T}} \operatorname{vec}(\boldsymbol{B}(\varepsilon)) \tag{20}$$

Proof. We first note that

$$\begin{aligned} \boldsymbol{Q}_1(\varepsilon)\boldsymbol{S}_{n+k}(\varepsilon)\boldsymbol{Q}_2(\varepsilon) &= \varepsilon \boldsymbol{Q}_1 \boldsymbol{S}_{n+k} \boldsymbol{Q}_2(\varepsilon) = \varepsilon \boldsymbol{Q}_1 \boldsymbol{S}_{n+k}(\boldsymbol{Q}_2(0) + o(\varepsilon)) \\ &= \varepsilon(\boldsymbol{Q}_1 \boldsymbol{S}_{n+k} + o(\varepsilon)) \end{aligned}$$

where the last three equalities are due to Lemma 5.2. By vectorizing, rows of $\boldsymbol{A}(\varepsilon)$ can be expressed as

$$\begin{aligned}\mathrm{vec}(\boldsymbol{Q}_1\boldsymbol{S}_{n+1}\boldsymbol{Q}_2(\varepsilon)) &= \varepsilon(\boldsymbol{Q}_2^{\mathrm{T}}(\varepsilon)\otimes\boldsymbol{Q}_1)\mathrm{vec}(\boldsymbol{S}_{n+k})\\ &= \varepsilon\big((\boldsymbol{I}\otimes\boldsymbol{Q}_1)\mathrm{vec}(\boldsymbol{S}_{n+k})+o(\varepsilon)\big)\end{aligned}$$

which, after transposition, yields the desired result for $\boldsymbol{A}(\varepsilon)$. Similarly, for $\boldsymbol{b}(\varepsilon)$ we get

$$(\boldsymbol{b}(\varepsilon))_k = -\mathrm{vec}(\boldsymbol{S}_{n+k}(\varepsilon))^{\mathrm{T}}\mathrm{vec}(\boldsymbol{B}(\varepsilon)) = -\varepsilon\mathrm{vec}(\boldsymbol{S}_{n+k})^{\mathrm{T}}\mathrm{vec}(\boldsymbol{B}(\varepsilon))$$

which completes the proof. □

To use (12) and (13), we require that $\boldsymbol{A}(\varepsilon)$ is full row rank. The proof relies on Lemma 5.3 to show that the main term of (19) is of full rank.

Proposition 5.5. *Under Assumptions 3.1–3.3,* $\mathrm{rank}\{\boldsymbol{A}(\varepsilon)\} = m$ *in some neighborhood of* 0 *(for all* $\varepsilon \in (0, \varepsilon_1)$ *for some* ε_1*).*

Proof of Proposition 5.5. The matrix $\boldsymbol{A}(\varepsilon)$ has m rows, and therefore, by (19), it is sufficient to show that $\mathrm{rank}\{\boldsymbol{S}'^{\mathrm{T}}(\boldsymbol{I}\otimes\boldsymbol{Q}_1^{\mathrm{T}})\} = m$, which, by semi-continuity of the rank would imply $\mathrm{rank}\{\boldsymbol{A}(\varepsilon)\} = m$ in a neighborhood of 0.

To do that, we make use of Lemma 5.3. Let us take the parameterization of feasible $\widehat{\boldsymbol{p}}$ in Equation (RMIN) $\widehat{\boldsymbol{p}} = \widehat{\boldsymbol{p}}(\boldsymbol{z}) = (\boldsymbol{p}_0, \boldsymbol{z})$, for $\boldsymbol{z}\in\mathbb{C}^m$ and denote $\boldsymbol{x}_0 = \mathrm{vec}S(\widehat{\boldsymbol{p}}(\boldsymbol{0}))$. Then, by applying vectorization, we have that

$$\mathrm{vec}(S(\widehat{\boldsymbol{p}})) = \boldsymbol{x}_0 + \boldsymbol{S}'\boldsymbol{z}$$

Therefore, the condition (16) can be equivalently rewritten as $(\boldsymbol{I}\otimes\boldsymbol{Q}_1)(\boldsymbol{S}'\boldsymbol{z}+\boldsymbol{x}_0) = 0$, or after regrouping the terms,

$$(\boldsymbol{I}\otimes\boldsymbol{Q}_1)\boldsymbol{S}'\boldsymbol{z} = -(\boldsymbol{I}\otimes\boldsymbol{Q}_1)\boldsymbol{x}_0 \tag{21}$$

By Assumption 3.1, the solution of (21) exists and is unique, therefore, the matrix $(\boldsymbol{I}\otimes\boldsymbol{Q}_1)\boldsymbol{S}'$ must be full column rank, otherwise there is another solution to (21) (and therefore, another solution to (16)). But the matrix $((\boldsymbol{I}\otimes\boldsymbol{Q}_1)\boldsymbol{S}')$ is just the transpose of $(\boldsymbol{S}')^{\mathrm{T}}(\boldsymbol{I}\otimes\boldsymbol{Q}_1^{\mathrm{T}})$, therefore, the proof is complete. □

We finish this section with a lemma on the form of $\boldsymbol{b}(\varepsilon)$ from Equation (20).

Lemma 5.6. *Under Assumptions 3.1–3.3, the vector* $\boldsymbol{b}(\varepsilon)$ *is analytic in a small neighborhood of* 0 *and has an expansion*:

$$\boldsymbol{b}(\varepsilon) = \varepsilon^2(\boldsymbol{b}_2 + o(\varepsilon)) \tag{22}$$

Proof. We first show that $\boldsymbol{B}(\varepsilon)$ has a convergent power series expansion for real ε in a neighborhood of 0. For this we first note that (thanks to Lemma 5.2) there is a semi-unitary (i.e., $\boldsymbol{W}^{\mathrm{H}}\boldsymbol{W} = \boldsymbol{I}$) matrix $\boldsymbol{W}\in\mathbb{C}^{L\times r}$ spanning the column space of $S_\varepsilon(\widehat{\boldsymbol{p}})$ for any ε. Then we have that

$$S_\varepsilon(\widehat{\boldsymbol{p}}) = \boldsymbol{W}\boldsymbol{W}^{\mathrm{H}}S_\varepsilon(\widehat{\boldsymbol{p}}) \tag{23}$$

where $\boldsymbol{Z}(\varepsilon) := \boldsymbol{W}^{\mathrm{H}}S_\varepsilon(\widehat{\boldsymbol{p}})$ is full row rank for all $\varepsilon\in\mathbb{R}$. Therefore, the polar factor (for real ε) can be expressed as (see, e.g., [30])

$$\boldsymbol{B}(\varepsilon) = \boldsymbol{W}(\boldsymbol{Z}(\varepsilon)\boldsymbol{Z}^{\mathrm{H}}(\varepsilon))^{-\frac{1}{2}}\boldsymbol{Z}(\varepsilon)$$

and therefore has a convergent power series expansion for real ε in a neighborhood of 0:

$$\boldsymbol{B}(\varepsilon) = \boldsymbol{B}_0 + \varepsilon\boldsymbol{B}_1 + o(\varepsilon) \tag{24}$$

Now recall that the right block of $S_0(\widehat{\boldsymbol{p}})$ is zero by (3), hence,

$$\boldsymbol{B}_0 = \boldsymbol{W}(\boldsymbol{Z}(0)\boldsymbol{Z}^{\mathrm{H}}(0))^{-\frac{1}{2}}\boldsymbol{W}^{\mathrm{H}}S_0(\widehat{\boldsymbol{p}}) = \begin{bmatrix}* & \boldsymbol{0}_{L\times\kappa}\end{bmatrix}$$

Therefore, by Lemma 5.2 (i.e., from Assumption 3.3), we have

$$\langle\boldsymbol{B}_0, \boldsymbol{S}_k\rangle = 0 \text{ for all } k\in\{n+1,\ldots,n+m\}$$

Combining (20) and (24), we get $\boldsymbol{b}(\varepsilon) = \varepsilon^2(\boldsymbol{b}_2 + o(\varepsilon))$ with

$$\boldsymbol{b}_2 = -(\boldsymbol{S}')^{\mathrm{T}}\mathrm{vec}(\boldsymbol{B}_1) \tag{25}$$

which completes the proof. □

5.3 | The Proof of the Main Theorem

Proof of Theorem 3.3. Now let us denote by $\boldsymbol{M}^*(\varepsilon)$ the candidate dual certificate (Definition 4.2) for the scaled case. Thanks to Lemma 5.5, we have $\mathrm{rank}\{\boldsymbol{A}(\varepsilon)\} = m$ for small ε, therefore, by (13), we have that

$$\|\boldsymbol{M}^*(\varepsilon)\|_F^2 = \boldsymbol{b}^{\mathrm{H}}(\varepsilon)(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{H}}(\varepsilon))^{-1}\boldsymbol{b}(\varepsilon)$$

Note that from Equation (19), we have

$$\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{H}}(\varepsilon) = \varepsilon^2((\boldsymbol{S}')^{\mathrm{T}}(\boldsymbol{I}\otimes\boldsymbol{Q}_1^{\mathrm{T}})(\boldsymbol{S}') + o(\varepsilon))$$

in a neighborhood of 0, with $(\boldsymbol{S}')^{\mathrm{T}}(\boldsymbol{I}\otimes\boldsymbol{Q}_1^{\mathrm{T}})(\boldsymbol{S}')$ nonsingular by Lemma 5.5.

Therefore, we get that, from Equation (22):

$$\begin{aligned}\|\boldsymbol{M}^*(\varepsilon)\|_F^2 &= \varepsilon^2(\boldsymbol{b}_2+o(\varepsilon))^{\mathrm{H}}\big(\boldsymbol{A}(\varepsilon)\boldsymbol{A}^{\mathrm{H}}(\varepsilon)\big)^{-1}\varepsilon^2(\boldsymbol{b}_2+o(\varepsilon))\\ &= \varepsilon^2\Big(\boldsymbol{b}_2^{\mathrm{H}}\big((\boldsymbol{S}')^{\mathrm{T}}(\boldsymbol{I}\otimes\boldsymbol{Q}_1^{\mathrm{T}})(\boldsymbol{S}')\big)^{-1}\boldsymbol{b}_2 + o(1)\Big)\end{aligned}$$

which is analytic in a neighborhood of 0. This implies that there exists a neighborhood of 0 such that

$$\|\boldsymbol{M}^*(\varepsilon)\|_2^2 \le \|\boldsymbol{M}^*(\varepsilon)\|_F^2 < 1$$

which concludes the proof.

Finally, we give the proof of Proposition 4.3.

Proposition 5.7 (Proposition 4.3, reformulated). *Under Assumptions 3.1–3.3 let* $\widehat{\boldsymbol{p}} = \widehat{\boldsymbol{p}}^{(\mathrm{RMIN})}$*, and let* $\boldsymbol{M}^*(\varepsilon)$ *denote the candidate dual certificate for the scaled problem*(NNMIN-ε)*. Then, for*

small $\varepsilon > 0$, the candidate dual certificate exists, is unique, and its squared Frobenius norm (see (13)) has the expansion

$$\|\boldsymbol{M}^*(\varepsilon)\|_F^2 = m_0\varepsilon^2 + o(\varepsilon^2)$$

where m_0 is a constant defined as follows:

$$m_0 = \boldsymbol{b}_2^{\mathrm{H}}((\boldsymbol{S}')^{\mathrm{T}}(\boldsymbol{I} \otimes \boldsymbol{Q}_1^{\mathrm{T}})\boldsymbol{S}')^{-1}\boldsymbol{b}_2$$

The term $\boldsymbol{b}_2$ has the expansion $\boldsymbol{b}_2 = -(\boldsymbol{S}')^{\mathrm{T}}\mathrm{vec}(\boldsymbol{B}_1)$ (see (25), the coefficient corresponding to ε^2 from Equation (22)), where $\boldsymbol{B}_1$ is the first-order term in the expansion (24), which can be found as

$$\boldsymbol{B}_1 = \boldsymbol{U}_0\boldsymbol{\Sigma}_0^{-1}\boldsymbol{U}_0^{\mathrm{H}}\boldsymbol{\Delta} \tag{26}$$

where in Equation (26), $\boldsymbol{U}_0$ and $\boldsymbol{\Sigma}_0$ are the terms of the SVD of $\mathcal{S}_0(\widehat{\boldsymbol{p}}) = \boldsymbol{U}_0\boldsymbol{\Sigma}_0\boldsymbol{V}_0^{\mathrm{H}}$ and

$$\boldsymbol{\Delta} := \left(\mathcal{S}(\widehat{\boldsymbol{p}})\begin{bmatrix}\boldsymbol{0}_{K-\kappa} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_\kappa\end{bmatrix}\right)$$

is the notation that we introduce for the unscaled block of $\mathcal{S}_\varepsilon(\widehat{\boldsymbol{p}})$ (so that $\mathcal{S}_\varepsilon(\widehat{\boldsymbol{p}}) = \mathcal{S}_0(\widehat{\boldsymbol{p}}) + \varepsilon\boldsymbol{\Delta}$).

Proof. To show that $\boldsymbol{B}_1$ has the form (26), we denote $\boldsymbol{Z}(\varepsilon)$ as in Equation (23), with a particular choice of $\boldsymbol{W} = \boldsymbol{U}_0$. Recall that $\boldsymbol{Z}(\varepsilon)$ is a perturbation of a full-row-rank matrix $\boldsymbol{Z}_0$

$$\boldsymbol{Z}(\varepsilon) = \boldsymbol{Z}_0 + \varepsilon\boldsymbol{Z}_1$$

with

$$\boldsymbol{Z}_0 = \boldsymbol{U}_0^{\mathrm{H}}\mathcal{S}_0(\widehat{\boldsymbol{p}}), \quad \boldsymbol{Z}_1 = \boldsymbol{U}_0^{\mathrm{H}}\boldsymbol{\Delta}$$

where $\langle \boldsymbol{Z}_0, \boldsymbol{Z}_1\rangle_F = 0$. Then from [30, Lemma 3.2, equation (28)], we get that the polar factor of $\boldsymbol{Z}(\varepsilon)$ has expansion

$$\mathcal{P}(\boldsymbol{Z}(\varepsilon)) = \mathcal{P}(\boldsymbol{Z}_0) + \varepsilon(\boldsymbol{Z}_0\boldsymbol{Z}_0^{\mathrm{H}})^{-\frac{1}{2}}\boldsymbol{Z}_1 + o(\varepsilon)$$

Bringing it all together and using $\boldsymbol{Z}_0 = \boldsymbol{U}_0^{\mathrm{H}}\mathcal{S}_0(\widehat{\boldsymbol{p}}) = \boldsymbol{\Sigma}_0\boldsymbol{V}_0^{\mathrm{H}}$, we get

$$\boldsymbol{B}_1 = \boldsymbol{U}_0(\boldsymbol{Z}_0\boldsymbol{Z}_0^{\mathrm{H}})^{-\frac{1}{2}}\boldsymbol{Z}_1 = \boldsymbol{U}_0(\boldsymbol{\Sigma}_0^2)^{-\frac{1}{2}}\boldsymbol{Z}_1 = \boldsymbol{U}_0(\boldsymbol{\Sigma}_0)^{-1}\boldsymbol{U}_0^{\mathrm{H}}\boldsymbol{\Delta}$$

which completes the proof. □

6 | Numerical Examples

In this section, we describe some expository numerical examples describing how the scaling proposed enables exact recovery. We later focus on forecasting time series and data-driven simulation to offer just two application areas. Exact recovery guarantees obtained via nuclear norm minimization for structured matrices can be extended beyond these examples based on Hankel-type structures to other formats that naturally appear in diverse applications. For example, Toeplitz matrices—which often model convolution operators in imaging, communications, and system identification—arise when linear systems exhibit time-invariant or shift-invariant properties. Circulant matrices, which are a special case of Toeplitz matrices where each row is a cyclic shift of the previous one, are common in the analysis of digital filters and channel estimation problems because they diagonalize under the discrete Fourier transform. Such structures also emerge in array processing and radar signal detection, where block-Toeplitz or block-circulant matrices characterize multi-channel or spatial correlations. In these contexts, exact recovery results can be used to robustly interpolate missing measurements or reconstruct degraded signals while preserving the inherent matrix structure. This not only leads to more efficient computational algorithms (by exploiting fast Fourier transform techniques in the circulant case) but also reinforces the reliability of system identification and deconvolution methods in applications ranging from advanced medical imaging to network communications. Several application areas for which the result in this paper has significant consequences are also described in Reference [7] and the references therein.

In our current MATLAB implementation, we leverage CVX [31, 32], a software for convex optimization, which provides a rapid and versatile framework for our computations. This setup enables us to efficiently handle a variety of problem instances without extensive tailoring to specific applications. Due to the non-specific optimization routines we have used in this paper, we may experience some loss in efficiency since it is not specifically optimized for any particular problem or application. This generality means that while our approach can be applied to a wide range of scenarios, it might not be as finely tuned or perform as well as methods designed for specific tasks. However, if needed, our method can serve as an effective initial step for non-convex optimization routines.

Efficient algorithms for matrix completion leverage a variety of advanced computational techniques to improve performance and scalability. One primary approach involves first-order optimization methods, such as those described in Reference [33]. Randomized linear algebra also plays a crucial role in efficient matrix completion [34]. By using random sampling and projections, these methods can approximate the leading singular values and vectors with high accuracy while significantly reducing the computational load compared to traditional deterministic algorithms. The efficient computation of SVD is particularly important in matrix completion. Structured matrices further enhance the efficiency of SVD computations (e.g., [35]). Algorithms that exploit these structures can perform SVD more rapidly.

6.1 | The Rank-1 Case

In this example, we revisit the numerical study of [10], and take $p_k = \lambda^k$, $\lambda \in \mathbb{R}$, with structure as described in Equation (1). The aim is to investigate for which range of λ and ε the solutions of (RMIN) and (NNMIN-ε) coincide. It is shown in Reference [10, Thm. 6] that for the case with no matrix scaling (i.e., $\varepsilon = 1$) if $|\lambda| < 1$ then the solution of (NNMIN) is unique and coincides with the solution of (RMIN), namely $p_{n+k} = \lambda^{n+k}$. Otherwise, the solutions do not coincide. In this numerical exercise, we show that the scaling (2) and consequent solving of the scaled optimization problem (NNMIN-ε) indeed increases the range of $|\lambda|$ for which we get exact (perfect) recovery of missing values. We use (12) to compute candidate dual certificates to evidence this.

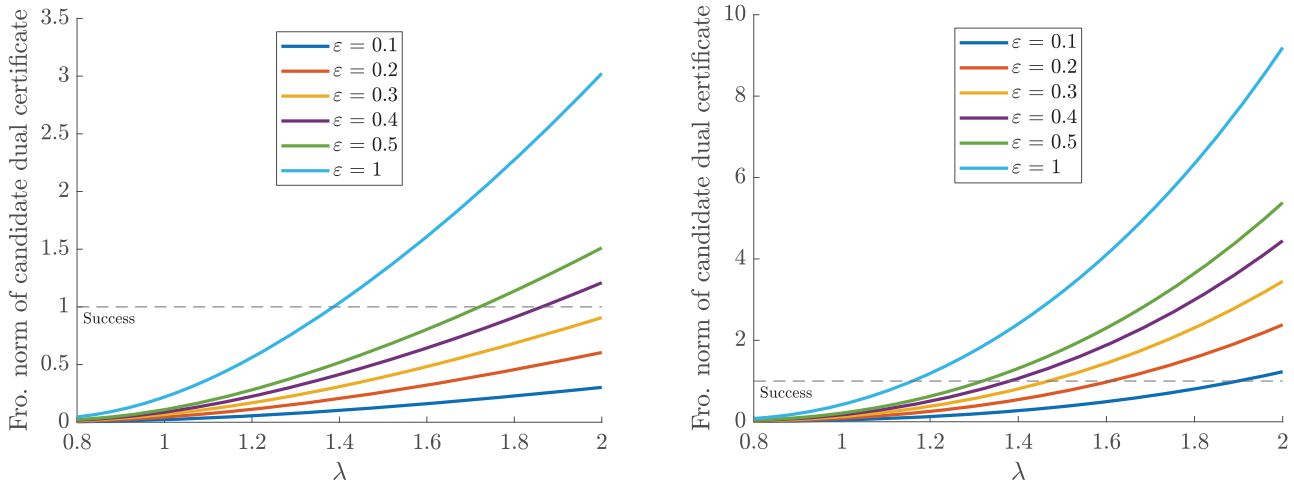


FIGURE 1 | Plot of the Frobenius norm of the candidate dual certificate as given in Equation (13) against λ for the form of $\boldsymbol{p}$ as described, with $n = 10$, $L = 4$, and $\kappa = m$.

Figure 1 is a plot of the Frobenius norm of the candidate dual certificate as given in Equation (13) against λ for the form of $\boldsymbol{p}$ as described, with $n = 10$, $L = 4$, $\kappa = m$, for different numbers of unknown observations m and scaling ϵ, as stated within the figure. The horizontal "dashed" lines denote the value when the Frobenius norm of the candidate dual certificate is one. By Lemma 4.1, when this norm is strictly less than one then the solution of (NNMIN-ε) is unique, coincides with that of (RMIN), and thus we have exact recovery of the m missing values. The range of λ for which we have exact recovery is a function of the scaling parameter ϵ; the range increases as more scaling (in the sense of taking ϵ small) is applied. For $\boldsymbol{p}$ with exponential growth, the more observations that are unknown, the more scaling is required for exact recovery. Note that the point at which the solution with no scaling ($\epsilon = 1$) intersects the horizontal "success" line, is the value of λ as given in Reference [13, Corollary 4.3], which gives bounds on λ for which the solution of (NNMIN) coincides with that of (RMIN).

6.2 | The Rank-r Case

We extend the example of the previous section for a more challenging problem, and instead take $p_k = \sum_{j=1}^{r} \lambda_j^k$, $\lambda_j \in \mathbb{R}$, with again Hankel structure as described in Equation (1). We take $m = 1$, $L = 10$, $n = 20$, $\kappa = 1$ and each λ_j is independently sampled uniformly in the range $(0, u)$, for some selected $u > 0$. We explore the influence of u and r and how they interact with the necessary amount of scaling (dictated by the parameter ϵ) to yield the Frobenius norm of the candidate dual certificate as given in Equation (13) to be smaller than 1.

Figure 2 contains plots (3-d and contour) of the proportion of times this Frobenius norm of the candidate dual certificate (13) exceeds one, with ϵ plotted against $r = 1, 2, \ldots, 8$, for different u. We take 10,000 Monte-Carlo simulations. The amount of scaling required for exact recovery of the missing values increases with r and u. This is consistent with the intuition as written in References [10] and [13], that the observed vector needs to be sufficiently damped for (NNMIN) to yield an identical solution to (RMIN).

In the examples considered so far, we have opted to compute the Frobenius norm of the candidate dual certificate, $\|\boldsymbol{M}^*\|_F$, as given in Equation (13), instead of its spectral norm $\|\boldsymbol{M}^*\|_2$ by virtue of Theorem 3.3, thus obtaining a weaker sufficient condition. We now investigate the gap between these norms. We take the same signal as described at the opening of this section, with $L = 10$, $n = 20$, $r = 3$, and $\kappa = m$, and inspect the ratio $\|\boldsymbol{M}^*\|_F / \|\boldsymbol{M}^*\|_2$ for different m and u.

Figures 3 and 4 contain boxplots of the ratio $\|\boldsymbol{M}^*\|_F / \|\boldsymbol{M}^*\|_2$ for different m and u (as stated in the caption), $\epsilon = 0.5$ and $\epsilon = 0.1$ respectively, taken over 10,000 Monte-Carlo simulations. The ratio appears more variable as the number of missing values m increases, and for small u the ratio seems to be a piece-wise linear function of m, increasing until beginning to (approximately) plateau after a particular number of missing observations. For small ϵ, the ratio is smaller.

We now discuss the result of Proposition 5.7. Using the notation as given in this proposition, Figure 5 is a plot of the ratio $\|\boldsymbol{M}^*\|_F / \epsilon\sqrt{m_0}$ against ϵ and r for the example considered with $\kappa = 1$, for different u, again taken over 10,000 Monte-Carlo simulations. The leading term approximation of the Frobenius norm of the candidate dual certificate $\boldsymbol{M}^*$ can be seen to be an excellent approximation, with deviation only for more "complex" time series for large r or u, but this deviation can be mitigated with a selection of a smaller ϵ.

6.3 | Forecasting Time Series

In this section, we provide figures directly showing the impact of scaling upon forecasting a given signal/time series. Take $p_k = \exp(\lambda k)\sin(2\pi\omega k)$, $\lambda = \omega = 0.1$, with Hankel structure as described in Equation (1). We take $m = 5$, $L = 4$, $n = 10$, $\kappa = 5$.

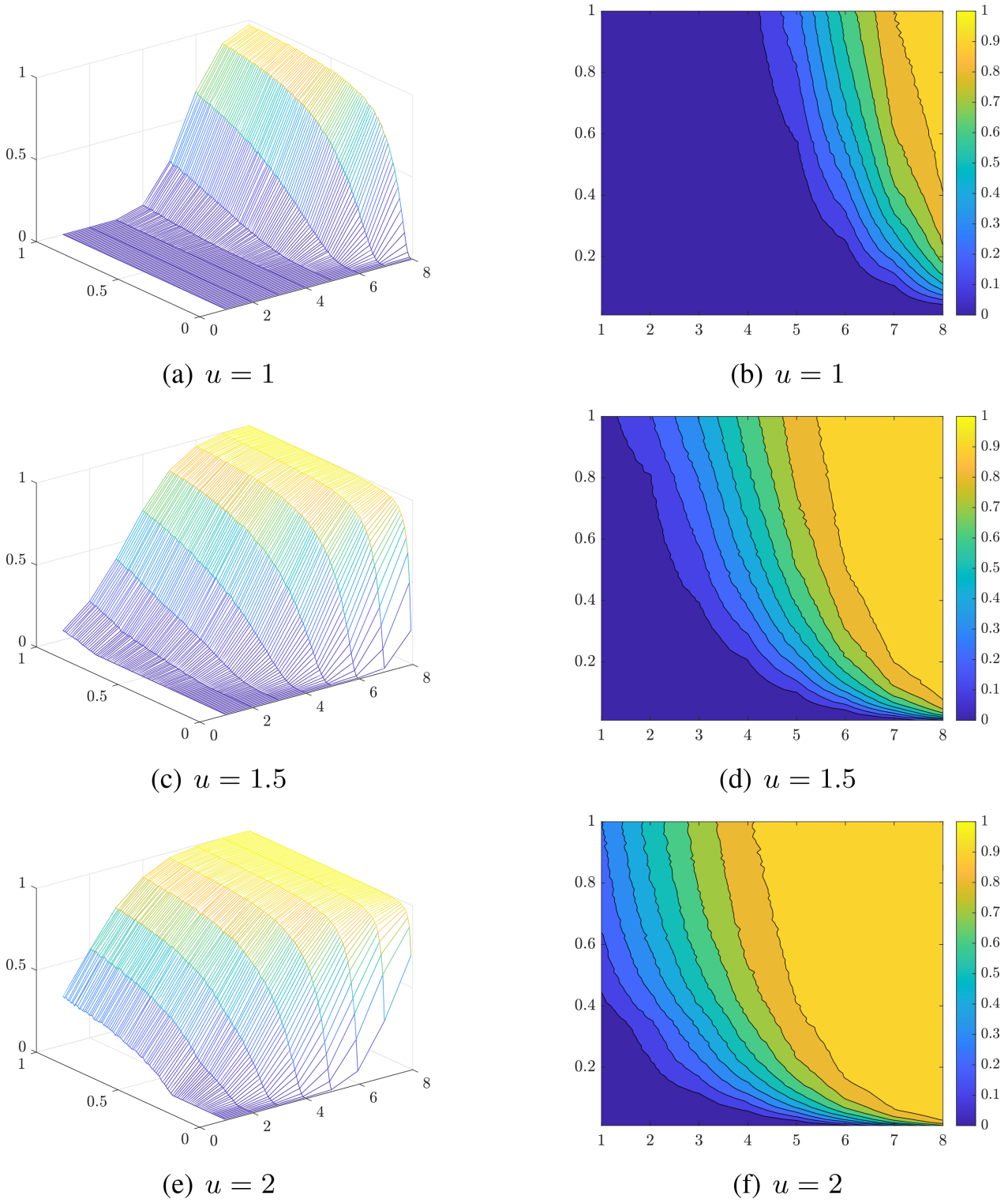
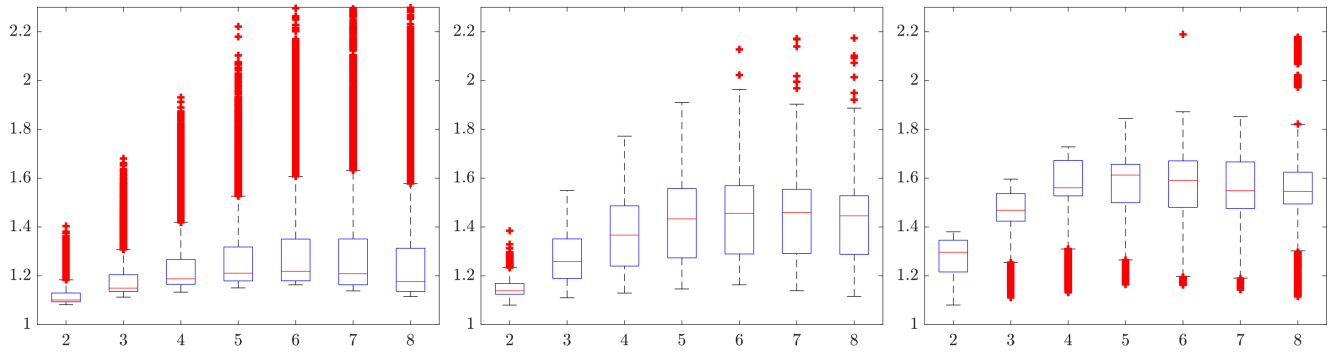


FIGURE 2 | Plots (3-d and contour) of the proportion of times the Frobenius norm of the candidate dual certificate as given in Equation (13) exceeds one, with ε (lhs) against $r = 1, 2, \ldots, 8$ (bottom), for different u. Parameter settings are $m = 1$, $L = 10$, $n = 20$, and $\kappa = 1$.

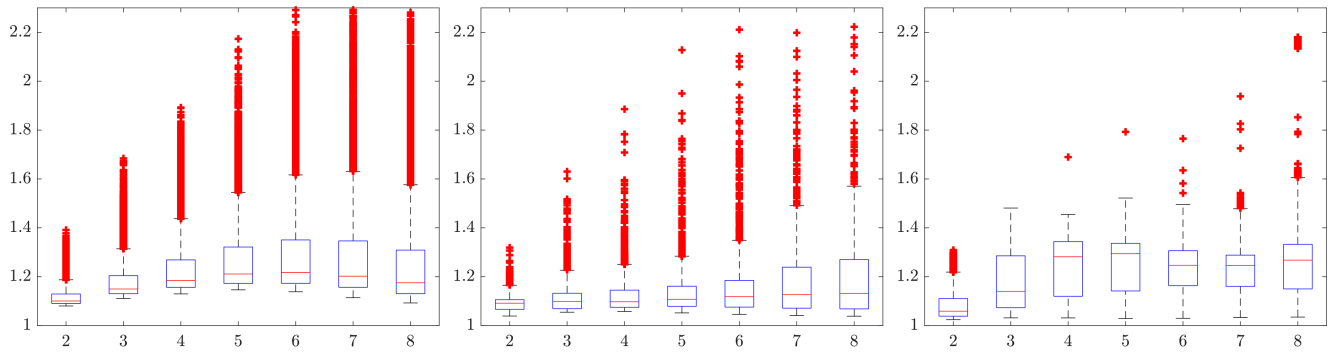
Practically, one may view the problem considered here as forecasting five observations ahead, having already recorded ten exact observations of a signal/time series.

Figure 6 contains plots of the signal $\boldsymbol{p}$ with the solutions obtained from Equations (NNMIN-ε) and (NNMIN) (i.e., nuclear norm minimization with and without scaling, respectively). The failure of (NNMIN) to exactly recover the time series is clearly evident, and the introduction of scaling gives exact recovery of the $m = 5$ missing values once ε is taken sufficiently small. The final value of $\varepsilon \approx 0.0694$ is that found from Equation (8), showing its value in selecting the scaling parameter ε in practice.



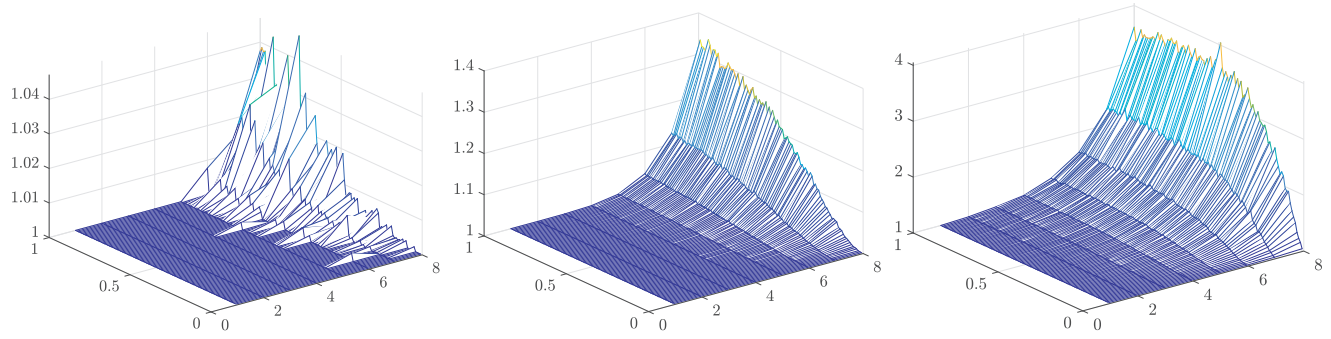
(a) $u = 1$ (b) $u = 3$ (c) $u = 7$

FIGURE 3 | Boxplots of the ratio $|\boldsymbol{M}^*|_F/|\boldsymbol{M}^*|_2$ for different m and u (as stated in the caption), $\epsilon = 0.5$.



(a) $u = 1$ (b) $u = 3$ (c) $u = 7$

FIGURE 4 | Boxplots of the ratio $\|\boldsymbol{M}^*\|_F/\|\boldsymbol{M}^*\|_2$ for different m and u (as stated in the caption), $\epsilon = 0.1$.

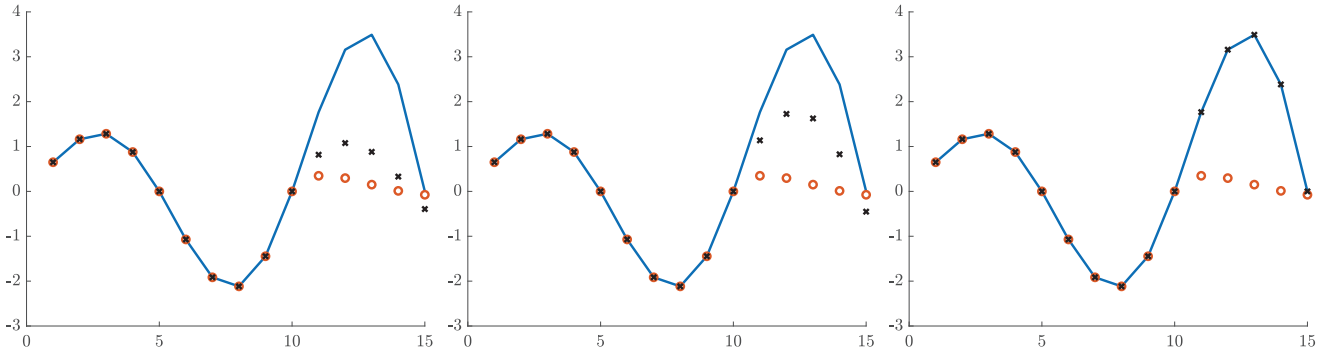


(a) $u = 0.5$ (b) $u = 1$ (c) $u = 2$

FIGURE 5 | Ratio $\|\boldsymbol{M}^*\|_F/\epsilon\sqrt{m_0}$ against ϵ and r for the example considered with $\kappa = 1$ and different u.

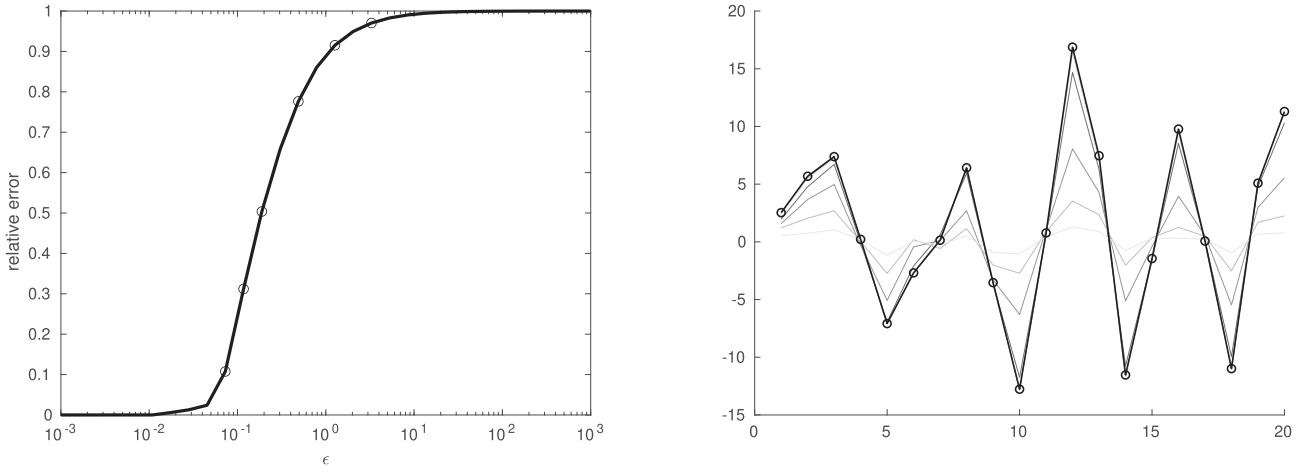
6.4 | Data-Driven Simulation

In this subsection, we describe an example of structured matrix completion that appears in modeling (linear time-invariant) dynamical systems (with p inputs and m outputs) that transform an input sequence $u(t) \in \mathbb{R}^p$ to the output sequence $y(t) \in \mathbb{R}^m$. The simulation problem [16] aims at computing the output time series for a given dynamical system, input, and initial conditions; the system parameters often have to be estimated prior to simulation. The data-driven approach to simulation [7, 27] avoids the step of estimating system parameters and is rather based on previously recorded data of inputs and outputs produced by the system. As shown in References [7, 27], data-driven simulation can be reformulated as structured matrix completion. More details on the data-driven simulation problem are provided in Appendix, and we give below just the necessary details to formulate the matrix completion problem.



(a) $\varepsilon = 0.25$ (b) $\varepsilon = 0.15$ (c) $\varepsilon = 0.0694$

FIGURE 6 | Plots of the signal $\boldsymbol{p}$ (blue curve) with the solutions obtained from Equation (NNMIN-ε) (black cross, with ϵ as given) and (NNMIN) (orange circle).



(a) Relative simulation error (b) Simulated output

FIGURE 7 | The relative reconstruction error in the data-driven simulation problem exhibits a transition to a zero error for sufficiently small values of the scaling parameter ϵ. This effect is shown on the simulated signal where large values of $\epsilon \gg 1$ result in a zero simulation output (relative error of 1), while small values of $\epsilon \ll 1$ lead to a perfect reconstruction of the signal in the noiseless case, indicated by the bold line connecting the circled points. We additionally show simulation results for intermediate values of ϵ in gray lines that have a nonzero relative error—their corresponding relative errors are indicated on the left panel by circles.

Let $\boldsymbol{u}_d \in \mathbb{R}^{m\times T_d}$ and $\boldsymbol{y}_d \in \mathbb{R}^{p\times T_d}$ be some known input-output pair of time series ("data")

$$\boldsymbol{u}_d = \big(\, \boldsymbol{u}_d(1), \ldots, \boldsymbol{u}_d(T_d) \,\big), \quad \boldsymbol{y}_d = \big(\, \boldsymbol{y}_d(1), \ldots, \boldsymbol{y}_d(T_d) \,\big)$$

and a new input time series $\boldsymbol{u}_s \in \mathbb{R}^{m\times T_s}$ the goal is to find the corresponding output $\boldsymbol{y}_s \in \mathbb{R}^{p\times T_s}$

$$\boldsymbol{u}_s = \big(\, \boldsymbol{u}_s(1), \ldots, \boldsymbol{u}_s(T_s) \,\big),$$

$$\boldsymbol{y}_s = \Big(\, \boldsymbol{y}_s(1), \ldots, \boldsymbol{y}_s(\ell), \underbrace{\boldsymbol{y}_s(\ell+1), \ldots, \boldsymbol{y}_s(T_s)}_{\text{unknown}} \,\Big)$$

assuming the knowledge of its first $\ell \ll T_s$ values (initial conditions).

By choosing the parameter $L \geq \ell + 1$ and denoting $K_d = T_d - L + 1$ and $K_s = T_s - L + 1$, the approach of [27] consists in minimizing the rank of the following $L(m+p)\times(K_d+K_s)$ matrix

$$\mathcal{M}_L = \begin{bmatrix}
\boldsymbol{u}_d(1) & \boldsymbol{u}_d(2) & \cdots & \boldsymbol{u}_d(K_d) & \boldsymbol{u}_s(1) & \boldsymbol{u}_s(2) & \iddots & \iddots & \cdots & \boldsymbol{u}_s(K_s) \\
\boldsymbol{u}_d(2) & \boldsymbol{u}_d(3) & \cdots & \vdots & \boldsymbol{u}_s(2) & \boldsymbol{u}_s(3) & \iddots & \iddots & \cdots & \vdots \\
\vdots & \iddots & \cdots & \vdots & \vdots & \iddots & \iddots & \iddots & \cdots & \vdots \\
\boldsymbol{u}_d(L) & \iddots & \cdots & \boldsymbol{u}_d(T_d) & \boldsymbol{u}_s(L) & \iddots & \iddots & \iddots & \cdots & \boldsymbol{u}_s(T_s) \\
\hline
\boldsymbol{y}_d(1) & \boldsymbol{y}_d(2) & \cdots & \boldsymbol{y}_d(K_d) & \boldsymbol{y}_s(1) & \cdots & \boldsymbol{y}_s(\ell) & \boldsymbol{y}_s(\ell+1) & \cdots & \boldsymbol{y}_s(K_s) \\
\boldsymbol{y}_d(2) & \boldsymbol{y}_d(3) & \cdots & \vdots & \vdots & \iddots & \iddots & \iddots & \cdots & \vdots \\
\vdots & \iddots & \cdots & \vdots & \boldsymbol{y}_s(\ell) & \iddots & \iddots & \iddots & \cdots & \vdots \\
\vdots & \iddots & \cdots & \vdots & \boldsymbol{y}_s(\ell+1) & \iddots & \iddots & \iddots & \cdots & \vdots \\
\vdots & \iddots & \cdots & \vdots & \vdots & \iddots & \iddots & \iddots & \cdots & \vdots \\
\boldsymbol{y}_d(L) & \iddots & \cdots & \boldsymbol{y}_d(T_d) & \boldsymbol{y}_s(L) & \iddots & \iddots & \iddots & \cdots & \boldsymbol{y}_s(T_s)
\end{bmatrix} \tag{27}$$

We consider a randomly generated single-input single-output ($m = p = 1$) linear time-invariant system of order $\ell = 6$ (generated by MATLAB's function DRSS). The input is a standard normal zero-mean white Gaussian noise sequence u_d of length $T_d = 40$, and we denote the corresponding (noise-free) output as y_d.

The simulation input u_s of length $T_s = 20 + \ell$ is generated in the same way, and the (noise-free) simulation output is denoted by $\tilde{y}_s$. All simulations assume to start from zero initial conditions. We set the number of block rows L of the Hankel matrix to $L = \ell + 1$.

To study the effect of the data scaling, we replace the data $\boldsymbol{w}_s = (\boldsymbol{u}_s, \boldsymbol{y}_s)$ by $\varepsilon \boldsymbol{w}_s$ and investigate the results for values of ε in the range $[10^{-3}, 10^3]$. The nuclear norm minimization using the proposed block-column scaling (NNMIN-ε) returns the (scaled) simulation output $\hat{\boldsymbol{y}}_s$, which is then compared to the true simulation output $\tilde{y}_s$ after rescaling the result by multiplying it with ε^{-1}. Figure 7 contains the result of this analysis, showing the relative simulation error $\|\hat{\boldsymbol{y}}_s - \tilde{\boldsymbol{y}}_s\|/\|\tilde{\boldsymbol{y}}_s\|$ with $\tilde{\boldsymbol{y}}_s$ the true simulation output. At large values $\varepsilon \gg 1$, the simulation signal is close to a zero signal (and the relative error is one). As ε decreases, the reconstruction error gradually decreases, and for a sufficiently small value $\varepsilon \ll 1$ the method returns the true simulation output $\tilde{y}_s$.

7 | Conclusion

We have introduced a technique to achieve exact recovery of missing entries of structured matrices using the commonly used nuclear norm relaxation of low-rank matrix completion, where the observations are recorded without noise. To do this, we use block-column scaling of the given matrix. As well as describing the theoretical basis of this approach, we have offered several examples showing the usefulness and merit of our proposed technique. We provide a result suggesting a value for our scaling parameter, which will be helpful for the practical use of our results. There is significant potential for synergizing our theoretical advancements with several practical applications in fields such as time-series analysis, data-driven simulation, and more. Future work will consider the hard problem of when observations are noisy and not exact, and derive conditions for when Assumption 3.1 is verifiable for general affine matrix structures.

Data Availability Statement

The authors have nothing to report.

Endnotes

[1] Missing values in the middle of $\boldsymbol{p}$ can be also treated, see discussion in Section 3.3.

References

1. D. Lazzaro, "A Nonconvex Approach to Low-Rank Matrix Completion Using Convex Optimization," *Numerical Linear Algebra with Applications* 23, no. 5 (2016): 801–824.

2. E. J. Candès and B. Recht, "Exact Matrix Completion via Convex Optimization," *Communications of the ACM* 55, no. 6 (2012): 111–119.

3. E. J. Candès and T. Tao, "The Power of Convex Relaxation: Near-Optimal Matrix Completion," *IEEE Transactions on Information Theory* 56, no. 5 (2010): 2053–2080.

4. Y.-H. Xiao and Z.-F. Jin, "An Alternating Direction Method for Linear-Constrained Matrix Nuclear Norm Minimization," *Numerical Linear Algebra with Applications* 19, no. 3 (2012): 541–554.

5. M. Carlsson, D. Gerosa, and C. Olsson, "An Unbiased Approach to Low Rank Recovery," *SIAM Journal on Optimization* 32, no. 4 (2022): 2969–2996.

6. I. Markovsky, "Structured Low-Rank Approximation and Its Applications," *Automatica* 44, no. 4 (2008): 891–909.

7. I. Markovsky, *Low-Rank Approximation* (Springer, 2019).

8. I. Markovsky, "How Effective is the Nuclear Norm Heuristic in Solving Data Approximation Problems?," *IFAC Proceedings Volumes* 45, no. 16 (2012): 316–321.

9. L. Dai and K. Pelckmans, "On the Nuclear Norm Heuristic for a Hankel Matrix Completion Problem," *Automatica* 51 (2015): 268–272.

10. K. Usevich and P. Comon, "Hankel Low-Rank Matrix Completion: Performance of the Nuclear Norm Relaxation," *IEEE Journal of Selected Topics in Signal Processing* 10, no. 4 (2016): 637–646.

11. E. J. Candès and Y. Plan, "Matrix Completion With Noise," *Proceedings of the IEEE* 98, no. 6 (2010): 925–936.

12. H. Butcher and J. Gillard, "Simple Nuclear Norm Based Algorithms for Imputing Missing Data and Forecasting in Time Series," *Statistics and Its Interface* 10, no. 3 (2017): 19–25.

13. J. Gillard and K. Usevich, "Structured Low-Rank Matrix Completion for Forecasting in Time Series Analysis," *International Journal of Forecasting* 34, no. 4 (2018): 582–597.

14. J. Gillard and K. Usevich, "Hankel Low-Rank Approximation and Completion in Time Series Analysis and Forecasting: A Brief Review," *Statistics and Its Interface* 16, no. 2 (2023): 287–303.

15. N. Golyandina, V. Nekrutkin, and A. A. Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques* (CRC press, 2001).

16. I. Markovsky, "A Missing Data Approach to Data-Driven Filtering and Control," *IEEE Transactions on Automatic Control* 62 (2017): 1972–1978.

17. P. Jain, S. Oh, P. Netrapalli, and S. Sanghavi, "Low-Rank Matrix Completion Using Alternating Minimization," in *Proceedings of the 45th Annual ACM Symposium on Theory of Computing* (STOC, 2013), 665–674.

18. B. Vandereycken, "Low-Rank Matrix Completion by Riemannian Optimization," *SIAM Journal on Optimization* 23, no. 2 (2013): 1214–1236.

19. J. Tanner and K. Wei, "Low Rank Matrix Completion by Alternating Steepest Descent Methods," *Applied and Computational Harmonic Analysis* 40 (2016): 417–429.

20. N. Srebro and R. R. Salakhutdinov, "Collaborative Filtering in a Non-Uniform World: Learning With the Weighted Trace Norm," in *Advances in Neural Information Processing Systems* (NIPS, 2010), 23.

21. Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward, "Coherent Matrix Completion," in *International Conference on Machine Learning* (PMLR, 2014), 674–682.

22. S. Foucart, D. Needell, R. Pathak, Y. Plan, and M. Wootters, "Weighted Matrix Completion From Non-Random, Non-Uniform Sampling Patterns," *IEEE Transactions on Information Theory* 67, no. 2 (2021): 1264–1290.

23. A. Tasissa and R. Lai, "Low-Rank Matrix Completion in a General Non-Orthogonal Basis," *Linear Algebra and Its Applications* 625 (2021): 81–112.

24. G. Heinig and K. Rost, *Algebraic Methods for Toeplitz-Like Matrices and Operators* (Birkhäuser, 1984).

25. I. Markovsky and F. Dörfler, "Data-Driven Dynamic Interpolation and Approximation," *Automatica* 135 (2022): 110008.

26. J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. M. De Moor, "A Note on Persistency of Excitation," *Systems & Control Letters* 54, no. 4 (2005): 325–329.

27. P. Dreesen and I. Markovsky, "Data-Driven Simulation Using the Nuclear Norm Heuristic," in *Proceedings of the 2019 IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP, 2019), 8207–8211.

28. S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, 2004).

29. T. Kato, *Perturbation Theory for Linear Operators*, 2nd corrected ed. (Springer-Verlag, 1995).

30. E. S. Gawlik and M. Leok, "Iterative Computation of the Fréchet Derivative of the Polar Decomposition," *SIAM Journal on Matrix Analysis and Applications* 38, no. 4 (2017): 1354–1379.

31. M. Grant and S. Boyd, "Graph Implementations for Nonsmooth Convex Programs," in *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, ed. V. Blondel, S. Boyd, and H. Kimura (Springer-Verlag Limited, 2008), 95–110, http://stanford.edu/~boyd/graph_dcp.html.

32. M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, Version 2.1," 2014, https://cvxr.com/cvx.

33. A. Beck, *First-Order Methods in Optimization* (SIAM, 2017).

34. N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding Structure With Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions," *SIAM Review* 53, no. 2 (2011): 217–288.

35. F. T. Luk and S. Qiao, "A Fast Eigenvalue Algorithm for Hankel Matrices," *Linear Algebra and Its Applications* 316, no. 1–3 (2000): 171–182.

36. J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis* (Springer Verlag, 2001).

37. G. A. Watson, "Characterization of the Subdifferential of Some Matrix Norms," *Linear Algebra and Its Applications* 170, no. 1 (1992): 33–45.

38. J. W. Polderman and J. C. Willems, *Introduction to Mathematical Systems Theory: A Behavioral Approach* (Springer, 1998).

Appendix A

Forecasting for Hankel Matrices

Proof of Lemma 3.5. Note that $S_0(\boldsymbol{p})$ depends only on $\boldsymbol{p}_0' = (\boldsymbol{p}_0)_{1:n+m-\kappa}$

$$S_0(\boldsymbol{p}) := \begin{bmatrix} \mathcal{H}_L(\boldsymbol{p}_0') & \boldsymbol{0} \end{bmatrix}$$

where $\mathcal{H}_L(\boldsymbol{p}_0')$ is a subblock of $\mathcal{H}_L(\boldsymbol{p}_0)$. Under the assumption of the lemma, thanks to [15, §5.3], the vector $\boldsymbol{p}_0'$ has a unique minimal rank completion for any number of forecasted values, and so does $\boldsymbol{p}_0$.

Subdifferential of the Nuclear Norm

Definition (Subdifferential of a function f at $\boldsymbol{x}$). For a convex function $f : \mathbb{R}^N \to \mathbb{R}$, the subdifferential [36, p. 167] of f at $\boldsymbol{x}$, denoted by $\partial f(\boldsymbol{x})$, is defined as the set

$$\partial f(\boldsymbol{x}) \stackrel{\text{def}}{=} \left\{ \boldsymbol{z} \in \mathbb{R}^N \,:\, f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{z}, \boldsymbol{y} - \boldsymbol{x} \rangle, \text{ for all } \boldsymbol{y} \in \mathbb{R}^N \right\}$$

Notice that if f is differentiable at $\boldsymbol{x}$, then the subdifferential contains a single element, namely the gradient, that is, $\partial f(\boldsymbol{x}) = \boldsymbol{\nabla} f(\boldsymbol{x})$.

The subdifferential of the nuclear norm of a matrix $\boldsymbol{X}$ is related to the SVD of $\boldsymbol{X}$ as follows. Consider a rank-r matrix $\boldsymbol{X} \in \mathbb{R}^{L \times K}$, and its compact SVD as

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$$

withorthonormal $\boldsymbol{U} \in \mathbb{R}^{L \times r}$ and $\boldsymbol{V} \in \mathbb{R}^{K \times r}$, and $\boldsymbol{\Sigma} \in \mathbb{R}^{r \times r}$ a diagonal matrix with the nonzero singular values. Let also $\boldsymbol{U}_\perp \in \mathbb{R}^{L \times (L-r)}$ and $\boldsymbol{V}_\perp \in \mathbb{R}^{K \times (K-r)}$ be the matrices containing the remaining left and right singular vectors (spanning the left and right nullspaces of $\boldsymbol{X}$, respectively). Then the subdifferential of the nuclear norm [37, p. 41] at $\boldsymbol{X}$ is the following set of matrices:

$$\partial \|\boldsymbol{X}\|_* = \left\{ \begin{bmatrix} \boldsymbol{U} & \boldsymbol{U}_\perp \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{Z} \end{bmatrix} \begin{bmatrix} \boldsymbol{V} & \boldsymbol{V}_\perp \end{bmatrix}^{\mathrm{T}} \,:\, \boldsymbol{Z} \in \mathbb{R}^{(L-r)\times(K-r)}, \|\boldsymbol{Z}\|_2 \leq 1 \right\}$$

In what follows, we use a slightly more compact notation. Denote by $\boldsymbol{B} \stackrel{\text{def}}{=} \boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}$ the *polar factor* of $\boldsymbol{X}$ and by $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ the orthogonal projectors on the left and right nullspace of $\boldsymbol{X}$ respectively:

$$\boldsymbol{Q}_1 = \boldsymbol{U}_\perp \boldsymbol{U}_\perp^{\mathrm{T}} = \boldsymbol{I}_L - \boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}, \quad \boldsymbol{Q}_2 = \boldsymbol{V}_\perp \boldsymbol{V}_\perp^{\mathrm{T}} = \boldsymbol{I}_K - \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}$$

Then the subdifferential of the nuclear norm at $\boldsymbol{X}$ can be expressed as

$$\partial \|\boldsymbol{X}\|_* = \left\{ \boldsymbol{B} + \boldsymbol{Q}_1 \boldsymbol{M} \boldsymbol{Q}_2 \,:\, \boldsymbol{M} \in \mathbb{R}^{L \times K} \text{ with } \|\boldsymbol{M}\|_2 \leq 1 \right\} \tag{A1}$$

Let $f(\boldsymbol{p}) = \|S(\boldsymbol{p})\|_*$ denote the function that maps coefficients $\boldsymbol{p}$ to the nuclear norm of the associated structured matrix $S(\boldsymbol{p})$. By the chain rule, the subdifferential of f at $\boldsymbol{p}$ is (see [10])

$$\partial f(\boldsymbol{p}) = \left\{ \begin{bmatrix} \langle \boldsymbol{S}_1, \boldsymbol{H} \rangle_F & \cdots & \langle \boldsymbol{S}_N, \boldsymbol{H} \rangle_F \end{bmatrix}^{\mathrm{T}} \right. \\ \left. : \, \boldsymbol{H} \in \partial \|\boldsymbol{X}\|_* \text{ and } \boldsymbol{X} = S(\boldsymbol{p}) \right\} \tag{A2}$$

where $\langle \cdot, \cdot \rangle_F$ denotes the Frobenius inner product. Combining (A1) with (A2) and the standard necessary and sufficient conditions for convex optimization problems, we can easily obtain conditions for the structured nuclear norm minimization case.

Linear Dynamical Systems and Block-Hankel Matrices

Willems' behavioral system theory [38] defines a system $\mathcal{B}$ as the set of its admissible trajectories $\boldsymbol{w}$. A trajectory $\boldsymbol{w}$ of a discrete-time q-variate system $\mathcal{B}$ (in shorthand notation "$\boldsymbol{w} \in \mathcal{B}$") is a sequence $\boldsymbol{w} = (\boldsymbol{w}(1), \ldots, \boldsymbol{w}(T))$, with $\boldsymbol{w}(t) \in \mathbb{R}^q$ for $t = 1, \ldots, T$. In this paper, we consider the class of q-variate linear time-invariant (LTI) systems $\mathcal{L}^q$. An LTI system $\mathcal{B} \in \mathcal{L}^q$ (with m inputs and p outputs, such that $q = p + m$) has a kernel representation

$$\mathcal{B} = \left\{ \boldsymbol{w} \mid \boldsymbol{R}_0 \boldsymbol{w}(t) + \cdots + \boldsymbol{R}_\ell \boldsymbol{w}(t + \ell) = \boldsymbol{0}, \text{ for } t \geq 1 \right\} \tag{A3}$$

where $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_0 & \boldsymbol{R}_1 & \ldots & \boldsymbol{R}_\ell \end{bmatrix}$, in which $\boldsymbol{R}_i \in \mathbb{R}^{p \times q}$ is a kernel parameter that specifies the system. This formulation can be viewed as a difference equation describing the admissible trajectories $\boldsymbol{w} \in \mathcal{B}$. The minimal value for ℓ for which (A3) holds, is an invariant of the system $\mathcal{B}$ and is called the *lag*.

For a trajectory $\boldsymbol{w}$, the block-Hankel matrix is defined similarly to (1) as

$$\mathcal{H}_L(\boldsymbol{w}) \stackrel{\text{def}}{=} \begin{bmatrix} \boldsymbol{w}(1) & \boldsymbol{w}(2) & \cdots & \boldsymbol{w}(T-L+1) \\ \boldsymbol{w}(2) & \boldsymbol{w}(3) & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ \boldsymbol{w}(L) & \boldsymbol{w}(L+1) & \cdots & \boldsymbol{w}(T) \end{bmatrix} \in \mathbb{R}^{qL \times (T-L+1)} \tag{A4}$$

The block-Hankel matrix captures into the language of linear algebra the linear time-invariance of $\mathcal{B}$, and is closely connected to the kernel representation of $\mathcal{B} \in \mathcal{L}^q$. For a trajectory $\boldsymbol{w} \in \mathcal{B}$, the block-Hankel matrix $\boldsymbol{H}_L(\boldsymbol{w})$ with $L \geq \ell + 1$ is (row) rank-deficient, since $\boldsymbol{R}\boldsymbol{H}_{\ell+1}(\boldsymbol{w}) = \boldsymbol{0}$. More precisely, the rank of the block-Hankel matrix $\boldsymbol{H}_L(\boldsymbol{w})$ is bounded [7]

$$\operatorname{rank} \boldsymbol{H}_L(\boldsymbol{w}) \leq mL + p\ell \tag{A5}$$

for $T \gg L$, and under the persistency of excitation condition.

A trajectory $\boldsymbol{w} \in \mathcal{B}$ admits a partitoning into inputs and outputs: that is, there exists two time series $\boldsymbol{u}(t) \in \mathbb{R}^m$ (input) and $\boldsymbol{y}(t) \in \mathbb{R}^p$ (output), so

that the vector $\boldsymbol{w}(t)$ is a permutation of the elements of the vector $\begin{bmatrix}\boldsymbol{u}(t)\\ \boldsymbol{y}(t)\end{bmatrix}$; we will write $\boldsymbol{w} = (\boldsymbol{u}, \boldsymbol{y})$ for short and assume that the permutation is known.

Remark A2 (SISO LTI systems). The case $m = p = 1$ is referred to as single-input-single-output (SISO); in this case, the lag ℓ is equal to the system order. Moreover, the block-Hankel matrix has rank $\operatorname{rank} \boldsymbol{H}_{\ell+1}(\boldsymbol{w}) \leq 2\ell + 1$ (rank defeciency 1), with generically equality.

Data-Driven Simulation as Matrix Completion

The simulation of a system for a given input signal can be stated as follows. Given a system $\mathcal{B} \in \mathcal{L}^q$, an input $\boldsymbol{u}_s$, and initial conditions $\boldsymbol{w}_i$

$$\boldsymbol{w}_i = \big(\ \boldsymbol{w}_i(-\ell+1), \boldsymbol{w}_i(-\ell+2), \ldots, \boldsymbol{w}_i(-1), \boldsymbol{w}_i(0)\ \big) \tag{A6}$$

find the output $\boldsymbol{y}_s$ such that $\boldsymbol{w}_i \wedge \boldsymbol{w}_s = \boldsymbol{w}_i \wedge (\boldsymbol{u}_s, \boldsymbol{y}_s) \in \mathcal{B}$, where $\wedge$ denotes concatenation of time series (trajectories). This is a basic problem in system theory and is studied in various formulations. In *data-driven simulation*, the system $\mathcal{B}$ is defined implicitly by a given trajectory $\boldsymbol{w}_d = \big(\ \boldsymbol{w}_d(1), \ldots, \boldsymbol{w}_d(T_d)\ \big) \in \mathcal{B}$, where "$d$" stands for data. In this context, the simulation of the output $\boldsymbol{y}_s = \big(\ \boldsymbol{y}_s(1), \ldots, \boldsymbol{y}_s(T_s)\ \big)$ for a given input $\boldsymbol{u}_s = \big(\ \boldsymbol{u}_s(1), \ldots, \boldsymbol{u}_s(T_s)\ \big)$, where "$s$" stands for simulation, is done without first identifying (i.e., estimating parameters of) the system $\mathcal{B}$. We assume that the input $\boldsymbol{u}_d$ of the given trajectory $\boldsymbol{w}_d = (\boldsymbol{u}_d, \boldsymbol{y}_d)$ is persistently exciting, so $\boldsymbol{w}_d$ completely specifies the system $\mathcal{B}$ [26]. The data-driven simulation problem can then formally be stated as follows.

Problem A3 (Exact data-driven simulation). **Given** a trajectory $\boldsymbol{w}_d = (\boldsymbol{u}_d, \boldsymbol{y}_d) \in \mathcal{B} \in \mathcal{L}^q$, an input u_s, and initial conditions $\boldsymbol{w}_i$, **find** the output y_s, such that $\boldsymbol{w}_i \wedge \boldsymbol{w}_s = \boldsymbol{w}_i \wedge (\boldsymbol{u}_s, \boldsymbol{y}_s) \in \mathcal{B}$.

Remark A4. If we assume that the initial conditions are zero, that is, $\boldsymbol{w}_i \equiv \boldsymbol{0}$, the specification of initial conditions is done by prepending the trajectory $\boldsymbol{w}$ with ℓ zeros.

The rank deficiency of the Hankel matrix $\boldsymbol{H}_L(\boldsymbol{w})$, for $L \geq \ell + 1$, is closely related to the kernel representation of the system $\mathcal{B}$ and its interpretation in terms of difference equations. The mosaic Hankel matrix $\big[\boldsymbol{H}_L(\boldsymbol{w}')\ \ \boldsymbol{H}_L(\boldsymbol{w}'')\big]$ built from two trajectories $\boldsymbol{w}', \boldsymbol{w}'' \in \mathcal{B}$ has the same rank as $\boldsymbol{H}_L(\boldsymbol{w})$, since $\boldsymbol{w}'$ and $\boldsymbol{w}''$ satisfy the same difference equations, so $\operatorname{rank}\ \big[\boldsymbol{H}_L(\boldsymbol{w}')\ \ \boldsymbol{H}_L(\boldsymbol{w}'')\big] \leq mL + p\ell$.

This observation is crucial for formulating the data-driven simulation problem as a matrix completion problem. Putting $\boldsymbol{w}_d = (\boldsymbol{u}_d, \boldsymbol{y}_d)$ and $\boldsymbol{w}_s = (\boldsymbol{u}_s, \boldsymbol{y}_s)$ (where $\boldsymbol{w}_s$ is only partially known), the unknown output trajectory $\boldsymbol{y}_s$ should be determined such that

$$\operatorname{rank}\ \Big[\boldsymbol{H}_L(\boldsymbol{w}_d)\ \ \boldsymbol{H}_L(\boldsymbol{w}_s)\Big] \leq mL + p\ell, \ \text{for}\ L \geq \ell + 1 \tag{A7}$$

implying that both trajectories belong to the same system $\mathcal{B}$. Remark that again we need to specify initial conditions to uniquely determine the simulation output $\boldsymbol{y}_s$. If we denote by $\boldsymbol{w}_s' = \boldsymbol{w}_i \wedge \boldsymbol{w}_s$ the simulation trajectory prepended with initial conditions, then data-driven simulation problem can be phrased as the following block-Hankel matrix completion problem.

Problem A5 (Data-driven simulation via Hankel completion). **Given** a trajectory $\boldsymbol{w}_d = (\boldsymbol{u}_d, \boldsymbol{y}_d) \in \mathcal{B} \in \mathcal{L}^q$, an input $\boldsymbol{u}_s$, and initial conditions $\boldsymbol{w}_i$, **find** the output $\boldsymbol{y}_s$ from the following minimization problem.

$$\underset{\boldsymbol{y}_s}{\text{minimize}}\ \operatorname{rank}\ \Big[\boldsymbol{H}_L(\boldsymbol{w}_d)\ \ \boldsymbol{H}_L(\boldsymbol{w}_s')\Big] \tag{A8}$$

where $\boldsymbol{w}_s = (\boldsymbol{u}_s, \boldsymbol{y}_s)$.

By rearranging the rows of the mosaic Hankel matrix in Equation (A8), it is easy to see that Problem A.5 is equivalent to rank minimization of the matrix in Equation (27) (where the fixed element in the right blocks in Equation (27) are due to initial conditions).