



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Weighted Matrix Completion and Recovery with Prior Subspace Information

Citation for published version:

Eftekhari, A, Yang, D & Wakin, MB 2016 'Weighted Matrix Completion and Recovery with Prior Subspace Information'.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Early version, also known as pre-print

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Weighted Matrix Completion and Recovery with Prior Subspace Information

Armin Eftekhari, Dehui Yang, and Michael B. Wakin

Abstract—A low-rank matrix with “diffuse” entries can be efficiently reconstructed after observing a few of its entries, at random, and then solving a convex program. In many applications, in addition to these measurements, potentially valuable prior knowledge about the column and row spaces of the matrix is also available to the practitioner. In this paper, we incorporate this prior knowledge in matrix completion—by minimizing a weighted nuclear norm—and precisely quantify any improvements. In particular, in theory, we find that reliable prior knowledge reduces the sample complexity of matrix completion by a logarithmic factor; the observed improvement is considerably more magnified in numerical simulations. We also present similar results for the closely related problem of matrix recovery from generic linear measurements.

I. INTRODUCTION

Matrix completion is commonly defined as the problem of recovering a low-rank matrix $M \in \mathbb{R}^{n_1 \times n_2}$ from a fraction of its entries, observed on an often random index set [1], [2]. More concretely,¹ let $n_1 = n_2 = n$ and set $r = \text{rank}(M)$ for short. Also let $M = U_r \Sigma_r V_r^*$ be the singular value decomposition (SVD) of M , where $U_r, V_r \in \mathbb{R}^{n \times r}$ have orthonormal columns and the diagonal matrix $\Sigma_r \in \mathbb{R}^{r \times r}$ contains the singular values of M .

In a typical low-rank matrix completion problem, each entry of M is observed with a probability of $p \in (0, 1]$ so that, in expectation, pn^2 entries of M are revealed. Let $Y = \mathcal{R}_p(M) \in \mathbb{R}^{n \times n}$ contain the observed entries of M (with zeros everywhere else) with $\mathcal{R}_p(\cdot)$ representing the measurement process. In fact, with overwhelming probability, M can be successfully reconstructed from the measurements Y by solving the convex program

$$\begin{cases} \min_X \|X\|_* , \\ \text{subject to } \mathcal{R}_p(X) = Y, \end{cases} \quad (1)$$

provided that²

$$\frac{\eta(M)r \log^2 n}{n} \lesssim p \leq 1. \quad (2)$$

AE is with the Alan Turing Institute in London. DY and MBW are with the Department of Electrical Engineering and Computer Science at Colorado School of Mines. (Email: aeftekhari@turing.ac.edu, dyang@mines.edu, mwakin@mines.edu). This work was partially supported by NSF grant CCF-1409258 and NSF CAREER grant CCF-1149225.

¹In an attempt to make the introduction as accessible as possible, technical details are kept to a minimum in this section. More details are deferred to Section II.

²Throughout, we often use \lesssim to simplify the presentation by suppressing universal constants. We should also point out that, to improve accessibility, notation in the Introduction is slightly different from the rest of the paper and the mathematics is less rigorous.

Above, the nuclear norm $\|X\|_*$ returns the sum of singular values of a matrix X . In addition, $\eta(M)$, the *coherence* of M , measures how “spiky” M is, as precisely defined later in Section II. Roughly speaking, then, one can expect to successfully recover M from $O(\eta(M) \cdot rn \log^2 n)$ uniform samples [3], [4].

A. Incorporating Prior Knowledge

Let $U_r = \text{span}(U_r)$ and $V_r = \text{span}(V_r)$ be the column and row spaces of $M = U_r \Sigma_r V_r^*$.³ Suppose that we have been presented with some prior knowledge about M in the form of estimates for the subspaces U_r and V_r . More specifically, let the r -dimensional subspaces \tilde{U}_r and \tilde{V}_r be the initial estimates of the column and row spaces of M , respectively, made available to us.

For example, \tilde{U}_r and \tilde{V}_r might represent the similarities among users and movies, respectively, in the “Netflix challenge.” The rows and columns of the popular Netflix matrix correspond to the Netflix subscribers and available movies, respectively. The Netflix matrix is sparsely populated with the ratings assigned by its users. The challenge is then to complete the Netflix matrix given only the ratings available, namely given only a small fraction of its entries. A more realistic setup for the Netflix challenge should perhaps incorporate the changes in the preferences of Netflix users over time which, for example, are significantly altered by child-rearing. Here, \tilde{U}_r and \tilde{V}_r might incorporate prior information about the users and movies. Similar problems arise in tracking changes in videos or updating the Laplacian of a graph with time-variant connectivity. See also [5] for more examples in the context of *collaborative filtering* and related topics.

As another example, in *exploration seismology*, large and often incomplete matrices are acquired and processed in order to determine the subsurface structure of an area. Each matrix is comprised of responses from many sources (at a certain frequency) recorded at many receivers, where some recordings are missing. In this context, information from adjacent frequency bands might help enhance matrix completion. That is, one might set \tilde{U}_r and \tilde{V}_r to be the column and row spaces of the response matrix from adjacent frequency bands [6].

Lastly, consider the problem of *subspace tracking* from limited measurements, in which we are interested in recovering a subspace after partially observing a sequence of generic vectors in that subspace. Here, we might consider $\tilde{U}_r = U_r^{(k-1)}$ to be our current estimate of the underlying subspace U_r , which we wish to update in the k th iteration. Before moving on, we

³We reserve upright letters for the corresponding subspaces.

point out that the related literature is more carefully studied later in Section V.

Motivated by such scenarios, it is perhaps natural to ask:

- **Question:** How should we incorporate in matrix completion any prior knowledge about column and row spaces?

We approach this question with the aid of a weighted nuclear norm as follows. Let $P_{\tilde{U}_r}, P_{\tilde{U}_r^\perp} \in \mathbb{R}^{n \times n}$ be the orthogonal projections onto the subspace \tilde{U}_r and its orthogonal complement \tilde{U}_r^\perp , respectively. For some weight $w \in (0, 1]$, define

$$Q_{\tilde{U}_r, w} := w \cdot P_{\tilde{U}_r} + P_{\tilde{U}_r^\perp} \in \mathbb{R}^{n \times n}. \quad (3)$$

Likewise, define $Q_{\tilde{V}_r} \in \mathbb{R}^{n \times n}$ and let us modify Program (1) to read

$$\begin{cases} \min_X \left\| Q_{\tilde{U}_r} \cdot X \cdot Q_{\tilde{V}_r} \right\|_*, \\ \text{subject to } \mathcal{R}_p(X) = Y. \end{cases} \quad (4)$$

In a sense, the weight w reflects our uncertainty in the prior knowledge. The smaller w , the more confident we are that $U_r \approx \tilde{U}_r$ and $V_r \approx \tilde{V}_r$ and, in turn, the more penalty Program (4) places on feasible matrices with column or row spaces orthogonal to, respectively, \tilde{U}_r or \tilde{V}_r . In contrast, when our prior information is not reliable, we might set $w = 1$, in which case $Q_{\tilde{U}_r} = Q_{\tilde{V}_r} = I_n$ and Program (4) reduces to standard matrix completion (namely Program (1)), thereby completely ignoring any prior knowledge about the problem.

A more general form of Program (4) is discussed in Section II in greater detail. For now, let us briefly compare the two Programs (1) and (4) in practice. Fix $n = 20, r = 3$, and set $M = U_r \Sigma_r V_r^*$ where $U_r, V_r \in \mathbb{R}^{n \times r}$ are generic matrices with orthonormal columns, and $\Sigma_r \in \mathbb{R}^{r \times r}$ is diagonal with its nonzero entries drawn independently from the uniform distribution on the interval $(0, 1]$. As for the prior information, we set \tilde{U}_r to be the column span of $U_r + 0.01G$, with $G \in \mathbb{R}^{n \times r}$ populated by independent zero-mean random Gaussian variables with unit variance. \tilde{V}_r is constructed likewise. Lastly, we set $w = 0.1$. As the probability p of observing every entry of M varies in $(0, 1]$, we solve both Programs (1) and (4) and record the results. The success rates for both programs, averaged over 100 trials, is shown in Figure 1. A trial is considered successful if it recovers M up to a relative error of 10^{-3} . Observe how reliable prior knowledge, when used properly, allows for successful matrix completion from substantially fewer measurements.

B. Simplified Main Result

One of our main results in this paper concerns the performance of Program (4), and might be considered as a special case of the more general result presented in Theorem 2 (in Section III).

Consider a rank- r matrix $M \in \mathbb{R}^{n \times n}$, and let $U_r = \text{span}(M)$ and $V_r = \text{span}(M^*)$ be the column and row spaces of M . Let also $\eta(M)$ be the coherence of M , which we briefly introduced earlier.

Suppose also that the r -dimensional subspaces \tilde{U}_r and \tilde{V}_r , with orthonormal bases $\tilde{U}_r, \tilde{V}_r \in \mathbb{R}^{n \times r}$, represent our prior

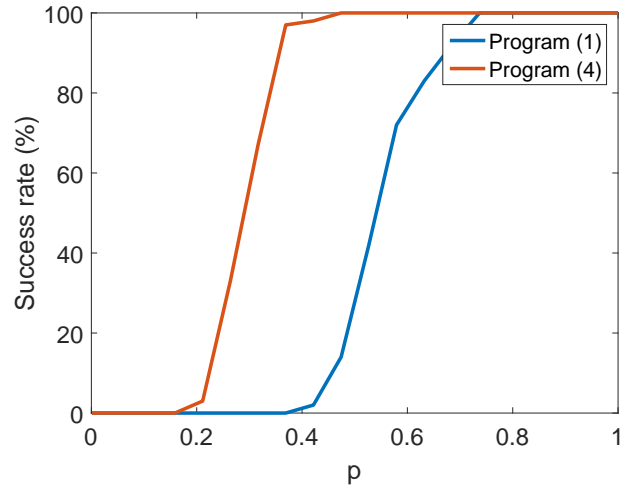


Figure 1: Leveraging prior information can substantially assist in matrix completion, as demonstrated in this figure. As the probability p of observing every entry of a generic rank-3 matrix $M \in \mathbb{R}^{20 \times 20}$ varies in $(0, 1]$, we plot the success rate of standard matrix completion (Program (1)) and matrix completion aided by prior knowledge (Program (4)) over many trials. See Section I-A for more details.

knowledge about the column and row spaces of M , respectively. In particular, let $u = \angle[U_r, \tilde{U}_r]$ and $v = \angle[V_r, \tilde{V}_r]$ denote the largest principal angles between each pair of subspaces.⁴ We set $\theta = \max[u, v]$. Moreover, consider the probability $p \in (0, 1]$ and let $Y = \mathcal{R}_p(M)$ be the matrix of measurements, defined earlier. Lastly, for weight $w \in (0, 1]$, let \tilde{M} be a solution of Program (4). Then, $\tilde{M} = M$, except with a probability of $o(n^{-19})$ and provided that⁵

$$1 \geq p \gtrsim \max[\log(\alpha_1 \cdot n), 1] \cdot \frac{\eta(M)r \log n}{n} \cdot \text{inter-coherence factor},$$

$$\alpha_2 \leq \frac{1}{8}, \quad (5)$$

where the “inter-coherence factor” above (to be made precise later) is often not large and reflects the interaction between the coherence of prior subspaces and the true ones. Additionally,

$$\alpha_1 := \frac{w^4 \cos^2 \theta + \sin^2 \theta}{w^2 \cos^2 \theta + \sin^2 \theta}, \quad \alpha_2 := \frac{3\sqrt{1-w^2} \sin \theta}{\sqrt{w^2 \cos^2 \theta + \sin^2 \theta}}.$$

In particular, when $w = 1$, then the prior information $(\tilde{U}_r, \tilde{V}_r)$ is ignored and Program (4) reduces to Program (1). In this case, $\alpha_1 = 1, \alpha_2 = 0$ and (5) reduces to (2), save for the typically small coherence factor.⁶

On the other hand, when our prior knowledge is reliable, namely when θ is small, the proper choice of w in Program

⁴Principal angle between subspaces generalizes the notion of angle between lines and, for the sake of completeness, is defined later in Section VI.

⁵Throughout, we will occasionally use the standard “little- o ” and “big- O ” notation.

⁶In fact, this could be easily sharpened by slightly modifying the proof to yield $p \gtrsim \eta(M)r \log^2 n/n$, which precisely matches (2). We however opted for the more elegant (but slightly loose) bound presented in (5).

(4) leads to substantial improvement over Program (1). For example, for sufficiently small θ and with $w = \sqrt{\tan \theta}$, observe that $\alpha_1 = O(\sin \theta)$, $\alpha_2 = O(\sqrt{\sin \theta})$, and (5) now reads

$$1 \geq p \gtrsim \frac{\eta(M)r \log n}{n}, \quad (6)$$

save for the often small coherence factor. The lower bound in (6) is, by a logarithmic factor, better than (2).⁷

C. Matrix Recovery with Prior Knowledge

Matrix completion, discussed above, is a special case of the more general *matrix recovery*, in which the objective is to recover a matrix from more generic (and often random) linear measurements. This problem is reviewed in Section II, and Theorem 1 (in Section III) concerns leveraging prior information in this context.

D. Organization

The rest of this paper is organized as follows. In Section II, we briefly review low-rank matrix recovery and completion, and further motivate the use of prior knowledge in these contexts. The main results are summarized in Section III and followed by a thorough discussion. Section IV offers some numerical evidence to support the theory and the related literature is highlighted in Section V. Technical details are postponed to Sections VI-VIII and appendices. In particular, Section VI collects the technical tools common to the analysis of both matrix recovery and completion. Sections VII and VIII then contain the arguments specialized to matrix recovery and completion, respectively.

II. PROBLEM STATEMENT

Consider a matrix $M \in \mathbb{R}^{n \times n}$ and its SVD $M = U\Sigma V^*$. Here, $U, V \in \mathbb{R}^{n \times n}$ are orthonormal bases, and the diagonal matrix $\Sigma \in \mathbb{R}^{n \times n}$ collects the singular values of M in a non-increasing order, namely $\sigma_1(M) \geq \sigma_2(M) \geq \dots \geq \sigma_n(M)$. For any integer $r \leq n$, let $U_r, V_r \in \mathbb{R}^{n \times r}$ comprise of the first r columns of U, V , respectively, and let $\Sigma_r \in \mathbb{R}^{r \times r}$ contain the first r (largest) singular values of M . Ties are broken arbitrarily. Then, $M_r = U_r \Sigma_r V_r^* \in \mathbb{R}^{n \times n}$ is a rank- r truncation of M . We also set $M_{r+} = M - M_r$ to be the residual.

Suppose that we can only access $M \in \mathbb{R}^{n \times n}$ through a linear operator $\mathcal{R}_m(\cdot)$ that collects m measurements from M . More specifically, we let

$$y = \mathcal{R}_m(M + E) \in \mathbb{R}^m, \quad \|\mathcal{R}_m(E)\|_2 \leq e, \quad (7)$$

be the vector of m (possibly noisy) measurements. Here, $E \in \mathbb{R}^{n \times n}$ and $e \geq 0$ represent the noise. *Matrix recovery* is the problem of (approximately) reconstructing M from the measurement vector y .

⁷It is worth noting that, under different coherence assumptions, the very recent work [7] reduced the logarithmic factor in (2) from $\log^2 n$ to $\log n$. It remains to be seen if (2) could also be improved under the conventional coherence assumptions, in which case this improvement might, in turn, conceivably reflect in (6).

The case where the entries of M are randomly observed is of particular importance in practice, where here we pragmatically assume that a measurement operator $\mathcal{R}_p(\cdot)$ observes each entry of M with a probability of $p \in (0, 1]$.⁸ We set $p = m/n^2$ so that, in expectation, $\mathcal{R}_p(M)$ contains m entries of M . To be more specific, $\mathcal{R}_p(\cdot)$ takes $M \in \mathbb{R}^{n \times n}$ to $\mathcal{R}_p(M) \in \mathbb{R}^{n \times n}$ defined as

$$\mathcal{R}_p(M) = \sum_{i,j=1}^n \frac{\epsilon_{ij}}{p} \cdot M[i, j] \cdot C_{ij}, \quad (8)$$

where $\{\epsilon_{ij}\}$ is a sequence of independent Bernoulli random variables taking 1 with probability of p (and 0 otherwise). Throughout, $C_{ij} \in \mathbb{R}^{n \times n}$ is the $[i, j]$ th canonical matrix, so that $C_{ij}[i, j] = 1$ is its only nonzero entry. We also let

$$Y = \mathcal{R}_p(M + E) \in \mathbb{R}^{n \times n}, \quad \|\mathcal{R}_p(E)\|_F \leq e, \quad (9)$$

be the (possibly noisy) matrix of measurements. As before, E and e represent the noise. *Matrix completion* is the problem of (approximately) recovering M from Y .

A. Standard Low-Rank Matrix Recovery and Completion

In general, both matrix recovery and completion problems are ill-posed when $m \leq n^2$ and, to rectify this issue, it is common to impose that M is (nearly) low-rank. Let us briefly review both *low-rank* matrix completion and recovery next.

In low-rank matrix recovery, the *restricted isometry property* (RIP) plays a key role by ensuring that the measurement operator preserves the geometry of the set of low-rank matrices. More specifically, for $\delta_r \in [0, 1)$, we say that $\mathcal{R}_m(\cdot)$ satisfies the (r, δ_r) -RIP (or simply δ_r -RIP when there is no ambiguity) if

$$(1 - \delta_r) \|X\|_F \leq \|\mathcal{R}_m(X)\|_2 \leq (1 + \delta_r) \|X\|_F, \quad (10)$$

for every $X \in \mathbb{R}^{n \times n}$ with $\text{rank}(X) \leq r$. It is perhaps remarkable that, when the number of measurements m is sufficiently large, a “generic” linear operator from $\mathbb{R}^{n \times n}$ to \mathbb{R}^m satisfies the RIP. For example, suppose that $G \in \mathbb{R}^{n \times n}$ is populated with independent zero-mean Gaussian random variables with variance $1/m$. Then, $\langle G, X \rangle$ collects one linear measurement from X . The measurement operator formed from m independent copies of $\langle G, X \rangle$ is known to satisfy δ_r -RIP when $m \gtrsim rn \log n / \delta_r^2$. When the residual M_{r+} is small (i.e., $M = M_r + M_{r+}$ is nearly low-rank) and $\mathcal{R}_m(\cdot)$ satisfies the RIP, we can in fact (approximately) recover M by solving the following convex program:

$$\begin{cases} \min_X \|X\|_*, \\ \text{subject to } \|\mathcal{R}_m(X) - y\|_2 \leq e. \end{cases} \quad (11)$$

Above, with $\{\sigma_i(X)\}_i$ standing for the singular values of matrix X , $\|X\|_* = \sum_i \sigma_i(X)$ and $\|X\|_F = (\sum_i \sigma_i^2(X))^{\frac{1}{2}}$

⁸Alternatively, one may observe m entries of M at random. However, the advantage of the measurement model in (8) is that, by assigning different probabilities to each entry, the uniform sampling operator $\mathcal{R}_p(\cdot)$ may easily be modified for *leveraged sampling*, where each entry of M is observed according to its leverage (importance).

are the nuclear and Frobenius norms of X , respectively. The recovery error of Program (11) is summarized next [8], [9].

Proposition 1. (Matrix recovery) *For an integer $r \leq n$ and matrix $M \in \mathbb{R}^{n \times n}$, let $M_r \in \mathbb{R}^{n \times n}$ be a rank- r truncation of M , and let $M_{r+} = M - M_r$ be the residual. Suppose that the linear measurement operator $\mathcal{R}_m : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^m$ satisfies δ_{5r} -RIP with $\delta_{5r} \leq 0.1$. Let also $\widehat{M} \in \mathbb{R}^{n \times n}$ be a solution of Program (11). Then, it holds that*

$$\left\| \widehat{M} - M \right\|_F \lesssim \frac{\|M_{r+}\|_*}{\sqrt{r}} + e. \quad (12)$$

In low-rank matrix completion, on the other hand, $\mathcal{R}_p(\cdot)$ does *not* satisfy the RIP unless $p \approx 1$, in which case nearly every entry of M is observed anyway. Indeed, $\mathcal{R}_p(C_{11}) = 0$ with a probability of $1 - p$, where C_{11} is the first canonical matrix in $\mathbb{R}^{n \times n}$. However, $\mathcal{R}_p(\cdot)$ does preserve the geometry of the set of low-rank and *incoherent* matrices, provided that p is sufficiently large. More specifically, let $M_r = U_r \Sigma_r V_r^*$ be the SVD of the rank- r matrix M_r . Then, the *coherence* of M_r (denoted by $\eta(M_r)$ throughout) is defined as

$$\eta(M_r) = \frac{n}{r} \max \left[\|U_r\|_{2 \rightarrow \infty}^2, \|V_r\|_{2 \rightarrow \infty}^2 \right], \quad (13)$$

where $\|X\|_{2 \rightarrow \infty}$ returns the largest ℓ_2 norm of the rows of X . It is not difficult to verify that $\eta(M_r) \in [1, \frac{n}{r}]$, and that $\eta(M_r)$ depends only on the column and row spaces of M_r . When $\eta(M_r)$ is small, entries of M_r tend to be diffuse and we say that M_r is *incoherent*. At the other extreme, when $\eta(M_r)$ is large, M_r is often “spiky” and we say that M_r is *coherent*. Indeed, when M_{r+} is small (i.e., when M is nearly low-rank), we can (approximately) recover M by solving the convex program

$$\begin{cases} \min_X \|X\|_*, \\ \text{subject to } \|\mathcal{R}_p(X) - Y\|_F \leq e, \end{cases} \quad (14)$$

for which a bound for the recovery error is obtained by slightly modifying Theorem 7 in [3] to fit our setup [10, Proposition 2].

Proposition 2. (Matrix completion) *For an integer $r \leq n$ and matrix $M \in \mathbb{R}^{n \times n}$, let $M_r \in \mathbb{R}^{n \times n}$ be a rank- r truncation of M , and let $M_{r+} = M - M_r$ be the residual. Let $\widehat{M} \in \mathbb{R}^{n \times n}$ be a solution of Program (14). Then, except with a probability of at most $o(n^{-19})$, it holds that*

$$\left\| \widehat{M} - M \right\|_F \lesssim \frac{\|M_{r+}\|_*}{\sqrt{p}} + e\sqrt{pn}, \quad (15)$$

provided that

$$\frac{\eta(M_r) r \log^2 n}{n} \lesssim p \leq 1. \quad (16)$$

For instance, when M_r is incoherent, say $\eta_r(M_r) \approx 1$, solving Program (14) approximately completes M after observing only $O(rn \log^2 n)$ of its samples (in expectation). It is worth pointing out that, under different coherence assumptions on M , (15) has been recently improved to $O(rn \log n)$ samples [7].

B. Incorporating Prior Knowledge

Ideally, if the column and row spaces of a rank- r matrix M_r were known *a priori*, only r^2 linear measurements of M_r would suffice for exact recovery (in the absence of noise). Indeed, if rank- r matrices $A_r, B_r \in \mathbb{R}^{n \times r}$ span the column and row spaces of M_r , then $A_r^* M_r B_r \in \mathbb{R}^{r \times r}$ contains all the necessary information to reconstruct M_r .

More generally, consider $M \in \mathbb{R}^{n \times n}$ and let M_r be a rank- r truncation of M as before. If available, suppose that the r -dimensional subspaces \widetilde{U}_r and \widetilde{V}_r represent our prior knowledge about the column and row spaces of $M_r = U_r \Sigma_r V_r^*$. In order to incorporate this prior knowledge into matrix recovery and completion, we propose the following approach.

Let $P_{\widetilde{U}_r} \in \mathbb{R}^{n \times n}$ and $P_{\widetilde{U}_r^\perp} \in \mathbb{R}^{n \times n}$ be the orthogonal projections onto the subspace \widetilde{U}_r and its complement \widetilde{U}_r^\perp , respectively. Likewise, we define $n \times n$ projection matrices $P_{\widetilde{V}_r}$ and $P_{\widetilde{V}_r^\perp}$. For (left and right) weights $\lambda, \rho \in [0, 1]$, set

$$\begin{aligned} Q_{\widetilde{U}_r, \lambda} &:= \lambda \cdot P_{\widetilde{U}_r} + P_{\widetilde{U}_r^\perp} \in \mathbb{R}^{n \times n}, \\ Q_{\widetilde{V}_r, \rho} &:= \rho \cdot P_{\widetilde{V}_r} + P_{\widetilde{V}_r^\perp} \in \mathbb{R}^{n \times n}. \end{aligned} \quad (17)$$

In order to leverage the prior information $(\widetilde{U}_r, \widetilde{V}_r)$ in low-rank matrix recovery, we modify Program (11) as follows:

$$\begin{cases} \min_X \left\| Q_{\widetilde{U}_r, \lambda} \cdot X \cdot Q_{\widetilde{V}_r, \rho} \right\|_*, \\ \text{subject to } \|\mathcal{R}_m(X) - y\|_2 \leq e. \end{cases} \quad (18)$$

Note that the weights $\lambda, \rho \in [0, 1]$ reflect our uncertainty (or lack of confidence) in available prior knowledge, as the following examples might help clarify.

Example 1. *Consider the rank- r matrix $M_r = U_r \Sigma_r V_r^* \in \mathbb{R}^{n \times n}$ and suppose that $\widetilde{U}_r = U_r = \text{span}(M_r)$ and $\widetilde{V}_r = V_r = \text{span}(M_r^*)$, i.e., our prior information about M is perfectly accurate. To represent the lack of uncertainty in this knowledge, we set $\lambda = \rho = 0$ so that $Q_{\widetilde{U}_r} = P_{\widetilde{U}_r^\perp} = P_{\widetilde{U}_r}$ and $Q_{\widetilde{V}_r} = P_{\widetilde{V}_r^\perp}$, which in turn penalizes the component of solution orthogonal to the column and row spaces of M_r .*

Example 2. *At the other extreme, suppose that \widetilde{U}_r and \widetilde{V}_r are poor estimates of the true column and row spaces of $M_r = U_r \Sigma_r V_r^*$. To represent our uncertainty about the available prior information, we might set $\lambda = \rho = 1$, in which case Program (18) reduces to standard matrix recovery, namely Program (11).*

Similarly, for low-rank matrix completion, given the prior information $(\widetilde{U}_r, \widetilde{V}_r)$, we consider the following modification of Program (14):

$$\begin{cases} \min_X \left\| Q_{\widetilde{U}_r, \lambda} \cdot X \cdot Q_{\widetilde{V}_r, \rho} \right\|_*, \\ \text{subject to } \|\mathcal{R}_p(X) - Y\|_F \leq e. \end{cases} \quad (19)$$

To what extent, does prior knowledge help (or hurt) matrix recovery and completion? To find out, we next study the recovery error of Programs (18) and (19).

III. MAIN RESULTS

In Section II-B, we proposed Programs (18) and (19) in order to leverage available prior knowledge in matrix recovery and completion, respectively. Our first main result, proved in Section VII, is concerned with the performance of Program (18).

Theorem 1. (Matrix recovery with prior knowledge) *For an integer r and matrix $M \in \mathbb{R}^{n \times n}$, let $M_r \in \mathbb{R}^{n \times n}$ be a rank- r truncation of M , and let $M_{r+} = M - M_r$ be the residual. Let also $U_r = \text{span}(M_r)$ and $V_r = \text{span}(M_r^*)$ denote the column and row spaces of M_r , respectively. Additionally, let $\eta(M_r)$ be the coherence of M_r (see (13)). Suppose that the r -dimensional subspaces \tilde{U}_r and \tilde{V}_r represent the prior knowledge about U_r and V_r , respectively. Let*

$$u = \angle [U_r, \tilde{U}_r], \quad v = \angle [V_r, \tilde{V}_r],$$

denote the largest principal angles between each pair of subspaces.

For integer m , suppose that the linear sensing operator $\mathcal{R}_m(\cdot)$ satisfies δ_{32r} -RIP with

$$\delta_{32r} \leq \frac{0.9 - \max[\alpha_3, \alpha_4]/\sqrt{30}}{0.9 + \max[\alpha_3, \alpha_4]/\sqrt{30}}, \quad (20)$$

and acquire the (possibly noisy) measurement vector $y = \mathcal{R}_m(M + E) \in \mathbb{R}^m$ where $\|\mathcal{R}_m(E)\|_2 \leq e$. Lastly, for weights $\lambda, \rho \in (0, 1]$, let \hat{M} be a solution of Program (18). Then, it holds that

$$\|\hat{M} - M\|_F \lesssim \frac{\|M_{r+}\|_*}{\sqrt{r}} + e. \quad (21)$$

Above, α_3 and α_4 are set to be

$$\alpha_3 := \sqrt{\frac{\lambda^4 \cos^2 u + \sin^2 u}{\lambda^2 \cos^2 u + \sin^2 u}} + \sqrt{\frac{\rho^4 \cos^2 v + \sin^2 v}{\rho^2 \cos^2 v + \sin^2 v}},$$

$$\alpha_4 := \sqrt{\frac{2(1 - \lambda^2) \sin^2 u}{\lambda^2 \cos^2 u + \sin^2 u}} + \sqrt{\frac{2(1 - \rho^2) \sin^2 v}{\rho^2 \cos^2 v + \sin^2 v}}. \quad (22)$$

A few remarks are in order to help clarify Theorem 1.

Remark 1. (Connection to standard low-rank matrix recovery) If we set $\lambda = \rho = 1$, Program (18) reduces to Program (11) for standard low-rank matrix recovery, which entirely ignores the prior knowledge (\tilde{U}_r, \tilde{V}_r). In this case, $\alpha_3 = 2, \alpha_4 = 0$, and (20) reads $\delta_{32r} \leq 0.42$. It is known that $\delta_t \leq \frac{t-1}{s-1} \cdot \delta_s$ for $t \geq s > 1$ [11, Exercise 6.10]. Therefore, $\delta_{32r} \leq 7.75 \cdot \delta_{5r}$, so that $\delta_{5r} \leq 0.05$ implies $\delta_{32r} \leq 0.42$. This bound is slightly more conservative than $\delta_{5r} \leq 0.1$ in Proposition 1, as we made no attempts to optimize the constants.

On the other hand, even with the conservative bounds in Theorem 1, we observe that Program (18) indeed outperforms Program (11) when the prior knowledge is reliable (namely, the principal angles u, v are small) and when the weights λ, ρ are selected small (to reflect our confidence about the available information). For example, suppose that $u = v = \theta$ and $\lambda = \rho = \sqrt{\tan \theta}$ which gives $\alpha_3, \alpha_4 \leq 2\sqrt{2} \sin \theta$. Then, if we repeat the calculations at the end of Section

VII-B (to find the tightest bound here), we find that matrix recovery is successful (namely, (21) holds) if $\delta_{5r} \leq 0.75$ and $\theta \leq 0.0248$. This requirement is substantially better than $\delta_{5r} \leq 0.1$ in Proposition 1. That is, the bound for Program (18) considerably improves upon the bound for Program (11) and it does so by leveraging the available prior information.

Remark 2. (Different weights for column and row spaces)

Note that the formulation in Program (18) allows for assigning different weights to the column and row spaces (by selecting $\lambda \neq \rho$). This enables the user to handle scenarios when the uncertainty about \tilde{U}_r and \tilde{V}_r are different. For example, if one expects $u = \angle[U_r, \tilde{U}_r] \approx \frac{\pi}{2}$ and $v = \angle[V_r, \tilde{V}_r] \approx 0$, one might naturally choose $\lambda \approx 1$ and $\rho \approx 0$ to best handle this scenario.

Remark 3. (On choosing the weights) Ideally, the weights $\lambda, \rho \in (0, 1]$ must reflect our uncertainty (or lack of confidence) in the prior information (\tilde{U}_r, \tilde{V}_r). To loosen the restriction on the isometry constant for $\mathcal{R}_m(\cdot)$ (see (20)), inaccurate prior knowledge must be given lower influence in Program (19) and vice versa. As a concrete example, suppose that $u = v = \theta$ and $\lambda = \rho$. Then, if $\theta \approx \frac{\pi}{2}$ for example, the prior information is obviously unreliable, and it is wise to choose $\lambda = \rho \approx 1$ so as to give less influence to \tilde{U}_r and \tilde{V}_r in Program (19). On the contrary, if $\theta \approx 0$, the prior information is reliable, and it is best to take $\lambda = \rho \approx 0$ to reflect our confidence in the prior knowledge.

More specifically, given the principal angle $u = v = \theta$ (or its estimate), one might naturally ask: What is the optimal choice of weights $\lambda = \rho$ in Program (18)?

For a fixed angle $\theta \neq 0$, it is not difficult to verify that $\max[\alpha_3, \alpha_4]$ is minimized by the choice of $\lambda^2 = \rho^2 = \sqrt{\tan^4 \theta + \tan^2 \theta} - \tan^2 \theta$. This choice in turn maximizes the right hand side of (20). In particular, when the principal angle is small ($\theta \approx 0$) this suggests the choice of $\lambda = \rho \approx \sqrt{\tan \theta}$.

Our second main result in this paper, proved in Section VIII, quantifies the performance of Program (19) for low-rank matrix completion with prior knowledge.

Theorem 2. (Matrix completion with prior knowledge)

Recall the first paragraph of Theorem 1 and let $\eta(M_r) = \eta(U_r V_r^*)$ denote the coherence of M_r (see (13)). Additionally, let \tilde{U} and \tilde{V} be orthonormal bases for $\text{span}([U_r, \tilde{U}_r])$ and $\text{span}([V_r, \tilde{V}_r])$, respectively. For $p \in (0, 1]$ and recalling (8), acquire the (possibly noisy) measurement matrix $Y = \mathcal{R}_p(M + E)$ where $\|\mathcal{R}_p(E)\|_F \leq e$ for noise level $e \geq 0$. Lastly, for $\lambda, \rho \in (0, 1]$, let \hat{M} be a solution of Program (19). Then, it holds that

$$\|\hat{M} - M\|_F \lesssim \frac{\|M_{r+}\|_*}{\sqrt{p}} + e\sqrt{pn}, \quad (23)$$

except with a probability of $o(n^{-19})$, and provided that

$$1 \geq p \gtrsim \max[\log(\alpha_5 \cdot n), 1] \cdot \frac{\eta(M_r) r \log n}{n} \cdot \max \left[\alpha_6 \left(1 + \sqrt{\frac{\eta(\tilde{U}\tilde{V}^*)}{\eta(U_r V_r^*)}} \right), 1 \right],$$

$$\alpha_7 \leq \frac{1}{8}, \quad (24)$$

where $\eta(\check{U}\check{V}^*)$ is the coherence of $\check{U}\check{V}^*$. Above, we also set

$$\begin{aligned} \alpha_5 &:= \sqrt{\frac{\lambda^4 \cos^2 u + \sin^2 u}{\lambda^2 \cos^2 u + \sin^2 u}} \cdot \sqrt{\frac{\rho^4 \cos^2 v + \sin^2 v}{\rho^2 \cos^2 v + \sin^2 v}}, \\ \alpha_6 &:= \left(\sqrt{\frac{\lambda^2 \cos^2 u + \sin^2 u}{\rho^2 \cos^2 v + \sin^2 v}} + \sqrt{\frac{\rho^2 \cos^2 v + \sin^2 v}{\lambda^2 \cos^2 u + \sin^2 u}} \right) \\ &\quad \cdot \left(\sqrt{\lambda^4 \cos^2 u + \sin^2 u} + \sqrt{\rho^4 \cos^2 v + \sin^2 v} \right), \\ \alpha_7 &:= \frac{3\sqrt{1-\lambda^2} \sin u}{2\sqrt{\lambda^2 \cos^2 u + \sin^2 u}} + \frac{3\sqrt{1-\rho^2} \sin v}{2\sqrt{\rho^2 \cos^2 v + \sin^2 v}}. \end{aligned}$$

A few remarks are in order about Theorem 2.

Remark 4. (Connection to standard low-rank matrix completion) Note that, by taking $\lambda = \rho = 1$, Program (19) reduces to Program (14) for standard matrix completion, thereby ignoring any prior information. In this special case, $\alpha_5 = 1$, $\alpha_6 = 4$, $\alpha_7 = 0$, and (24) reads

$$1 \geq p \gtrsim \frac{\eta(M_r)r \log^2 n}{n} \cdot \left(1 + \sqrt{\frac{\eta(\check{U}\check{V}^*)}{\eta(U_r V_r^*)}} \right),$$

which is worse than (16) in Proposition 2 because of the term $\eta(\check{U}\check{V}^*)/\eta(U_r V_r^*)$. However, employing a slightly sharper bound in Appendix E gives $p \gtrsim \eta r \log^2 n/n$, which precisely matches (16). We opted for the looser bound in (24) to keep the bound compact.

As was the case in matrix recovery, Program (19) improves over Program (14) when the prior knowledge is reliable and our confidence is reflected in the small choice of weights. For example, suppose again that $u = v = \theta$ and $\lambda = \rho = \sqrt{\tan \theta}$. Then, a simple calculation shows that

$$\alpha_5 = \frac{2 \sin \theta}{\sin \theta + \cos \theta} \leq 2 \sin \theta.$$

Therefore, if $\sin \theta \leq 1/n$, the logarithmic factors in (24) reduce to merely $\log n$. If also $\eta(\check{U}\check{V}^*) \approx \eta(U_r V_r^*)$, then the lower bound on sampling probability in (24) improves over that in standard matrix completion, where prior knowledge is not utilized, by reducing $\log^2 n$ to $\log n$.

Note also that, in the extreme case of $\theta = 0$ (i.e., when row and column spaces are exactly known *a priori*), one might recover M from only r^2 samples (or, equivalently, $p = r^2/n^2$).

Remark 5. (On choosing the weights) This remark, similar to Remark 3, discusses the optimal choice of weights $\lambda, \rho \in (0, 1]$ now in Program (19). For simplicity, again assume that $u = v = \theta$ and $\lambda = \rho$. Then, α_6 and α_7 are non-increasing and non-decreasing in $\lambda = \rho$, respectively. However, α_5 is minimized with the choice of $\lambda^2 = \rho^2 = \sqrt{\tan^4 \theta + \tan^2 \theta} - \tan^2 \theta$. In particular, when θ is small, α_5 is minimized with the choice of $\lambda = \rho \approx \sqrt{\tan \theta}$.

IV. SIMULATIONS

This section is intended to provide some numerical evidence to support the weighted matrix recovery and completion schemes introduced earlier.

Starting with matrix recovery, we assume in Program (18) for simplicity that there is no measurement noise ($e = 0$) and the left and right weights are equal ($\lambda = \rho = w$). Fix $n = 20, r = 3$, and set $M = U_r \Sigma_r V_r^*$ where $U_r \in \mathbb{R}^{n \times r}$ with orthonormal columns span the column space of a random Gaussian matrix $G \in \mathbb{R}^{n \times r}$. Also, $\Sigma_r \in \mathbb{R}^{r \times r}$ is diagonal with nonzero entries drawn from the uniform distribution on the interval $(0, 1]$. Lastly, $V_r \in \mathbb{R}^{n \times r}$ is similarly generated from an independent copy of G .

First, consider the scenario when the prior knowledge about the column and row spaces of M is reliable. More specifically, set \tilde{U}_r to be the column span of $U_r + \sigma G'$, where $\sigma = 0.01$ and $G' \in \mathbb{R}^{n \times r}$ is an independent copy of G . Construct \tilde{V}_r in a similar manner. We measure M by computing its inner product against m independent random Gaussian matrices, namely $y[i] = \langle M, G_i \rangle$ for $i \in [1 : m]$. As the number of measurements m increases and for various values of weight $\lambda = \rho = w$, we solve Program (18) (with $e = 0$). The success rates, averaged over 100 trials, are shown in Figure 2a. A trial is considered successful if M is recovered up to a relative error of 10^{-3} . Note that, with a proper choice of $w < 1$, we correctly place our confidence in the prior knowledge, which in turn results in better matrix recovery. On the other hand, by choosing $w = 1$, Program (18) fails to take advantage of the reliable prior information about M .

Next, consider the case when the prior knowledge is unreliable by setting $\sigma = 1$ above and repeating the same experiment. The results are illustrated in Figure 2b. Note that, by choosing w to be small, we misplace our confidence in the prior knowledge, and that leads to a poor performance. As the figure suggests, it is wise in fact to take w to be large to reflect the uncertainty and ignore the unreliable prior information.

The same experiments are repeated for matrix completion. Instead of m , the probability p of observing every entry of M varies in $(0, 1]$ and, instead of Program (18), we solve Program (19) (with $e = 0$ and $\lambda = \rho = w$). As with matrix recovery, the results in the case of matrix completion, summarized in Figure 3, corroborate the theoretical findings in Section III.

V. RELATED WORK

Programs similar to (18) and (19) have appeared in the literature before, and we wish to summarize here some of the related work. In [12], [13], the authors incorporate side information for matrix completion using nuclear norm minimization. However, their works differ from ours in that they assume perfect subspace information, which they use to reduce the dimension of the low-rank recovery problem (as well as the sample complexity). A later paper [14] supplements this recovery program with a correction term to account for imperfect side information. Theory is again provided which allows for a reduction in sample complexity; however, this theory is limited to randomly generated matrices and uses

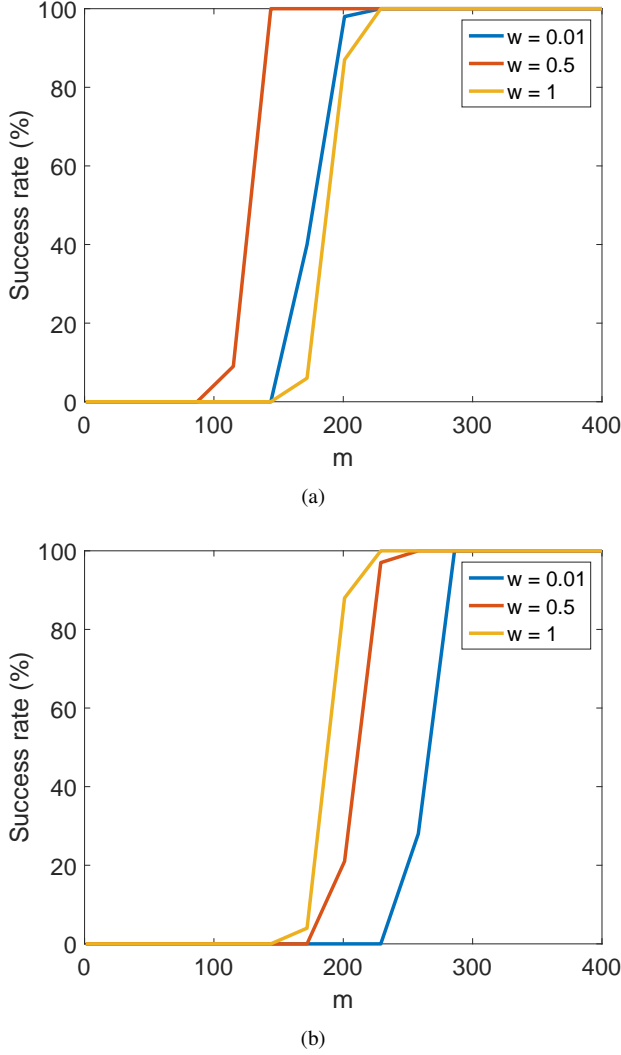


Figure 2: Matrix recovery with good and poor prior knowledge (top and bottom figures, respectively) and for various weights w in Program (18). Refer to Section IV for details.

a different characterization of subspace accuracy than the principal angles we consider.

In [15], the authors considered the following non-convex program for rank minimization:

$$\begin{cases} \min_X \log(\det(X)) \\ X \in \mathcal{C}. \end{cases} \quad (25)$$

Here, the feasible set $\mathcal{C} \subset \mathbb{R}^{n \times n}$ is assumed to be convex. (Local) linearization of the above objective function leads to a *majorization-minimization* algorithm to solve Program (25), in which the k -th iteration takes the form of

$$\begin{cases} \min_X \|W_1^k X W_2^k\|_* \\ X \in \mathcal{C}, \end{cases} \quad (26)$$

for certain weight matrices W_1^k and W_2^k . Convergence of this *reweighted* algorithm to a local minimum of Program (25) is known. See also [16] for a related problem.

In [5], the authors study the following program:

$$\min_{X \in \mathbb{R}^{n \times n}} \frac{1}{m} \|P_\Omega(X - Y)\|_F^2 + \lambda_N \|AXB\|_*. \quad (27)$$

Here, $\Omega \subseteq [1 : n]^2$ is a random index set of size m , and $P_\Omega(X) \in \mathbb{R}^{n \times n}$ retains the entries of X on the index set Ω and sets the rest to zero. In addition, $Y = P_\Omega(M + E)$ where we take $M \in \mathbb{R}^{n \times n}$ to be rank- r for simplicity, and the entries of $E \in \mathbb{R}^{n \times n}$ are independent zero-mean Gaussian random variables with variance σ^2/n . Lastly, $\lambda_N > 0$ and $A, B \in \mathbb{R}^{n \times n}$ are both assumed to be invertible. Let $\widehat{M} \in \mathbb{R}^{n \times n}$ be a solution of Program (27). Then, Theorem 2 in the same reference then establishes that

$$\|\widehat{M} - M\|_F^2 \lesssim \alpha_N^2 \max[1, \sigma^2] \frac{r \log n}{m}, \quad (28)$$

with high probability and provided that $\lambda_N \gtrsim \sqrt{\frac{n \log n}{m}}$. Above,

$$\alpha_N := n \cdot \frac{\|AMB^*\|_\infty}{\|AMB^*\|_F}. \quad (29)$$

By setting $A = Q_{\widetilde{U}_{r,\lambda}}$ and $B = Q_{\widetilde{V}_{r,\rho}}$ (see (17)), Program (27) will be equivalent to Program (19) for the right choice of λ_N . To compare (28) and Theorem 2, let $Q_{\widetilde{U}_{r,\lambda}} M Q_{\widetilde{V}_{r,\rho}}^* = U_{N,r} \Sigma_{N,r} V_{N,r}^*$ be the SVD of AMB and note that

$$\begin{aligned} \alpha_N^2 &= n^2 \cdot \frac{\|Q_{\widetilde{U}_{r,\lambda}} M Q_{\widetilde{V}_{r,\rho}}^*\|_\infty^2}{\|Q_{\widetilde{U}_{r,\lambda}} M Q_{\widetilde{V}_{r,\rho}}^*\|_F^2} \\ &= n^2 \cdot \frac{\|U_{N,r} \Sigma_{N,r} V_{N,r}^*\|_\infty^2}{\|U_{N,r} \Sigma_{N,r} V_{N,r}^*\|_F^2} \\ &\leq n^2 \cdot \frac{\|U_{N,r}\|_{2 \rightarrow \infty}^2 \cdot \|\Sigma_{N,r}\|^2 \cdot \|V_{N,r}\|_{2 \rightarrow \infty}^2}{\|\Sigma_{N,r}\|_F^2} \\ &= n^2 \cdot \frac{r^2 \cdot \eta^2(Q_{\widetilde{U}_{r,\lambda}} M Q_{\widetilde{V}_{r,\rho}}^*) \|\Sigma_{N,r}\|^2}{n^2 \|\Sigma_{N,r}\|_F^2} \\ &\leq \frac{r \cdot \eta^2(Q_{\widetilde{U}_{r,\lambda}} M Q_{\widetilde{V}_{r,\rho}}^*) \|\Sigma_{N,r}\|^2}{\sigma_r^2(\Sigma_{N,r})} \\ &= r \cdot \eta^2(Q_{\widetilde{U}_{r,\lambda}} M Q_{\widetilde{V}_{r,\rho}}^*) \kappa^2(Q_{\widetilde{U}_{r,\lambda}} M Q_{\widetilde{V}_{r,\rho}}^*), \end{aligned} \quad (30)$$

where $\eta(\cdot)$ and $\kappa(\cdot)$ return the coherence and condition number of a matrix, respectively. In particular, the above inequalities hold with equality when $U_{N,r}, V_{N,r}$ are columns of the Fourier basis and the condition number of $Q_{\widetilde{U}_{r,\lambda}} M Q_{\widetilde{V}_{r,\rho}}^*$ equals one. Since coherence and condition number are both never smaller than one, we conclude that the right-hand side of (28) scales with r^2 . This, in turn, forces m (number of measurements) to scale with r^2 . In contrast, the expected number of measurements required in Theorem 2 scales linearly with r . We must note that [5] itself was preceded by [17] where, among other contributions, a weighted program for matrix completion was studied with diagonal A and B in Program (27). A similar program was empirically studied in [18] in the context of collaborative filtering.

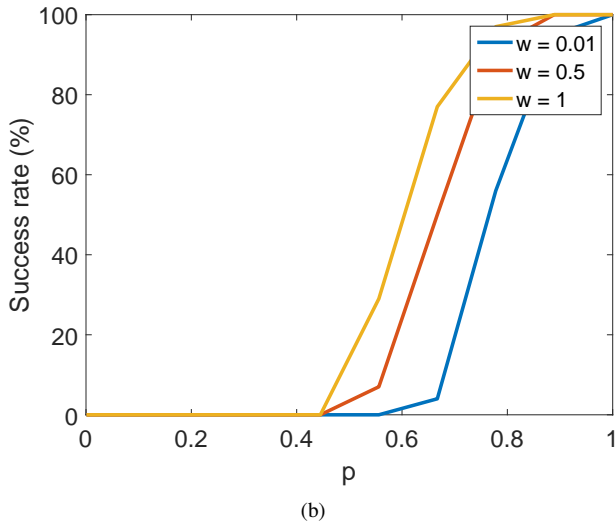
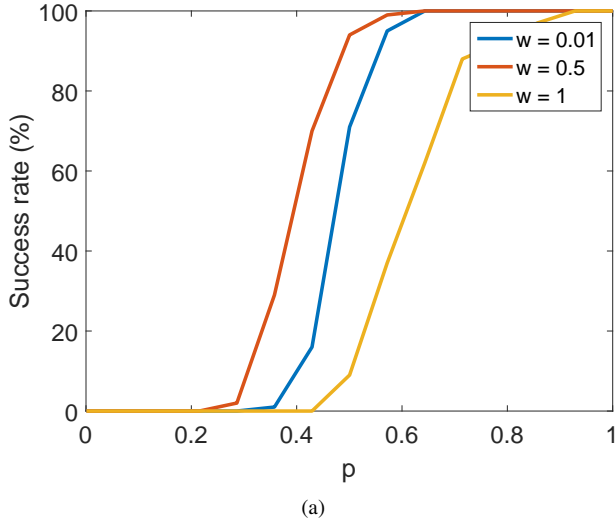


Figure 3: Matrix completion with good and poor prior knowledge (top and bottom figures, respectively) and for various weights w in Program (19). Refer to Section IV for details.

It must be noted that our interest in weighted matrix recovery was inspired by [6], which we have previously mentioned in Section 1.1. In [6], weighted nuclear norm minimization has been successfully but heuristically applied to a “frequency continuation” matrix completion problem in seismic signal processing.

The ideas in this work have a precedent in *compressive sensing* and, more generally, sparse regression. In [19], [20], weighted ℓ_1 minimization was proposed in order to incorporate partial information about the support; this was actually the inspiration for the weighted matrix completion algorithm in [6]. Originally, iterative reweighted ℓ_1 minimization to enhance sparse recovery appeared in [21]. As our analysis in Sections VI-VIII indicates, extending these ideas to matrices requires different techniques and is substantially more involved.

ACKNOWLEDGEMENTS

The authors took inspiration from [6], and MBW is grateful to Felix Herrmann and the Seismic Laboratory for Imaging and Modeling (SLIM) for their hospitality during part of his sabbatical. Part of this research was conducted when AE was a graduate fellow at the Statistical and Applied Mathematical Sciences Institute (SAMSI) and later a visitor at the Institute for Computational and Experimental Research in Mathematics (ICERM). AE is grateful for their hospitality and kindness, and also acknowledges Rachel Ward for helpful conversations about a closely related problem.

VI. COMMONS

In this section, we collect the necessary technical tools that are common to the analysis of both Programs (18) and (19).

A. Canonical Decomposition

Central to the analysis is a canonical way of decomposing M_r that takes the prior knowledge $(\tilde{U}_r, \tilde{V}_r)$ into account. This result is well-known and a short proof is given in Appendix A for the sake of completeness [22]. Throughout, the empty blocks of matrices should be interpreted as filled with zeros. Also, in our notation, $I_a \in \mathbb{R}^{a \times a}$ is the identity matrix and $0_a \in \mathbb{R}^{a \times a}$ and $0_{a \times b} \in \mathbb{R}^{a \times b}$ are filled with zeros.

Lemma 3. Consider a rank- r matrix $M_r \in \mathbb{R}^{n \times n}$, and let $U_r = \text{span}(M_r)$ be the column span of M_r . Let \tilde{U}_r be another r -dimensional subspace in \mathbb{R}^n . Then, there exists $U_r, \tilde{U}_r \in \mathbb{R}^{n \times r}$, $U'_r, \tilde{U}'_r \in \mathbb{R}^{n \times r}$, and $U''_{n-2r} \in \mathbb{R}^{n \times (n-2r)}$ such that

$$U_r = \text{span}(U_r), \quad \tilde{U}_r = \text{span}(\tilde{U}_r),$$

and

$$\begin{aligned} B_L &:= \begin{bmatrix} U_r & U'_r & U''_{n-2r} \end{bmatrix} \in \mathbb{R}^{n \times n}, \\ \tilde{B}_L &:= \begin{bmatrix} \tilde{U}_r & \tilde{U}'_r & U''_{n-2r} \end{bmatrix} \in \mathbb{R}^{n \times n}, \end{aligned} \quad (31)$$

are both orthonormal bases for \mathbb{R}^n . Moreover, it holds that

$$B_L^* \tilde{B}_L = \begin{bmatrix} \cos u & \sin u & & \\ -\sin u & \cos u & & \\ & & & I_{n-2r} \end{bmatrix}, \quad (32)$$

where $u \in \mathbb{R}^{r \times r}$ is diagonal and contains the principal angles between U_r and \tilde{U}_r , in a non-increasing order: $\pi/2 \geq u_1 \geq u_2 \geq \dots \geq u_r \geq 0$.⁹ The diagonal matrix $\cos u$ is naturally defined as

$$\cos u := \begin{bmatrix} \cos u_1 & & & \\ & \cos u_2 & & \\ & & \ddots & \\ & & & \cos u_r \end{bmatrix} \in \mathbb{R}^{r \times r},$$

⁹Note that, prior to Section VI, we had used u (rather than u_1) to denote the largest principal angle, in order to keep the notation light. From now on, we will adhere to setup of Lemma 3.

and $\sin u \in \mathbb{R}^{r \times r}$ is defined likewise. A similar construction exists for $V_r = \text{span}(M_r^*)$ and \tilde{V}_r , where we form the orthonormal bases $B_R, \tilde{B}_R \in \mathbb{R}^{n \times n}$ such that

$$B_R^* \tilde{B}_R = \begin{bmatrix} \cos v & \sin v & & \\ -\sin v & \cos v & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix}. \quad (33)$$

As before, the diagonal of $v \in \mathbb{R}^{r \times r}$ contains the principal angles between V_r and \tilde{V}_r in non-decreasing order.

Lemma 3 immediately implies that

$$\tilde{U}_r = B_L \begin{bmatrix} \cos u & & \\ -\sin u & & \\ 0_{(n-2r) \times r} & & \end{bmatrix},$$

which, in turn, allows us to derive the following expressions for orthogonal projections onto the subspace \tilde{U}_r and its complement:

$$\begin{aligned} P_{\tilde{U}_r} &= \tilde{U}_r \tilde{U}_r^* \\ &= B_L \begin{bmatrix} \cos^2 u & -\sin u \cdot \cos u & & \\ -\sin u \cdot \cos u & \sin^2 u & & \\ & & 0_{n-2r} & \\ & & & \end{bmatrix} B_L^*, \end{aligned}$$

$$\begin{aligned} P_{\tilde{U}_r^\perp} &= I_n - P_{\tilde{U}_r} \\ &= B_L \begin{bmatrix} \sin^2 u & \sin u \cdot \cos u & & \\ \sin u \cdot \cos u & \cos^2 u & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} B_L^*. \end{aligned}$$

It also follows that

$$\begin{aligned} Q_{\tilde{U}_r, \lambda} &= \lambda \cdot P_{\tilde{U}_r} + P_{\tilde{U}_r^\perp} \\ &= B_L \begin{bmatrix} \lambda \cos^2 u + \sin^2 u & (1-\lambda) \sin u \cdot \cos u & & \\ (1-\lambda) \sin u \cdot \cos u & \lambda \sin^2 u + \cos^2 u & & \\ & & 0_{(n-r) \times r} & \\ & & & \end{bmatrix} B_L^*, \end{aligned} \quad (34)$$

where we used (17). We next mold the above expression for $Q_{\tilde{U}_r, \lambda}$ into one that involves an upper-triangular matrix, as this will prove useful shortly. Define the orthonormal basis $O_L \in \mathbb{R}^{n \times n}$ as

$$O_L := \begin{bmatrix} (\lambda \cos^2 u + \sin^2 u) \Delta_L^{-1} & & & \\ (1-\lambda) \sin u \cdot \cos u \cdot \Delta_L^{-1} & & & \\ 0_{(n-r) \times r} & & & \\ -(1-\lambda) \sin u \cdot \cos u \cdot \Delta_L^{-1} & 0_{r \times (n-r)} & & \\ (\lambda \cos^2 u + \sin^2 u) \Delta_L^{-1} & 0_{r \times (n-r)} & & \\ 0_{(n-r) \times r} & & I_{n-2r} & \end{bmatrix},$$

$$\Delta_L := \sqrt{\lambda^2 \cos^2 u + \sin^2 u} \in \mathbb{R}^{r \times r}, \quad (35)$$

where Δ_L is invertible because $\lambda > 0$, by assumption. (It is easily verified that indeed $O_L O_L^* = I_n$.) We then rewrite (34) as

$$\begin{aligned} Q_{\tilde{U}_r, \lambda} &= B_L (O_L O_L^*) \\ &\cdot \begin{bmatrix} \lambda \cos^2 u + \sin^2 u & (1-\lambda) \sin u \cdot \cos u & & \\ (1-\lambda) \sin u \cdot \cos u & \lambda \sin^2 u + \cos^2 u & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} B_L^* \\ &= B_L O_L \begin{bmatrix} \Delta_L & (1-\lambda^2) \sin u \cdot \cos u \cdot \Delta_L^{-1} & & \\ & \lambda \Delta_L^{-1} & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} B_L^* \\ &=: B_L O_L \begin{bmatrix} L_{11} & L_{12} & & \\ & L_{22} & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} B_L^* \\ &=: B_L O_L L B_L^*, \end{aligned} \quad (36)$$

where $L \in \mathbb{R}^{n \times n}$ is an upper-triangular matrix with blocks $L_{11}, L_{12}, L_{22} \in \mathbb{R}^{r \times r}$ and defined as

$$\begin{aligned} L &:= \begin{bmatrix} L_{11} & L_{12} & & \\ & L_{22} & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} \\ &= \begin{bmatrix} \Delta_L & (1-\lambda^2) \sin u \cdot \cos u \cdot \Delta_L^{-1} & & \\ & \lambda \Delta_L^{-1} & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix}. \end{aligned} \quad (37)$$

In the third line of (36), we used the fact that $O_L O_L^* = I_n$. Because B_L, O_L are both orthonormal bases, we record that

$$\|Q_{\tilde{U}_r, \lambda}\| = \|L\| = 1. \quad (\text{see (17) and (36)}) \quad (38)$$

We can perform the same calculations for the row spaces and, in particular, define $R \in \mathbb{R}^{n \times n}$ as

$$\begin{aligned} R &:= \begin{bmatrix} R_{11} & R_{12} & & \\ & R_{22} & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} \\ &= \begin{bmatrix} \Delta_R & (1-\rho^2) \sin v \cdot \cos v \cdot \Delta_R^{-1} & & \\ & \rho \Delta_R^{-1} & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix}, \end{aligned} \quad (39)$$

with $\Delta_R = \sqrt{\rho^2 \cos^2 v + \sin^2 v} \in \mathbb{R}^{n \times n}$. With these calculations in mind, for an arbitrary matrix $H \in \mathbb{R}^{n \times n}$, we find the crucial decomposition

$$\begin{aligned} Q_{\tilde{U}_r, \lambda} \cdot H \cdot Q_{\tilde{V}_r, \rho} &= B_L O_L L (B_L^* H B_R) R^* O_R^* B_R^* \quad (\text{(36) and } Q_{\tilde{V}_r, \rho} = Q_{\tilde{V}_r, \rho}^*) \\ &=: B_L O_L L \bar{H} R^* O_R^* B_R^* \quad (\bar{H} := B_L^* H B_R) \\ &=: B_L O_L L \begin{bmatrix} \bar{H}_{11} & \bar{H}_{12} & \bar{H}_{13} \\ \bar{H}_{21} & \bar{H}_{22} & \bar{H}_{23} \\ \bar{H}_{31} & \bar{H}_{32} & \bar{H}_{33} \end{bmatrix} R^* O_R^* B_R^*, \end{aligned} \quad (40)$$

with $\bar{H}_{11}, \bar{H}_{22} \in \mathbb{R}^{r \times r}$ and $\bar{H}_{33} \in \mathbb{R}^{(n-2r) \times (n-2r)}$ being the diagonal blocks of \bar{H} . Moreover, from Lemma 3, recall that

$$\text{span}(M_r) = \text{span}(U_r), \quad \text{span}(M_r^*) = \text{span}(V_r), \quad (41)$$

which allows us to record that

$$\begin{aligned}
& Q_{\tilde{U}_r, \lambda} \cdot M_r \cdot Q_{\tilde{V}_r, \rho} \\
&= B_L O_L L (B_L^* M_r B_R) R^* O_R^* B_R^* \quad (\text{see (36)}) \\
&=: B_L O_L L \overline{M}_r R^* O_R^* B_R^* \quad (\overline{M}_r := B_L^* M_r B_R) \\
&=: B_L O_L L \begin{bmatrix} \overline{M}_{r,11} & \\ & 0_{n-r} \end{bmatrix} R^* O_R^* B_R^* \quad (\text{see (41)}) \\
&= B_L O_R \begin{bmatrix} L_{11} \overline{M}_{r,11} R_{11} & \\ & 0_{n-r} \end{bmatrix} \\
&\quad \cdot O_R^* B_R^*. \quad ((37) \text{ and } R_{11} = R_{11}^*) \quad (42)
\end{aligned}$$

In the last line above, we benefited from the fact that, by construction, both L and R are upper-triangular matrices. Note also that $\overline{M}_{r,11} = U_r^* M_r V_r$, as defined above, is *not* necessarily diagonal. For future reference, the following useful inequalities are proved in Appendix B.

Lemma 4. *With L and its blocks L_{11}, L_{12}, L_{22} defined in (37), it holds that*

$$\begin{aligned}
\|L_{11}\| &= \|\Delta_L\| \leq \sqrt{\lambda^2 \cos^2 u_1 + \sin^2 u_1}, \\
\|L_{12}\| &\leq \frac{(1 - \lambda^2) \sin u_1}{\sqrt{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \leq 1 - \lambda^2, \\
\|I_r - L_{22}\| &\leq \frac{\sqrt{1 - \lambda^2} \sin u_1}{\sqrt{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \leq \sqrt{1 - \lambda^2}, \\
\| [L_{11} \quad L_{12}] \|^2 &\leq \frac{\lambda^4 \cos^2 u_1 + \sin^2 u_1}{\lambda^2 \cos^2 u_1 + \sin^2 u_1} \leq 1, \\
\| [L_{22} - I_r \quad L_{12}] \|^2 &\leq \frac{2(1 - \lambda^2) \sin^2 u_1}{\lambda^2 \cos^2 u_1 + \sin^2 u_1}, \quad (43)
\end{aligned}$$

where u_1 is the largest principal angle between r -dimensional subspaces U_r and \tilde{U}_r . Similar bounds hold for R and its blocks, R_{11}, R_{12}, R_{22} .

B. Support

Let M_r be a rank- r truncation of $M \in \mathbb{R}^{n \times n}$ (obtained via SVD) and consider the decomposition

$$M = M_r + M_{r+} = U_r \overline{M}_{r,11} V_r^* + M_{r+},$$

where $U_r, V_r \in \mathbb{R}^{n \times r}$ (with orthonormal columns) span column and row spaces of M_r , and $\overline{M}_{11} \in \mathbb{R}^{r \times r}$ is rank- r but not necessary diagonal. Let $U_r = \text{span}(U_r) = \text{span}(M_r)$ and $V_r = \text{span}(V_r) = \text{span}(M_r^*)$. Then, the *support* of $M_r \in \mathbb{R}^{n \times n}$ is the linear subspace $\mathbb{T} \subset \mathbb{R}^{n \times n}$ defined as

$$\mathbb{T} = \{ Z \in \mathbb{R}^{n \times n} : Z = P_{U_r} \cdot Z + Z \cdot P_{V_r} - P_{U_r} \cdot Z \cdot P_{V_r} \}, \quad (44)$$

where $P_{U_r}, P_{V_r} \in \mathbb{R}^{n \times n}$ are orthogonal projection onto U_r, V_r , respectively. For the record, the orthogonal projection onto \mathbb{T} and its complement \mathbb{T}^\perp take $Z \in \mathbb{R}^{n \times n}$ to

$$\begin{aligned}
\mathcal{P}_{\mathbb{T}}(Z) &= P_{U_r} \cdot Z + Z \cdot P_{V_r} - P_{U_r} \cdot Z \cdot P_{V_r}, \\
\mathcal{P}_{\mathbb{T}^\perp}(Z) &= P_{U_r^\perp} \cdot Z \cdot P_{V_r^\perp},
\end{aligned}$$

respectively. As suggested above, throughout we reserve the calligraphic font for matrix operators. Note that, using Lemma 3, we can express \mathbb{T} equivalently as

$$\begin{aligned}
\mathbb{T} &= \left\{ Z \in \mathbb{R}^{n \times n} : Z = B_L \overline{Z} B_R^*, \quad \overline{Z} = \begin{bmatrix} \overline{Z}_{11} & \overline{Z}_{12} \\ \overline{Z}_{21} & 0_{n-r} \end{bmatrix} \right\} \\
&=: B_L \cdot \overline{\mathbb{T}} \cdot B_R^*, \quad (45)
\end{aligned}$$

where, to be clear, the new subspace $\overline{\mathbb{T}} \subset \mathbb{R}^{n \times n}$ is the support of $\overline{M}_r := B_L^* M_r B_R$ and is defined as

$$\overline{\mathbb{T}} = \left\{ \overline{Z} \in \mathbb{R}^{n \times n} : \overline{Z} = \begin{bmatrix} \overline{Z}_{11} & \overline{Z}_{12} \\ \overline{Z}_{21} & 0_{n-r} \end{bmatrix} \right\}. \quad (46)$$

Also note that, for arbitrary

$$\overline{Z} = \begin{bmatrix} \overline{Z}_{11} & \overline{Z}_{12} \\ \overline{Z}_{21} & \overline{Z}_{22} \end{bmatrix} \in \mathbb{R}^{n \times n},$$

with $\overline{Z}_{11} \in \mathbb{R}^{r \times r}, \overline{Z}_{22} \in \mathbb{R}^{(n-r) \times (n-r)}$, the orthogonal projection onto $\overline{\mathbb{T}}$ and its complement simply take \overline{Z} to

$$\begin{aligned}
\mathcal{P}_{\overline{\mathbb{T}}}(\overline{Z}) &= \begin{bmatrix} \overline{Z}_{11} & \overline{Z}_{12} \\ \overline{Z}_{21} & 0_{n-r} \end{bmatrix}, \\
\mathcal{P}_{\overline{\mathbb{T}}^\perp}(\overline{Z}) &= \begin{bmatrix} 0_r & \\ & \overline{Z}_{22} \end{bmatrix}, \quad (47)
\end{aligned}$$

respectively. Lastly, we record the following connection: For arbitrary $Z \in \mathbb{R}^{n \times n}$ and with $\overline{Z} = B_L^* Z B_R$, we have that

$$\begin{aligned}
\mathcal{P}_{\mathbb{T}}(Z) &= B_L \cdot \mathcal{P}_{\overline{\mathbb{T}}}(\overline{Z}) \cdot B_R^*, \\
\mathcal{P}_{\mathbb{T}^\perp}(Z) &= B_L \cdot \mathcal{P}_{\overline{\mathbb{T}}^\perp}(\overline{Z}) \cdot B_R^*. \quad (48)
\end{aligned}$$

VII. ANALYSIS FOR MATRIX RECOVERY

In this section, we study Program (18) in detail and eventually prove Theorem 1. First, in Section VII-A, we establish a variant of the well-known *nullspace property* for Program (18) which loosely states that the recovery error in Program (18) is concentrated along the subspace \mathbb{T} , namely the support of M_r . Using this property, then, we complete the proof of Theorem 1 in Section VII-B.

A. Nullspace Property

For solution \widehat{M} , let $H := \widehat{M} - M$ be the error. By feasibility of M and optimality of $\widehat{M} = M + H$ in Program (18), we have that

$$\| Q_{\tilde{U}_r, \lambda} (M + H) Q_{\tilde{V}_r, \rho} \|_* \leq \| Q_{\tilde{U}_r, \lambda} M Q_{\tilde{V}_r, \rho} \|_*. \quad (49)$$

Recall that $M = M_r + M_{r+}$. With the decomposition of M_r in (42) at hand, the right-hand side above is then bounded follows:

$$\begin{aligned}
& \left\| Q_{\tilde{U},\lambda} M Q_{\tilde{V},\rho} \right\|_* \\
& \leq \left\| Q_{\tilde{U},\lambda} M_r Q_{\tilde{V},\rho} \right\|_* + \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_* \\
& = \left\| B_L O_L \begin{bmatrix} L_{11} \bar{M}_{r,11} R_{11} & & \\ & 0_{n-r} & \\ & & \end{bmatrix} O_R^* B_R^* \right\|_* \\
& \quad + \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_* \quad (\text{see (42)}) \\
& = \left\| \begin{bmatrix} L_{11} \bar{M}_{r,11} R_{11} & & \\ & 0_{n-r} & \\ & & \end{bmatrix} \right\|_* + \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_* \\
& = \left\| L \bar{M} R^* \right\|_* + \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_*, \quad (\text{see (42)}) \quad (50)
\end{aligned}$$

where the first inequality above uses $M = M_r + M_{r+}$ and the triangle inequality. The second identity uses the rotational invariance of the nuclear norm. In the last line, we used the fact that B_L, B_R, O_L, O_R are all orthonormal bases. Using the decomposition of H in (40), the left hand side of (92) can also be bounded as follows:

$$\begin{aligned}
& \left\| Q_{\tilde{U},\lambda} (M + H) Q_{\tilde{V},\rho} \right\|_* \\
& \geq \left\| Q_{\tilde{U},\lambda} (M_r + H) Q_{\tilde{V},\rho} \right\|_* - \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_* \\
& = \left\| B_L O_L L (\bar{M}_r + \bar{H}) R^* O_R^* B_R^* \right\|_* - \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_* \\
& = \left\| L \bar{M}_r R^* + L \bar{H} R^* \right\|_* - \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_* \\
& = \left\| L \bar{M}_r R^* + L \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) R^* + L \mathcal{P}_{\bar{\mathbb{T}}^\perp}(\bar{H}) R^* \right\|_* \\
& \quad - \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_* \\
& = \left\| L \bar{M}_r R^* + L \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) R^* + L \mathcal{P}_{\bar{\mathbb{T}}^\perp}(\bar{H}) R^* \right. \\
& \quad \left. - \mathcal{P}_{\bar{\mathbb{T}}^\perp}(\bar{H}) + \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) \right\|_* - \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_*.
\end{aligned}$$

Above, the second line uses $M = M_r + M_{r+}$ and the triangle inequality. The third line uses (40) and (42). The fourth line uses the rotational invariance of the nuclear norm. We continue by writing that

$$\begin{aligned}
& \left\| Q_{\tilde{U},\lambda} (M + H) Q_{\tilde{V},\rho} \right\|_* \\
& = \left\| L \bar{M}_r R^* + L \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) R^* + L \begin{bmatrix} 0_r & & \\ & \bar{H}_{22} & \bar{H}_{23} \\ & \bar{H}_{32} & \bar{H}_{33} \end{bmatrix} R^* \right. \\
& \quad \left. - \begin{bmatrix} 0_r & & \\ & \bar{H}_{22} & \bar{H}_{23} \\ & \bar{H}_{32} & \bar{H}_{33} \end{bmatrix} + \mathcal{P}_{\bar{\mathbb{T}}^\perp}(\bar{H}) \right\|_* \\
& \quad - \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_* \quad (\text{see (47)}) \\
& =: \left\| L \bar{M}_r R^* + L \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) R^* + L \bar{H}' R^* - \bar{H}' + \mathcal{P}_{\bar{\mathbb{T}}^\perp}(\bar{H}) \right\|_* \\
& \quad - \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_* \quad (\text{see (51)}) \\
& \geq \left\| L \bar{M}_r R^* + \mathcal{P}_{\bar{\mathbb{T}}^\perp}(\bar{H}) \right\|_* - \left\| L \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) R^* \right\|_* \\
& \quad - \left\| L \bar{H}' R^* - \bar{H}' \right\|_* - \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_*,
\end{aligned}$$

where the last line uses the triangle inequality. To be concrete, above we defined

$$\bar{H}' := \begin{bmatrix} 0_r & & \\ & \bar{H}_{22} & \bar{H}_{23} \\ & \bar{H}_{32} & 0_{n-2r} \end{bmatrix} \in \mathbb{R}^{n \times n}. \quad (51)$$

Consequently,

$$\begin{aligned}
& \left\| Q_{\tilde{U},\lambda} (M + H) Q_{\tilde{V},\rho} \right\|_* \\
& = \left\| \begin{bmatrix} L_{11} \bar{M}_{r,11} R_{11} & & \\ & \bar{H}_{22} & \bar{H}_{23} \\ & \bar{H}_{32} & \bar{H}_{33} \end{bmatrix} \right\|_* - \left\| L \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) R^* \right\|_* \\
& \quad - \left\| L \bar{H}' R^* - \bar{H}' \right\|_* - \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_* \\
& = \left\| \begin{bmatrix} L_{11} \bar{M}_{r,11} R_{11} & & \\ & 0_r & \\ & & 0_{n-2r} \end{bmatrix} \right\|_* \\
& \quad + \left\| \begin{bmatrix} 0_r & & \\ & \bar{H}_{22} & \bar{H}_{23} \\ & \bar{H}_{32} & \bar{H}_{33} \end{bmatrix} \right\|_* - \left\| L \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) R^* \right\|_* \\
& \quad - \left\| L \bar{H}' R^* - \bar{H}' \right\|_* - \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_* \\
& = \left\| L \bar{M} R^* \right\|_* + \left\| \mathcal{P}_{\bar{\mathbb{T}}^\perp}(\bar{H}) \right\|_* - \left\| L \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) R^* \right\|_* \\
& \quad - \left\| L \bar{H}' R^* - \bar{H}' \right\|_* - \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_*. \quad (52)
\end{aligned}$$

Above, we used (42) and (47) twice. In the second identity in (52), we used the fact that $\|A + B\|_* = \|A\|_* + \|B\|_*$ whenever both column and row spaces of A are orthogonal to those of B , namely when $A^* B = A B^* = 0$. Now combining (49) with the bounds in (50) and (52) yields that

$$\left\| \mathcal{P}_{\bar{\mathbb{T}}^\perp}(\bar{H}) \right\|_* \leq \left\| L \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) R^* \right\|_* + \left\| L \bar{H}' R^* - \bar{H}' \right\|_* + 2 \left\| Q_{\tilde{U},\lambda} M_{r+} Q_{\tilde{V},\rho} \right\|_*. \quad (53)$$

We next simplify the terms in the above inequality. First, notice that

$$\begin{aligned}
& \begin{bmatrix} 0_r & & \\ & L_{22} & \\ & & I_{n-2r} \end{bmatrix} \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) \begin{bmatrix} 0_r & & \\ & R_{22} & \\ & & I_{n-2r} \end{bmatrix} \\
& = \begin{bmatrix} 0_r & & \\ & L_{22} & \\ & & I_{n-2r} \end{bmatrix} \begin{bmatrix} \bar{H}_{11} & \bar{H}_{12} & \bar{H}_{13} \\ \bar{H}_{21} & 0_r & \\ \bar{H}_{31} & & 0_{n-2r} \end{bmatrix} \\
& \quad \cdot \begin{bmatrix} 0_r & & \\ & R_{22} & \\ & & I_{n-2r} \end{bmatrix} \quad (\text{see (47)}) \\
& = 0_n, \quad (54)
\end{aligned}$$

which, in turn, allows us to simplify the first norm on the right-hand side of (53) as follows:

$$\begin{aligned}
& \left\| L \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) R^* \right\|_* \\
& = \left\| \begin{bmatrix} L_{11} & L_{12} & \\ & L_{22} & \\ & & I_{n-2r} \end{bmatrix} \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) \begin{bmatrix} R_{11} & & \\ R_{12}^* & R_{22} & \\ & & I_{n-2r} \end{bmatrix} \right\|_* \\
& = \left\| \begin{bmatrix} 0_r & & \\ & L_{22} & \\ & & I_{n-2r} \end{bmatrix} \mathcal{P}_{\bar{\mathbb{T}}}(\bar{H}) \begin{bmatrix} 0_r & & \\ & R_{22} & \\ & & I_{n-2r} \end{bmatrix} \right\|_*,
\end{aligned}$$

where we used (54). Then, we continue by writing that

$$\begin{aligned} & \|L\mathcal{P}_{\overline{T}}(\overline{H})R^*\|_* \\ &= \left\| \begin{bmatrix} L_{11} & L_{12} & & \\ & 0_r & & \\ & & 0_{n-2r} & \\ & & & \end{bmatrix} \mathcal{P}_{\overline{T}}(\overline{H})R \right. \\ &+ \left. \begin{bmatrix} 0_r & & & \\ & L_{22} & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} \mathcal{P}_{\overline{T}}(\overline{H}) \begin{bmatrix} R_{11} & & & \\ R_{12}^* & 0_r & & \\ & & 0_{n-2r} & \\ & & & \end{bmatrix} \right\|_* , \end{aligned}$$

in which we applied (56). Consequently,

$$\begin{aligned} & \|L\mathcal{P}_{\overline{T}}(\overline{H})R^*\|_* \\ &\leq \left\| \begin{bmatrix} L_{11} & L_{12} & & \\ & 0_r & & \\ & & 0_{n-2r} & \\ & & & \end{bmatrix} \mathcal{P}_{\overline{T}}(\overline{H})R \right\|_* \\ &+ \left\| \begin{bmatrix} 0_r & & & \\ & L_{22} & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} \mathcal{P}_{\overline{T}}(\overline{H}) \begin{bmatrix} R_{11} & & & \\ R_{12}^* & 0_r & & \\ & & 0_{n-2r} & \\ & & & \end{bmatrix} \right\|_* \\ &\leq \left\| \begin{bmatrix} L_{11} & L_{12} \\ & \end{bmatrix} \right\| \|\mathcal{P}_{\overline{T}}(\overline{H})\|_* \|R\| \\ &\quad + \max\{\|L_{22}\|, 1\} \|\mathcal{P}_{\overline{T}}(\overline{H})\|_* \left\| \begin{bmatrix} R_{11} & R_{12} \end{bmatrix} \right\| \\ &\leq \left\| \begin{bmatrix} L_{11} & L_{12} \\ & \end{bmatrix} \right\| \|\mathcal{P}_{\overline{T}}(\overline{H})\|_* \\ &\quad + \|\mathcal{P}_{\overline{T}}(\overline{H})\|_* \left\| \begin{bmatrix} R_{11} & R_{12} \end{bmatrix} \right\| \\ &= \left(\left\| \begin{bmatrix} L_{11} & L_{12} \\ & \end{bmatrix} \right\| + \left\| \begin{bmatrix} R_{11} & R_{12} \end{bmatrix} \right\| \right) \|\mathcal{P}_{\overline{T}}(\overline{H})\|_* . \end{aligned} \quad (55)$$

The second inequality above uses the fact that $\|AB\|_* \leq \|A\| \cdot \|B\|_*$ for all conforming matrices A, B . The last inequality uses (37), (38), and the observation that $\|L_{22}\| \leq \|L\|$. In the second inequality, we also used the so-called polarization identity

$$AZC - BZD = (A - B)ZC + BZ(C - D), \quad (56)$$

for conforming matrices A, B, C, D, Z . The second norm on the right-hand side of (53) may also be bounded as follows:

$$\begin{aligned} & \|L\overline{H}'R^* - \overline{H}'\|_* \\ &= \left\| \begin{bmatrix} L_{11} & L_{12} & & \\ & L_{22} & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} \overline{H}' \begin{bmatrix} R_{11} & & & \\ R_{12}^* & R_{22} & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} \right. \\ &\quad - \left. \begin{bmatrix} L_{11} & & & \\ & I_r & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} \overline{H}' \begin{bmatrix} R_{11} & & & \\ & I_r & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} \right\|_* \\ &= \left\| \begin{bmatrix} 0_r & L_{12} & & \\ & L_{22} - I_r & & \\ & & 0_{n-2r} & \\ & & & \end{bmatrix} \overline{H}'R^* \right. \\ &\quad - \left. \begin{bmatrix} L_{11} & & & \\ & I_r & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} \overline{H}' \begin{bmatrix} 0_r & & & \\ R_{12}^* & R_{22} - I_r & & \\ & & 0_{n-2r} & \\ & & & \end{bmatrix} \right\|_* , \end{aligned}$$

Above, the first identity uses (37). and (51). The second identity employs (56). We continue by applying the triangle

inequality to find that

$$\begin{aligned} & \|L\overline{H}'R^* - \overline{H}'\|_* \\ &\leq \left\| \begin{bmatrix} 0_r & L_{12} & & \\ & L_{22} - I_r & & \\ & & 0_{n-2r} & \\ & & & \end{bmatrix} \overline{H}'R^* \right\|_* \\ &+ \left\| \begin{bmatrix} L_{11} & & & \\ & I_r & & \\ & & I_{n-2r} & \\ & & & \end{bmatrix} \overline{H}' \begin{bmatrix} 0_r & & & \\ R_{12}^* & R_{22} - I_r & & \\ & & 0_{n-2r} & \\ & & & \end{bmatrix} \right\|_* , \end{aligned}$$

and, consequently,

$$\begin{aligned} & \|L\overline{H}'R^* - \overline{H}'\|_* \\ &\leq \left\| \begin{bmatrix} L_{12}^* & L_{22} - I_r \end{bmatrix} \right\| \|\overline{H}'\|_* \|R\| \\ &\quad + \max\{\|L_{11}\|, 1\} \|\overline{H}'\|_* \left\| \begin{bmatrix} R_{12}^* & R_{22} - I_r \end{bmatrix} \right\| \\ &\leq \left\| \begin{bmatrix} L_{12}^* & L_{22} - I_r \end{bmatrix} \right\| \|\overline{H}'\|_* \\ &\quad + \|\overline{H}'\|_* \left\| \begin{bmatrix} R_{12}^* & R_{22} - I_r \end{bmatrix} \right\| \\ &= \left(\left\| \begin{bmatrix} L_{12} & L_{22} - I_r \end{bmatrix} \right\| + \left\| \begin{bmatrix} R_{12} & R_{22} - I_r \end{bmatrix} \right\| \right) \|\overline{H}'\|_* . \end{aligned} \quad (57)$$

The first inequality above uses $\|AB\|_* \leq \|A\| \cdot \|B\|_*$. The second inequality applies (37), (38), and the fact that $\|L_{11}\| \leq \|L\|$. The last line benefits from the fact that $L_{12} = L_{12}^*$. By substituting (55) and (57) back into (53), we arrive at the following inequality:

$$\begin{aligned} & \|\mathcal{P}_{\overline{T}^\perp}(\overline{H})\|_* \\ &\leq \left(\left\| \begin{bmatrix} L_{11} & L_{12} \end{bmatrix} \right\| + \left\| \begin{bmatrix} R_{11} & R_{12} \end{bmatrix} \right\| \right) \|\mathcal{P}_{\overline{T}}(\overline{H})\|_* \\ &\quad + \left(\left\| \begin{bmatrix} L_{12} & L_{22} - I_r \end{bmatrix} \right\| + \left\| \begin{bmatrix} R_{12} & R_{22} - I_r \end{bmatrix} \right\| \right) \|\overline{H}'\|_* \\ &\quad + 2 \left\| Q_{\tilde{U}_{r,\lambda}} M_r + Q_{\tilde{V}_{r,\rho}} \right\|_* \\ &\leq \left(\sqrt{\frac{\lambda^4 \cos^2 u_1 + \sin^2 u_1}{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} + \sqrt{\frac{\rho^4 \cos^2 v_1 + \sin^2 v_1}{\rho^2 \cos^2 v_1 + \sin^2 v_1}} \right) \\ &\quad \cdot \|\mathcal{P}_{\overline{T}}(\overline{H})\|_* + \left(\sqrt{\frac{2(1-\lambda^2)\sin^2 u_1}{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \right. \\ &\quad \left. + \sqrt{\frac{2(1-\rho^2)\sin^2 v_1}{\rho^2 \cos^2 v_1 + \sin^2 v_1}} \right) \cdot \|\overline{H}'\|_* + 2 \left\| Q_{\tilde{U}_{r,\lambda}} M_r + Q_{\tilde{V}_{r,\rho}} \right\|_* \\ &=: \alpha_3(u_1, v_1, \lambda, \rho) \|\mathcal{P}_{\overline{T}}(\overline{H})\|_* + \alpha_4(u_1, v_1, \lambda, \rho) \|\overline{H}'\|_* \\ &\quad + 2 \left\| Q_{\tilde{U}_{r,\lambda}} M_r + Q_{\tilde{V}_{r,\rho}} \right\|_* . \end{aligned} \quad (58)$$

In the second inequality above, Lemma 4 is applied. Some additional manipulation of (58) is in order. First, owing to (48) and the rotational invariance of the nuclear norm, it holds that

$$\|\mathcal{P}_{\overline{T}}(H)\|_* = \|\mathcal{P}_{\overline{T}}(\overline{H})\|_* ,$$

$$\|\mathcal{P}_{\overline{T}^\perp}(H)\|_* = \|\mathcal{P}_{\overline{T}^\perp}(\overline{H})\|_* . \quad (59)$$

If we also define the linear subspace $\widetilde{\mathbb{T}} \subset \mathbb{T}^\perp \subset \mathbb{R}^{n \times n}$ as

$$\widetilde{\mathbb{T}} := \left\{ Z \in \mathbb{R}^{n \times n} : Z = B_L \begin{bmatrix} 0_r & & \\ & \overline{Z}_{22} & \overline{Z}_{23} \\ & \overline{Z}_{32} & 0_{n-2r} \end{bmatrix} B_R^* \right\}, \quad (60)$$

then we may write that

$$\|\overline{H}'\|_* = \|B_L \overline{H}' B_R^*\|_* = \|\mathcal{P}_{\widetilde{\mathbb{T}}}(H)\|_*, \quad (61)$$

by rotational invariance of the nuclear norm and in light of (51). Putting these all together, we may rewrite (58) as

$$\begin{aligned} \|\mathcal{P}_{\mathbb{T}^\perp}(H)\|_* &\leq \alpha_3 \|\mathcal{P}_{\mathbb{T}}(H)\|_* + \alpha_4 \|\mathcal{P}_{\widetilde{\mathbb{T}}}(H)\|_* \\ &\quad + 2 \left\| Q_{\widetilde{U}_{r,\lambda}} M_{r+} Q_{\widetilde{V}_{r,\rho}} \right\|_* \quad (\text{see (59) and (61)}) \\ &\leq \alpha_3 \|\mathcal{P}_{\mathbb{T}}(H)\|_* + \alpha_4 \|\mathcal{P}_{\widetilde{\mathbb{T}}}(H)\|_* + 2 \|M_{r+}\|_*, \end{aligned} \quad (62)$$

with α_3, α_4 as defined in (58). The last line above uses the inequality $\|AB\|_* \leq \|A\| \cdot \|B\|_*$ and (38). Note that (62) might be interpreted as an analog of the nullspace property in standard matrix recovery [8]. In particular, suppose that $M = M_r$ is rank- r and $\lambda = \rho = 1$ so that Program (18) reduces to Program (11). Then, in turn, (62) reduces to $\|\mathcal{P}_{\mathbb{T}^\perp}(H)\|_* \leq 2\|\mathcal{P}_{\mathbb{T}}(H)\|_*$ which is by a factor of two worse than the standard nullspace property.¹⁰ With (62) at hand, we are now prepared to prove Theorem 1.

B. Body of the Analysis

Let $\mathcal{P}_{\mathbb{T}^\perp}(H) = U' \Sigma' (V')^*$ be the SVD of $\mathcal{P}_{\mathbb{T}^\perp}(H)$ with $U', V' \in \mathbb{R}^{n \times (n-r)}$ and where the diagonal matrix $\Sigma' \in \mathbb{R}^{(n-r) \times (n-r)}$ contains the singular values of $\mathcal{P}_{\mathbb{T}^\perp}(H)$, in a non-increasing order. We partition the singular values into groups of size r' as follows, with integer r' to be set later. Using MATLAB's matrix notation, we form

$$\Sigma_i = \Sigma'[(i-1)r' + 1 : ir', (i-1)r' + 1 : ir'] \in \mathbb{R}^{r' \times r'},$$

$$U_i := U'[:, (i-1)r' + 1 : ir'] \in \mathbb{R}^{n \times r'},$$

$$V_i := V'[:, (i-1)r' + 1 : ir'] \in \mathbb{R}^{n \times r'},$$

$$H_i := U_i' \Sigma_i (V_i')^* \in \mathbb{R}^{n \times n},$$

for $i \geq 1$.¹¹ To unburden the notation, we also set $H_0 = \mathcal{P}_{\mathbb{T}}(H)$. This setup allows us to decompose the error H as

$$H = \mathcal{P}_{\mathbb{T}}(H) + \mathcal{P}_{\mathbb{T}^\perp}(H) = \sum_{i \geq 0} H_i. \quad (63)$$

Note that both row and column spans of H_i and H_j are orthogonal to one another when $i \neq j$, namely

$$H_i^* H_j = H_i H_j^* = 0_n, \quad i \neq j. \quad (64)$$

¹⁰The extra factor of two is likely an artifact of using the polarization identity.

¹¹The four last blocks might be smaller than others.

On the other hand, by feasibility of both M and \widehat{M} in Program (18), we find the so-called tube constraint:

$$\begin{aligned} \|\mathcal{R}_m(H)\|_F &= \left\| \mathcal{R}_m(\widehat{M} - M) \right\|_F \\ &\leq \left\| \mathcal{R}_m(\widehat{M}) - y \right\|_* + \|\mathcal{R}_m(M) - y\|_* \\ &\leq 2e. \quad (\text{see Program (18)}) \end{aligned} \quad (65)$$

Using (63), (65), and the triangle inequality, we may then write that

$$\|\mathcal{R}_m(H_0 + H_1)\|_F \leq \sum_{i \geq 2} \|\mathcal{R}_m(H_i)\|_F + 2e. \quad (66)$$

Recall (10) and suppose that the measurement operator $\mathcal{R}_m(\cdot)$ satisfies $\delta_{r''}$ -RIP with integer $r'' \geq 2r + r'$ to be set later. By construction,

$$\text{rank}(H_0 + H_1) = \text{rank}(\mathcal{P}_{\mathbb{T}}(H) + H_1) \leq 2r + r' \leq r'',$$

$$\text{rank}(H_i) \leq r' \leq r'', \quad i \geq 1,$$

and, therefore, (66) and the RIP together imply that

$$\begin{aligned} (1 - \delta_{r''}) \|H_0 + H_1\|_F &\leq (1 + \delta_{r''}) \sum_{i \geq 2} \|H_i\|_F + 2e \\ &\leq \frac{1 + \delta_{r''}}{\sqrt{r'}} \sum_{i \geq 1} \|H_i\|_* + 2e \\ &= \frac{1 + \delta_{r''}}{\sqrt{r'}} \left\| \sum_{i \geq 1} H_i \right\|_* + 2e \\ &= \frac{1 + \delta_{r''}}{\sqrt{r'}} \|\mathcal{P}_{\mathbb{T}^\perp}(H)\|_* + 2e, \end{aligned} \quad (67)$$

where, in the second line, we used the fact that $\|H_{i+1}\|_F \leq \frac{1}{\sqrt{r'}} \|H_i\|_*$ for every $i \geq 1$, which itself follows directly from the non-increasing order of the singular values in Σ' and the fact that $\text{rank}(H_i) \leq r'$ for every $i \geq 1$. The last line uses the fact that $\|A + B\|_* = \|A\|_* + \|B\|_*$ when $\text{span}(A) \perp \text{span}(B)$ and $\text{span}(A^*) \perp \text{span}(B^*)$. Then, invoking the nullspace property (62) below, we find that

$$\begin{aligned} \|H_0 + H_1\|_F &\leq \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{1}{r'}} \|\mathcal{P}_{\mathbb{T}^\perp}(H)\|_* + \frac{2e}{1 - \delta_{r''}} \\ &\leq \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{1}{r'}} (\alpha_3 \|\mathcal{P}_{\mathbb{T}}(H)\|_* + \alpha_4 \|\mathcal{P}_{\widetilde{\mathbb{T}}}(H)\|_* \\ &\quad + 2 \|M_{r+}\|_*) + \frac{2e}{1 - \delta_{r''}}, \end{aligned}$$

where the first and second inequalities use (67) and (62), respectively. We now continue by writing that

$$\begin{aligned}
& \|H_0 + H_1\|_F \\
& \leq \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{1}{r'}} (\alpha_3 \|\mathcal{P}_T(H)\|_* + \alpha_4 \|\mathcal{P}_{\tilde{T}}(H)\|_* \\
& \quad + 2 \|M_{r^+}\|_*) + \frac{2e}{1 - \delta_{r''}} \quad (\text{see (62)}) \\
& \leq \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{2r}{r'}} \cdot (\alpha_3 \|\mathcal{P}_T(H)\|_F + \alpha_4 \|\mathcal{P}_{\tilde{T}}(H)\|_F) \\
& \quad + 2 \cdot \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{1}{r'}} \|M_{r^+}\|_* + \frac{2e}{1 - \delta_{r''}} \\
& = \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{2r}{r'}} \cdot (\alpha_3 \|H_0\|_F + \alpha_4 \|\mathcal{P}_{\tilde{T}}(H)\|_F) \\
& \quad + 2 \cdot \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{1}{r'}} \|M_{r^+}\|_* + \frac{2e}{1 - \delta_{r''}} \\
& \leq \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{2r}{r'}} \cdot (\alpha_3 \|H_0\|_F + \alpha_4 \|H_1\|_F) \\
& \quad + 2 \cdot \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{1}{r'}} \|M_{r^+}\|_* + \frac{2e}{1 - \delta_{r''}} \\
& \leq \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{2r}{r'}} \cdot \max[\alpha_3, \alpha_4] \cdot (\|H_0\|_F + \|H_1\|_F) \\
& \quad + 2 \cdot \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{1}{r'}} \|M_{r^+}\|_* + \frac{2e}{1 - \delta_{r''}} \\
& \leq \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{2r}{r'}} \cdot \max[\alpha_3, \alpha_4] \cdot \sqrt{2 \|H_0\|_F^2 + 2 \|H_1\|_F^2} \\
& \quad + 2 \cdot \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{1}{r'}} \|M_{r^+}\|_* + \frac{2e}{1 - \delta_{r''}} \\
& = 2 \cdot \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{r}{r'}} \cdot \max[\alpha_3, \alpha_4] \cdot \|H_0 + H_1\|_F \\
& \quad + 2 \cdot \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{1}{r'}} \|M_{r^+}\|_* + \frac{2e}{1 - \delta_{r''}}, \quad (68)
\end{aligned}$$

which, as justified presently, holds as long as $r' \geq 2r$. Above, the second inequality holds because, by (44), $\text{rank}(\mathcal{P}_T(H)) \leq 2r$ and, by (60), $\text{rank}(\mathcal{P}_{\tilde{T}}(H)) \leq 2r$. The first identity there uses the fact that $H_0 = \mathcal{P}_T(H)$. The third inequality above follows because $\tilde{T}, T_1 \subset T^\perp$ and H_1 , by construction, is a rank- r' truncation of $\mathcal{P}_{T^\perp}(H)$. Therefore, as long as $r' \geq 2r \geq \text{rank}(\mathcal{P}_{\tilde{T}}(H))$, we have that $\|\mathcal{P}_{\tilde{T}}(H)\|_F \leq \|H_1\|_F$, as claimed above. Also, the last inequality uses the fact that $a + b \leq \sqrt{2a^2 + 2b^2}$ for scalars a, b , and the last identity holds because $H_0 = \mathcal{P}_T(H)$, $\text{span}(H_1) \subseteq T^\perp$, and consequently $\langle H_0, H_1 \rangle = \text{trace}(H_0^* H_1) = 0$. After rearranging the terms in (68), we find that if

$$2 \cdot \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \cdot \sqrt{\frac{r}{r'}} \max[\alpha_3, \alpha_4] \leq 0.9, \quad (69)$$

or equivalently if

$$\delta_{r''} \leq \frac{0.9 - 2 \max[\alpha_3, \alpha_4] \sqrt{\frac{r}{r'}}}{0.9 + 2 \max[\alpha_3, \alpha_4] \sqrt{\frac{r}{r'}}}, \quad (70)$$

then the following holds:

$$\|H_0 + H_1\|_F \leq 20 \cdot \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{1}{r'}} \|M_{r^+}\|_* + \frac{20e}{1 - \delta_{r''}}. \quad (71)$$

On the other hand, note that

$$\begin{aligned}
& \left\| \sum_{i \geq 2} H_i \right\|_F \\
& \leq \sum_{i \geq 2} \|H_i\|_F \quad (\text{triangle inequality}) \\
& \leq \frac{\|\mathcal{P}_{T^\perp}(H)\|_*}{\sqrt{r'}} \quad (\text{implicit in (67)}) \\
& \leq 2 \sqrt{\frac{r}{r'}} \cdot \max[\alpha_3, \alpha_4] \cdot \|H_0 + H_1\|_F \\
& \quad + \frac{2}{\sqrt{r'}} \|M_{r^+}\|_* \quad (\text{implicit in (68)}) \\
& \leq 2 \sqrt{\frac{r}{r'}} \cdot \max[\alpha_3, \alpha_4] \cdot \left(20 \cdot \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{1}{r'}} \|M_{r^+}\|_* \right. \\
& \quad \left. + \frac{20e}{1 - \delta_{r''}} \right) + \frac{2}{\sqrt{r'}} \|M_{r^+}\|_*, \quad (72)
\end{aligned}$$

where the last inequality uses (71). Lastly, (71) and (72) together imply that

$$\begin{aligned}
& \left\| \widehat{M} - M \right\|_F \\
& = \|H\|_F \\
& \leq \|H_0 + H_1\|_F + \left\| \sum_{i \geq 2} H_i \right\|_F \quad (\text{see (63)}) \\
& \leq \left(1 + 2 \sqrt{\frac{r}{r'}} \cdot \max[\alpha_3, \alpha_4] \right) \\
& \quad \cdot \left(20 \cdot \frac{1 + \delta_{r''}}{1 - \delta_{r''}} \sqrt{\frac{1}{r'}} \|M_{r^+}\|_* + \frac{20e}{1 - \delta_{r''}} \right) + \frac{2}{\sqrt{r'}} \|M_{r^+}\|_*,
\end{aligned}$$

provided that $r'' \geq 2r + r'$ and as long as (70) is met. The last inequality above uses (72). This completes the proof of Theorem 1 after taking $r' = 30r$ and $r'' = 32r$.

VIII. ANALYSIS FOR MATRIX COMPLETION

In this section, we analyze Program (19) and eventually prove Theorem 2. In fact, we prove a stronger result based on *leveraged sampling*, of which Theorem 2 is a special case. Let us begin with recalling the definition of *leverage scores* of a matrix.

A. Leverage Scores

For a rank- r matrix $M_r \in \mathbb{R}^{n \times n}$, let $U_r = \text{span}(U_r) = \text{span}(M_r)$ and $V_r = \text{span}(V_r) = \text{span}(M_r^*)$ be the column and row spaces of M_r with orthonormal bases $U_r, V_r \in \mathbb{R}^{n \times r}$, respectively. The leverage score corresponding to the i th row of M_r is defined as

$$\mu_i = \mu_i(U_r) := \frac{n}{r} \|U_r[i, :]\|_2^2, \quad i \in [1 : n], \quad (73)$$

where $U_r[i, :]$ is the i th row of U_r . Similarly, the leverage score corresponding to the j th column of M_r is defined as

$$\nu_j = \nu_j(V_r) := \frac{n}{r} \|V_r[:, j]\|_2^2, \quad j \in [1 : n]. \quad (74)$$

As our notation above suggests, leverage scores of a subspace are indeed independent of the choice of the orthonormal basis for subspace. In particular, notice that the coherence of a matrix is simply the largest leverage score of its column and row spans (see (13)), namely

$$\eta(M_r) = \max_i \mu_i(U_r) \vee \max_j \mu_j(V_r), \quad (75)$$

where $a \vee b = \max[a, b]$ is the shorthand for maximum. We also assign leverage scores to subspaces $\check{U} = \text{span}([U_r, \tilde{U}_r])$ and $\check{V} = \text{span}([V_r, \tilde{V}_r])$:

$$\begin{aligned} \check{\mu}_i &= \mu_i(\check{U}_r), & i &\in [1 : n], \\ \check{\nu}_j &= \nu_j(\check{V}_r), & j &\in [1 : n]. \end{aligned} \quad (76)$$

Throughout, we will continue using $\{\mu_i, \nu_i\}$ and $\{\check{\mu}_i, \check{\nu}_i\}$ as shorthand to ease the notation. To facilitate the calculations later, let us define the $n \times n$ diagonal matrix

$$\mu = \begin{bmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_n \end{bmatrix}. \quad (77)$$

The $n \times n$ matrices $\nu, \check{\mu}, \check{\nu}$ are defined similarly using $\{\nu_i, \check{\mu}_i, \check{\nu}_i\}$, respectively. Recall the $n \times r$ matrices U_r, U'_r, V_r, V'_r constructed in Lemma 3 and denote with $\|A\|_{2 \rightarrow \infty}$ the largest ℓ_2 norm of the rows of a matrix A . Assuming that $\mu_i, \nu_i \neq 0$ for all i , the relations below (which follow directly from earlier definitions) will prove useful later on:

$$\begin{aligned} \left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} U_r \right\|_{2 \rightarrow \infty} &= 1, & \left\| \left(\frac{\nu r}{n} \right)^{-\frac{1}{2}} V_r \right\|_{2 \rightarrow \infty} &= 1, \\ \left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} U'_r \right\|_{2 \rightarrow \infty} &\leq \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}}, \\ \left\| \left(\frac{\nu r}{n} \right)^{-\frac{1}{2}} V'_r \right\|_{2 \rightarrow \infty} &\leq \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}}. \end{aligned} \quad (78)$$

To be complete, let us verify the third relation above. Letting \check{U} denote an orthonormal basis for \check{U} , we write that

$$\begin{aligned} &\left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} U'_r \right\|_{2 \rightarrow \infty} \\ &= \max_i \frac{\|U'_r[i, :]\|_2}{\|U_r[i, :]\|_2} \quad (\text{see (77) and then (73)}) \\ &\leq \max_i \frac{\|\check{U}[i, :]\|_2}{\|U_r[i, :]\|_2} \\ &= \sqrt{\max_i \frac{\check{\mu}_i \cdot \dim(\check{U})}{\mu_i r}} \quad (\text{see (73)}) \\ &\leq \sqrt{\max_i \frac{\check{\mu}_i \cdot 2r}{\mu_i r}}, \quad (\dim(\check{U}) \leq 2r) \end{aligned}$$

where the third line above holds because $U'_r \subset \check{U}$, by construction in the proof of Lemma 3.

B. Measurement Operator

Next, we slightly modify the measurement operator in (8) to gain more versatility. Throughout Section VIII, for probabilities $\{p_{ij}\}_{i,j=1}^n \subset (0, 1]$, we assume that $\mathcal{R}_p(\cdot)$ takes $M \in \mathbb{R}^{n \times n}$ to $\mathcal{R}_p(M) \in \mathbb{R}^{n \times n}$ defined as

$$\mathcal{R}_p(M) = \sum_{i,j=1}^n \frac{\epsilon_{ij}}{p_{ij}} \cdot M[i, j] \cdot C_{ij}, \quad (79)$$

where each ϵ_{ij} is a Bernoulli random variable that takes 1 with a probability of p_{ij} (and 0 otherwise). Moreover, $\{\epsilon_{ij}\}$ are independent. Recall also that $C_{ij} \in \mathbb{R}^{n \times n}$ is the $[i, j]$ th canonical matrix. Throughout Section VIII, we will assume that $\{p_{ij}\} \subset [l, h]$ for some $0 < l \leq h \leq 1$. In particular, note that we retrieve the measurement operator in (8) by setting $p_{ij} = l = h = p$ for every i, j .

Through $\mathcal{R}_p(\cdot)$, we measure M . In particular, for noise level $e \geq 0$, let $Y = \mathcal{R}_p(M + E)$ with $\|\mathcal{R}_p(E)\|_F \leq e$ be the (possibly noisy) matrix of measurements. To (approximately) complete M given the measurement matrix Y and prior knowledge about column/row spaces of M , we solve

$$\begin{cases} \min_X \left\| Q_{\check{U}_r, \lambda} \cdot X \cdot Q_{\check{V}_r, \rho} \right\|_*, \\ \text{subject to } \|\mathcal{R}_p(X) - Y\|_F \leq e, \end{cases} \quad (80)$$

where $Q_{\check{U}_r, \lambda}, Q_{\check{V}_r, \rho} \in \mathbb{R}^{n \times n}$ encapsulate our prior knowledge about M and were defined in (17). In the rest of Section VIII, we analyze Program (80) with $\mathcal{R}_p(\cdot)$ defined in (79). Theorem 2 will follow as a special case, as explained later.

Understanding the properties of the measurement operator is imperative to the development of supporting theory. To list these properties, let us introduce the following norms which, respectively, measure the (weighted) largest entry and largest ℓ_2 norm of the rows of a matrix [4]: For a matrix $Z \in \mathbb{R}^{n \times n}$, we set

$$\begin{aligned} \|Z\|_{\mu(\infty)} &= \left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} \cdot Z \cdot \left(\frac{\nu r}{n} \right)^{-\frac{1}{2}} \right\|_{\infty} \\ &= \max_{i,j} \sqrt{\frac{n}{\mu_i r}} \cdot |Z[i, j]| \cdot \sqrt{\frac{n}{\nu_j r}}, \quad (\text{see (77)}) \end{aligned} \quad (81)$$

where $\|A\|_{\infty}$ returns the largest entry of matrix A in magnitude. Moreover, for $Z \in \mathbb{R}^{n \times n}$, we let

$$\begin{aligned} \|Z\|_{\mu(\infty, 2)} &= \left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} \cdot Z \right\|_{2 \rightarrow \infty} \vee \left\| \left(\frac{\nu r}{n} \right)^{-\frac{1}{2}} \cdot Z^* \right\|_{2 \rightarrow \infty} \\ &= \left(\max_i \sqrt{\frac{n}{\mu_i r}} \|Z[i, :]\|_2 \right) \vee \left(\max_j \sqrt{\frac{n}{\nu_j r}} \|Z[:, j]\|_2 \right), \end{aligned} \quad (82)$$

return the largest ℓ_2 norm of the columns and rows of Z after reweighting. Above, $\|A\|_{2 \rightarrow \infty}$ returns the largest ℓ_2 norm of the rows of matrix A .

Establishing the following results is a standard practice in the use of large deviation bounds, stated here without proof from [4].

Lemma 5. [4, Lemma 9] For probabilities $\{p_{ij}\} \subset (0, 1]$, consider the measurement operator $\mathcal{R}_p(\cdot)$ defined in (79). Let the subspace \mathbb{T} , defined in (44), be the support of $M_r \in \mathbb{R}^{n \times n}$ and let $\mathcal{P}_{\mathbb{T}}(\cdot)$ be the orthogonal projection onto \mathbb{T} . Then, except with a probability of at most n^{-20} , it holds that

$$\|(\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \circ \mathcal{R}_p \circ \mathcal{P}_{\mathbb{T}})(\cdot)\|_{F \rightarrow F} \leq \frac{1}{2},$$

provided that

$$\frac{(\mu_i + \nu_j) r \log n}{n} \lesssim p_{ij} \leq 1, \quad \forall i, j \in [1 : n]. \quad (83)$$

Above, $\|\mathcal{A}(\cdot)\|_{F \rightarrow F} = \sup_{\|X\|_F \leq 1} \|\mathcal{A}(X)\|_F$ is the operator norm of the linear map $\mathcal{A}(\cdot)$, and $(\mathcal{A} \circ \mathcal{B})(\cdot) = \mathcal{A}(\mathcal{B}(\cdot))$ stands for composition of operators $\mathcal{A}(\cdot)$ and $\mathcal{B}(\cdot)$.

Lemma 6. [4, Lemma 10] Consider the same setup as in Lemma 5 and fix a matrix $Z \in \mathbb{R}^{n \times n}$. Except with a probability of at most n^{-20} , it holds that

$$\|(\mathcal{I} - \mathcal{R}_p)(Z)\| \lesssim \|Z\|_{\mu(\infty)} + \|Z\|_{\mu(\infty, 2)},$$

provided that (83) holds. Here, $\mathcal{I}(\cdot)$ is the identity operator, so that $\mathcal{I}(Z) = Z$ for any Z . In particular, multiplying the far-left side of (83) by a factor of $\Delta^2 \geq 1$ will divide the right-hand side above by a factor of Δ .

Lemma 7. [4, Lemma 11] Consider the same setup as Lemma 5 and fix a matrix $Z \in \mathbb{T} \subset \mathbb{R}^{n \times n}$ (i.e., $\mathcal{P}_{\mathbb{T}}(Z) = Z$). Then, except with a probability of at most n^{-20} , it holds that

$$\begin{aligned} \|(\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \circ \mathcal{R}_p \circ \mathcal{P}_{\mathbb{T}})(Z)\|_{\mu(\infty, 2)} \\ \leq \frac{1}{2} \|Z\|_{\mu(\infty)} + \frac{1}{2} \|Z\|_{\mu(\infty, 2)}, \end{aligned}$$

as long as (83) holds.¹²

Lemma 8. [4, Lemma 12] Consider the same setup as in Lemma 5 and fix a matrix $Z \in \mathbb{T} \subset \mathbb{R}^{n \times n}$. Then, except with a probability of at most n^{-20} , it holds that

$$\|(\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \circ \mathcal{R}_p \circ \mathcal{P}_{\mathbb{T}})(Z)\|_{\mu(\infty)} \leq \frac{1}{2} \|Z\|_{\mu(\infty)},$$

as long as (83) holds.

At times, we will find it more convenient to work with the closely related operator $\overline{\mathcal{R}}_p(\cdot)$ that takes $\overline{Z} \in \mathbb{R}^{n \times n}$ to $\overline{\mathcal{R}}_p(\overline{Z}) \in \mathbb{R}^{n \times n}$, where

$$\overline{\mathcal{R}}_p(\overline{Z}) = B_L^* \cdot \mathcal{R}_p(B_L \overline{Z} B_R^*) \cdot B_R, \quad (84)$$

in which we applied (31) and (79). Corresponding to the measurement operator $\mathcal{R}_p(\cdot)$, we also define the orthogonal projection $\mathcal{P}_p(\cdot)$ that projects onto the support of $\mathcal{R}_p(\cdot)$. More specifically, $\mathcal{P}_p(\cdot)$ takes $Z \in \mathbb{R}^{n \times n}$ to $\mathcal{P}_p(Z) \in \mathbb{R}^{n \times n}$ defined as

$$\mathcal{P}_p(Z) = \sum_{i,j} \epsilon_{ij} Z[i, j] \cdot C_{ij}. \quad (85)$$

Similarly, $\overline{\mathcal{P}}_p(\cdot)$ is the orthogonal xprojection that takes $\overline{Z} \in \mathbb{R}^{n \times n}$ to $\overline{\mathcal{P}}_p(\overline{Z}) \in \mathbb{R}^{n \times n}$, defined as

$$\overline{\mathcal{P}}_p(\overline{Z}) = B_L^* \cdot \mathcal{P}_p(B_L \overline{Z} B_R^*) \cdot B_R. \quad (86)$$

¹²Lemmas 11 and 12 in [4] are in fact more general than the statements here.

We will only $\mathcal{P}_p(\cdot)$ and $\overline{\mathcal{P}}_p(\cdot)$ once. Below, we collect a few basic properties of all these operators which, for the sake of completeness, are proved in Appendix C.

Lemma 9. For an arbitrary $Z \in \mathbb{R}^{n \times n}$, with $\overline{Z} = B_L^* Z B_R \in \mathbb{R}^{n \times n}$, and for the operators $\mathcal{R}_p(\cdot)$, $\overline{\mathcal{R}}_p(\cdot)$, $\mathcal{P}_p(\cdot)$, $\overline{\mathcal{P}}_p(\cdot)$ defined above, it holds that

$$\langle \overline{Z}, \overline{\mathcal{R}}_p(\overline{Z}) \rangle = \langle Z, \mathcal{R}_p(Z) \rangle, \quad (87)$$

$$\|\overline{\mathcal{R}}_p(\overline{Z})\|_F = \|\mathcal{R}_p(Z)\|_F. \quad (88)$$

Additionally, if $\{p_{ij}\} \subset [l, h]$ with $0 < l \leq h \leq 1$, it holds true that

$$(\overline{\mathcal{R}}_p \circ \overline{\mathcal{R}}_p)(\cdot) \succcurlyeq \overline{\mathcal{R}}_p(\cdot). \quad (89)$$

$$\|\overline{\mathcal{R}}_p(\cdot)\|_{F \rightarrow F} = \|\mathcal{R}_p(\cdot)\|_{F \rightarrow F} \leq l^{-1}. \quad (90)$$

Above, for operators $\mathcal{A}(\cdot)$ and $\mathcal{B}(\cdot)$, $\mathcal{A}(\cdot) \succcurlyeq \mathcal{B}(\cdot)$ means that $\langle Z, \mathcal{A}(Z) \rangle \succcurlyeq \langle Z, \mathcal{B}(Z) \rangle$ for any matrix Z . Lastly,

$$\|\overline{\mathcal{P}}_p(\overline{Z})\|_F \leq h \|\overline{\mathcal{R}}_p(\overline{Z})\|_F. \quad (91)$$

We are now in position to study Program (80) in more detail.

C. Body of the Analysis

Assume for now that $M = M_r$ is rank- r and that $e = 0$, namely noise is absent. Extending to noise and nearly low-rank matrices is straightforward, as described later. For solution \widehat{M} , let $H := \widehat{M} - M$ be the error. In Program (80), by feasibility of M and optimality of $\widehat{M} = M + H$, we may write that

$$\left\| Q_{\tilde{U}_{r,\lambda}}(M + H) Q_{\tilde{V}_{r,\rho}} \right\|_* \leq \left\| Q_{\tilde{U}_{r,\lambda}} M Q_{\tilde{V}_{r,\rho}} \right\|_*. \quad (92)$$

The right-hand side above can itself be bounded as

$$\begin{aligned} & \left\| Q_{\tilde{U}_{r,\lambda}} M Q_{\tilde{V}_{r,\rho}} \right\|_* \\ &= \left\| Q_{\tilde{U}_{r,\lambda}} M_r Q_{\tilde{V}_{r,\rho}} \right\|_* \\ &\leq \|B_L O_L L \overline{M}_r R^* O_R^* B_R^*\|_* \\ &= \|L \overline{M}_r R^*\|_* \quad (\text{rotational invariance}) \\ &= \left\| \begin{bmatrix} L_{11} \overline{M}_{r,11} R_{11} & \\ & 0_{n-r} \end{bmatrix} \right\|_*, \quad (\text{see (42)}) \quad (93) \end{aligned}$$

where the third line applies (42) and then (38). Similarly, the left-hand side of (92) can be bounded from below as follows:

$$\begin{aligned} & \left\| Q_{\tilde{U}_{r,\lambda}}(M + H) Q_{\tilde{V}_{r,\rho}} \right\|_* \\ &= \left\| Q_{\tilde{U}_{r,\lambda}}(M_r + H) Q_{\tilde{V}_{r,\rho}} \right\|_* \\ &\geq \|B_L O_L L (\overline{M}_r + \overline{H}) R^* O_R^* B_R^*\|_* - \|M_r\|_* \\ &= \|L (\overline{M}_r + \overline{H}) R^*\|_* \quad (\text{rotational invariance}) \\ &= \left\| \begin{bmatrix} L_{11} \overline{M}_{r,11} R_{11} & \\ & 0_{n-r} \end{bmatrix} + L \overline{H} R^* \right\|_*, \quad (\text{see (42)}) \quad (94) \end{aligned}$$

where the third line uses (40) and then (38). By substituting (93) and (94) back in (92), and then using the convexity of nuclear norm, we arrive at the following:

$$\begin{aligned} & \langle L\bar{H}R^*, G \rangle \\ & \leq \left\| \begin{bmatrix} L_{11}\bar{M}_{r,11}R_{11} & & \\ & 0_{n-r} & \\ & & \end{bmatrix} + L\bar{H}R^* \right\|_* \\ & \quad - \left\| \begin{bmatrix} L_{11}\bar{M}_{r,11}R_{11} & & \\ & 0_{n-r} & \\ & & \end{bmatrix} \right\|_* \\ & \quad \forall G \in \partial \left\| \begin{bmatrix} L_{11}\bar{M}_{r,11}R_{11} & & \\ & 0_{n-r} & \\ & & \end{bmatrix} \right\|_*. \end{aligned} \quad (95)$$

Above, $\partial\|A\|_*$ stands for the sub-differential of the nuclear norm at A (e.g., [8, equation 2.9]). In order to fully characterize the sub-differential, we take the following steps. First, from (42), recall that $\text{rank}(\bar{M}_{r,11}) = \text{rank}(\bar{M}_r) = \text{rank}(M_r) = r$. Second, assume that $\lambda \cdot \rho \neq 0$ so that $\text{rank}(L_{11}\bar{M}_{11}R_{11}) = \text{rank}(\bar{M}_{11}) = r$ too (see (37) for the definitions of L_{11}, R_{11}). Third, consider the SVD

$$L_{11}\bar{M}_{r,11}R_{11} = \bar{U}_r \bar{\Lambda}_r \bar{V}_r^*, \quad \bar{U}_r, \bar{V}_r \in \mathbb{R}^{r \times r},$$

and define the *sign matrix* $S \in \mathbb{R}^{n \times n}$ as

$$S := \begin{bmatrix} S_{11} & & \\ & 0_{n-r} & \\ & & \end{bmatrix} := \begin{bmatrix} \bar{U}_r \bar{V}_r^* & & \\ & & \\ & & 0_{n-r} \end{bmatrix}. \quad (96)$$

Finally, the sub-differential in (95) is specified as

$$\begin{aligned} & \partial \left\| \begin{bmatrix} L_{11}\bar{M}_{r,11}R_{11} & & \\ & 0_{n-r} & \\ & & \end{bmatrix} \right\|_* \\ & = \left\{ G \in \mathbb{R}^{n \times n} : \right. \\ & \quad G = \begin{bmatrix} S_{11} \in \mathbb{R}^{r \times r}, & & \\ & G_{22} \in \mathbb{R}^{(n-r) \times (n-r)} & \\ & & \end{bmatrix} \\ & \quad \left. \text{and } \|G_{22}\| \leq 1 \right\} \\ & = \left\{ G \in \mathbb{R}^{n \times n} : \right. \\ & \quad \mathcal{P}_{\bar{T}}(G) = S = \begin{bmatrix} S_{11} & & \\ & 0_{n-r} & \\ & & \end{bmatrix}, \\ & \quad \left. \text{and } \|\mathcal{P}_{\bar{T}^\perp}(G)\| \leq 1 \right\}. \end{aligned} \quad (\text{see (46)}) \quad (97)$$

For the record, (96) also implies that

$$\begin{aligned} \text{rank}(S) &= \text{rank}(S_{11}) = r, \quad \|S\| = \|S_{11}\| = 1, \\ \|S\|_F &= \|S_{11}\|_F = \sqrt{r}. \end{aligned} \quad (98)$$

With the characterization of the sub-differential in (97), we rewrite (95) as

$$\left\langle L\bar{H}R^*, S + \begin{bmatrix} 0_r & & \\ & G_{22} & \\ & & \end{bmatrix} \right\rangle \leq 0,$$

for any $G_{22} \in \mathbb{R}^{(n-r) \times (n-r)}$ such that $\|G_{22}\| \leq 1$, from which it follows that

$$\begin{aligned} 0 & \geq \langle L\bar{H}R^*, S \rangle + \sup_{\|G_{22}\| \leq 1} \left\langle L\bar{H}R^*, \begin{bmatrix} 0_r & & \\ & G_{22} & \\ & & \end{bmatrix} \right\rangle \\ & = \langle L\bar{H}R^*, S \rangle + \sup_{\|G\| \leq 1} \langle \mathcal{P}_{\bar{T}^\perp}(L\bar{H}R^*), G \rangle \quad (\text{see (46)}) \\ & = \langle L\bar{H}R^*, S \rangle + \|\mathcal{P}_{\bar{T}^\perp}(L\bar{H}R^*)\|_* \\ & = \langle \bar{H}, L^* S R \rangle + \|\mathcal{P}_{\bar{T}^\perp}(L\bar{H}R^*)\|_* \\ & = \left\langle \bar{H}, \begin{bmatrix} L_{11}S_{11}R_{11} & L_{11}S_{11}R_{12} & \\ L_{12}^*S_{11}R_{11} & L_{12}^*S_{11}R_{12} & \\ & & 0_{n-2r} \end{bmatrix} \right\rangle \\ & \quad + \left\| \begin{bmatrix} 0_r & & \\ & L_{22}\bar{H}_{22}R_{22} & L_{22}\bar{H}_{23} \\ & \bar{H}_{32}R_{22} & \bar{H}_{33} \end{bmatrix} \right\|_* \\ & = \left\langle \bar{H}, \begin{bmatrix} L_{11}S_{11}R_{11} & L_{11}S_{11}R_{12} & \\ L_{12}^*S_{11}R_{11} & 0_r & \\ & & 0_{n-2r} \end{bmatrix} \right\rangle \\ & \quad + \langle \bar{H}_{22}, L_{12}^*S_{11}R_{12} \rangle \\ & \quad + \left\| \begin{bmatrix} 0_r & & \\ & L_{22}\bar{H}_{22}R_{22} & L_{22}\bar{H}_{23} \\ & \bar{H}_{32}R_{22} & \bar{H}_{33} \end{bmatrix} \right\|_* \\ & =: \langle \bar{H}, \bar{S}' \rangle + \langle \bar{H}_{22}, L_{12}^*S_{11}R_{12} \rangle + \|L' \cdot \mathcal{P}_{\bar{T}^\perp}(\bar{H}) \cdot R'\|_*, \end{aligned} \quad (99)$$

where the third line uses the duality of nuclear and spectral norms. The fourth identity applies (37) and (47). Above, we also conveniently defined $\bar{S}', L' \in \mathbb{R}^{n \times n}$ as

$$\begin{aligned} \bar{S}' &:= \begin{bmatrix} L_{11}S_{11}R_{11} & L_{11}S_{11}R_{12} & \\ L_{12}^*S_{11}R_{11} & 0_r & \\ & & 0_{n-2r} \end{bmatrix}, \\ L' &:= \begin{bmatrix} 0_r & & \\ & L_{22} & \\ & & I_{n-2r} \end{bmatrix}. \end{aligned} \quad (100)$$

We define $R' \in \mathbb{R}^{n \times n}$ similarly. The key feature in (99) is that $\bar{S}' \in \bar{T}$, namely $\bar{S}' = \mathcal{P}_{\bar{T}}(\bar{S}')$. Before going any further, let us record the following properties of \bar{S}', L', R' for future reference:

$$\begin{aligned} & \|\bar{S}'\|_F \\ & = \left\| \begin{bmatrix} L_{11}S_{11}R_{11} & L_{11}S_{11}R_{12} & \\ L_{12}^*S_{11}R_{11} & 0_r & \\ & & 0_{n-2r} \end{bmatrix} \right\|_F \quad (\text{see (100)}) \\ & \leq \left\| \begin{bmatrix} L_{11}S_{11}R_{11} & L_{11}S_{11}R_{12} & \\ L_{12}^*S_{11}R_{11} & L_{12}^*S_{11}R_{12} & \\ & & 0_{n-2r} \end{bmatrix} \right\|_F \\ & \leq \left\| \begin{bmatrix} L_{11} & L_{12} \end{bmatrix} \right\| \cdot \|S_{11}\|_F \cdot \left\| \begin{bmatrix} R_{11} & R_{12} \end{bmatrix} \right\| \\ & = \left\| \begin{bmatrix} L_{11} & L_{12} \end{bmatrix} \right\| \cdot \|S\|_F \cdot \left\| \begin{bmatrix} R_{11} & R_{12} \end{bmatrix} \right\| \quad (\text{see (96)}) \\ & = \sqrt{r} \left\| \begin{bmatrix} L_{11} & L_{12} \end{bmatrix} \right\| \cdot \left\| \begin{bmatrix} R_{11} & R_{12} \end{bmatrix} \right\| \quad (\text{see (98)}) \\ & \leq \sqrt{r} \cdot \sqrt{\frac{\lambda^4 \cos^2 u_1 + \sin^2 u_1}{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \cdot \sqrt{\frac{\rho^4 \cos^2 v_1 + \sin^2 v_1}{\rho^2 \cos^2 v_1 + \sin^2 v_1}} \\ & =: \sqrt{r} \cdot \alpha_5(u_1, v_1, \lambda, \rho), \end{aligned} \quad (101)$$

$$\|L'\| \leq \|L\| = 1. \quad (\text{see (37), (38), and (100)}) \quad (102)$$

The second inequality in (101) uses the fact that $\|AB\|_F \leq \|A\| \cdot \|B\|_F$ for conforming matrices A, B . The last inequality there uses Lemma 4. Let us now continue the line of argument in (99) by writing that

$$\begin{aligned} 0 &\geq \langle \bar{H}, \bar{S}' \rangle + \langle \bar{H}_{22}, L_{12}^* S_{11} R_{12} \rangle + \|L' \mathcal{P}_{\bar{T}^\perp}(\bar{H}) R'\|_* \\ &\geq \langle \bar{H}, \bar{S}' \rangle - \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \cdot \|L_{12}\| \|S_{11}\| \|R_{12}\| \\ &\quad + \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* - \|\mathcal{P}_{\bar{T}^\perp}(\bar{H}) - L' \mathcal{P}_{\bar{T}^\perp}(\bar{H}) R'\|_* \\ &= \langle \bar{H}, \bar{S}' \rangle + (1 - \|L_{12}\| \|R_{12}\|) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \\ &\quad - \|\mathcal{P}_{\bar{T}^\perp}(I_n) \cdot \mathcal{P}_{\bar{T}^\perp}(\bar{H}) \cdot \mathcal{P}_{\bar{T}^\perp}(I_n) - L' \mathcal{P}_{\bar{T}^\perp}(\bar{H}) R'\|_* \\ &\geq \langle \bar{H}, \bar{S}' \rangle + (1 - \|L_{12}\| \|R_{12}\|) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \\ &\quad - \|\mathcal{P}_{\bar{T}^\perp}(I_n) - L'\| \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \\ &\quad - \|L'\| \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \|\mathcal{P}_{\bar{T}^\perp}(I_n) - R'\| \\ &\geq \langle \bar{H}, \bar{S}' \rangle + (1 - \|L_{12}\| \|R_{12}\|) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \\ &\quad - \|I_r - L_{22}\| \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* - \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \|I_r - R_{22}\| \\ &= \langle \bar{H}, \bar{S}' \rangle + (1 - \|L_{12}\| \|R_{12}\| - \|I_r - L_{22}\| \\ &\quad - \|I_r - R_{22}\|) \cdot \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \\ &\geq \langle \bar{H}, \bar{S}' \rangle + \left(1 - \frac{3\sqrt{1 - \lambda^2} \sin u_1}{2\sqrt{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \right. \\ &\quad \left. - \frac{3\sqrt{1 - \rho^2} \sin v_1}{2\sqrt{\rho^2 \cos^2 v_1 + \sin^2 v_1}} \right) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \\ &=: \langle \bar{H}, \bar{S}' \rangle + (1 - \alpha_7(u_1, v_1, \lambda, \rho)) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_*. \quad (103) \end{aligned}$$

Above, the first inequality applies (99). The second inequality uses the Holder's inequality, and the fact that \bar{H}_{22} is a submatrix of $\mathcal{P}_{\bar{T}^\perp}(\bar{H})$ (see (46)), which yields $\|\bar{H}_{22}\|_* \leq \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_*$. An application of the triangle inequality also appears in the second inequality. The first identity above uses (98) and then (47). In the third inequality, we deployed the polarization identity (56), followed with an application of the triangle inequality, and then applied the fact that $\|AB\|_* \leq \|A\| \cdot \|B\|_*$ for conforming matrices A, B . In the fourth inequality, we used (100) and (102). In the last inequality above, we used Lemma 4 and the observation that $ab + a + b \leq \frac{3}{2}(a + b)$ for $a, b \in [0, 1]$. At this point, we introduce the dual certificate. Validating the claims below are postponed until Appendix D.

Lemma 10. (Dual certificate) *Let the subspace $\bar{T} \subset \mathbb{R}^{n \times n}$ be the support of \bar{M}_r , as defined in (46). Assume that $\min_{i,j} p_{ij} \geq l$, with $l^{-1} = 1/l(n) = \text{poly}(n)$, so that l^{-1} is bounded above by a polynomial in n (of finite degree). Recall also the operators $\bar{\mathcal{R}}_p(\cdot)$ and $\bar{\mathcal{P}}_p(\cdot)$ from (84) and (86), respectively. Then, as long as*

$$\begin{aligned} &\max[\log(\alpha_5 \cdot n), 1] \cdot \frac{(\mu_i + \nu_j)r \log n}{n} \\ &\cdot \max \left[\alpha_6 \left(1 + \max_i \frac{\check{\mu}_i}{\mu_i} + \max_j \frac{\check{\nu}_j}{\nu_j} \right), 1 \right] \lesssim p_{ij} \leq 1, \end{aligned}$$

for all $i, j \in [1 : n]$, the following statements are all true. First,

$$\|(\mathcal{P}_{\bar{T}} - \mathcal{P}_{\bar{T}} \circ \bar{\mathcal{R}}_p \circ \mathcal{P}_{\bar{T}})(\cdot)\|_{F \rightarrow F} \leq \frac{1}{2}, \quad (104)$$

except with a probability of $o(n^{-19})$. Moreover, there exists $\bar{\Lambda} \in \mathbb{R}^{n \times n}$ such that

$$\|\bar{S}' - \mathcal{P}_{\bar{T}}(\bar{\Lambda})\|_F \leq \frac{l}{4\sqrt{2}}, \quad (105)$$

$$\|\mathcal{P}_{\bar{T}^\perp}(\bar{\Lambda})\| \leq \frac{1}{2}, \quad (106)$$

$$\bar{\Lambda} = \bar{\mathcal{P}}_p(\bar{\Lambda}). \quad (107)$$

Here,

$$\begin{aligned} \alpha_5 &= \alpha_5(u_1, v_1, \lambda, \rho) \\ &:= \sqrt{\frac{\lambda^4 \cos^2 u_1 + \sin^2 u_1}{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \cdot \sqrt{\frac{\rho^4 \cos^2 v_1 + \sin^2 v_1}{\rho^2 \cos^2 v_1 + \sin^2 v_1}}, \end{aligned}$$

$$\begin{aligned} \alpha_6 &= \alpha_6(u_1, v_1, \lambda, \rho) \\ &:= \left(\sqrt{\frac{\lambda^2 \cos^2 u_1 + \sin^2 u_1}{\rho^2 \cos^2 v_1 + \sin^2 v_1}} + \sqrt{\frac{\rho^2 \cos^2 v_1 + \sin^2 v_1}{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \right) \\ &\quad \cdot \left(\sqrt{\lambda^4 \cos^2 u_1 + \sin^2 u_1} + \sqrt{\rho^4 \cos^2 v_1 + \sin^2 v_1} \right). \end{aligned}$$

Under Lemma 10, in particular, there exists $\bar{\Lambda}$ that satisfies (105-107). This allows us to continue the line of argument in (99) by writing that

$$\begin{aligned} 0 &\geq \langle \bar{H}, \bar{S}' \rangle + (1 - \alpha_7) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \quad (\text{see (99)}) \\ &= \langle \bar{H}, \mathcal{P}_{\bar{T}}(\bar{\Lambda}) \rangle + \langle \bar{H}, \bar{S}' - \mathcal{P}_{\bar{T}}(\bar{\Lambda}) \rangle \\ &\quad + (1 - \alpha_7) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \\ &= \langle \bar{H}, \bar{\Lambda} \rangle - \langle \bar{H}, \mathcal{P}_{\bar{T}^\perp}(\bar{\Lambda}) \rangle + \langle \bar{H}, \bar{S}' - \mathcal{P}_{\bar{T}}(\bar{\Lambda}) \rangle \\ &\quad + (1 - \alpha_7) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \\ &= -\langle \bar{H}, \mathcal{P}_{\bar{T}^\perp}(\bar{\Lambda}) \rangle + \langle \bar{H}, \bar{S}' - \mathcal{P}_{\bar{T}}(\bar{\Lambda}) \rangle \\ &\quad + (1 - \alpha_7) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \\ &\geq -\|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \|\mathcal{P}_{\bar{T}^\perp}(\bar{\Lambda})\| - \|\mathcal{P}_{\bar{T}}(\bar{H})\|_F \\ &\quad \cdot \|\bar{S}' - \mathcal{P}_{\bar{T}}(\bar{\Lambda})\|_F + (1 - \alpha_7) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \\ &\geq -\frac{1}{2} \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* - \frac{l}{4\sqrt{2}} \|\mathcal{P}_{\bar{T}}(\bar{H})\|_F \\ &\quad + (1 - \alpha_7) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \\ &= \left(\frac{1}{2} - \alpha_7 \right) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* - \frac{l}{4\sqrt{2}} \|\mathcal{P}_{\bar{T}}(\bar{H})\|_F, \quad (108) \end{aligned}$$

or, equivalently,

$$\left(\frac{1}{2} - \alpha_7 \right) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \leq \frac{l}{4\sqrt{2}} \|\mathcal{P}_{\bar{T}}(\bar{H})\|_F. \quad (109)$$

The bound above is nontrivial if $\alpha_7 = \alpha_7(u_1, v_1, \lambda, \rho) < \frac{1}{2}$. In (108), we used the Holder's inequality. Also, the third inequality in (108) uses (105) and (106). Lastly, the third

identity in (108) holds because $\langle \bar{H}, \bar{\Lambda} \rangle = 0$. To see why this is the case, first set $\bar{H} = B_L^* H B_R$ (see (40)). Then note that

$$\begin{aligned} \|\bar{\mathcal{R}}_p(\bar{H})\|_F &= \|\mathcal{R}_p(H)\|_F \quad (\text{see (88)}) \\ &= \left\| \mathcal{R}_p(\widehat{M} - M) \right\|_F \\ &= 0, \end{aligned} \quad (110)$$

where the last line uses the feasibility of M and \widehat{M} in Program (19). Therefore,

$$\begin{aligned} \langle \bar{H}, \bar{\Lambda} \rangle &= \langle \bar{H}, \bar{\mathcal{P}}_p(\bar{\Lambda}) \rangle \quad (\text{see (107)}) \\ &= \langle \bar{\mathcal{P}}_p(\bar{H}), \bar{\Lambda} \rangle \quad (\bar{\mathcal{P}}_p(\cdot) \text{ is self-adjoint}) \\ &= 0, \quad (\text{see (110)}) \end{aligned}$$

thereby verifying the fourth line of (109). On the other hand, to find a matching upper bound for (109), we reason as follows. First, note that

$$\begin{aligned} \|\bar{\mathcal{R}}_p(\mathcal{P}_{\bar{T}}(\bar{H}))\|_F &= \|\bar{\mathcal{R}}_p(\mathcal{P}_{\bar{T}^\perp}(\bar{H}))\|_F \quad (\text{see (110)}) \\ &\leq \|\bar{\mathcal{R}}_p(\cdot)\|_{F \rightarrow F} \cdot \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_F \\ &\leq \frac{1}{l} \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_F. \quad (\text{see (90)}) \end{aligned} \quad (111)$$

Under Lemma 10, $\bar{\mathcal{R}}_p(\cdot)$ acts as a near-isometry on the subspace \bar{T} , which allows us to find a matching lower bound for (111):

$$\begin{aligned} &\|\bar{\mathcal{R}}_p(\mathcal{P}_{\bar{T}}(\bar{H}))\|_F^2 \\ &= \langle \mathcal{P}_{\bar{T}}(\bar{H}), (\bar{\mathcal{R}}_p \circ \bar{\mathcal{R}}_p)(\mathcal{P}_{\bar{T}}(\bar{H})) \rangle \\ &\geq \langle \mathcal{P}_{\bar{T}}(\bar{H}), \bar{\mathcal{R}}_p(\mathcal{P}_{\bar{T}}(\bar{H})) \rangle \quad (\text{see (89)}) \\ &= \langle \mathcal{P}_{\bar{T}}(\bar{H}), \mathcal{P}_{\bar{T}}(\bar{H}) \rangle \\ &\quad + \langle \mathcal{P}_{\bar{T}}(\bar{H}), (\mathcal{P}_{\bar{T}} \circ \bar{\mathcal{R}}_p \circ \mathcal{P}_{\bar{T}} - \mathcal{P}_{\bar{T}}) \circ \mathcal{P}_{\bar{T}}(\bar{H}) \rangle \\ &\geq \|\mathcal{P}_{\bar{T}}(\bar{H})\|_F^2 - \|\mathcal{P}_{\bar{T}} \circ \bar{\mathcal{R}}_p \circ \mathcal{P}_{\bar{T}} - \mathcal{P}_{\bar{T}}\|_{F \rightarrow F} \cdot \|\mathcal{P}_{\bar{T}}(\bar{H})\|_F^2 \\ &\geq \frac{1}{2} \|\mathcal{P}_{\bar{T}}(\bar{H})\|_F^2. \quad (\text{see (104)}) \end{aligned} \quad (112)$$

Comparing (111) to (112) yields that

$$\|\mathcal{P}_{\bar{T}}(\bar{H})\|_F \leq \frac{\sqrt{2}}{l} \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_F. \quad (113)$$

The inequality above, when put together with (109), leads us to

$$\left(\frac{1}{2} - \alpha_7\right) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_* \leq \frac{l}{4\sqrt{2}} \|\mathcal{P}_{\bar{T}}(\bar{H})\|_F, \quad (\text{see (109)})$$

which, as long as $\alpha_7 = \alpha_7(u_1, v_1, \lambda, \rho) \leq \frac{1}{8}$, immediately yields that

$$\mathcal{P}_{\bar{T}^\perp}(\bar{H}) = 0. \quad (114)$$

We extend the error bound above to \bar{H} by noting that

$$\begin{aligned} \|\bar{H}\|_F &\leq \|\mathcal{P}_{\bar{T}}(\bar{H})\|_F + \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_F \\ &\leq \left(\frac{\sqrt{2}}{l} + 1\right) \|\mathcal{P}_{\bar{T}^\perp}(\bar{H})\|_F \quad (\text{see (113)}) \\ &= 0. \quad (\text{see (114)}) \end{aligned}$$

Since $\|H\|_F = \|\bar{H}\|_F$ (see (40)), we find that $\widehat{M} = M$. Extending to nearly low-rank matrices ($M_{r^+} \neq 0$) and

accounting for noise ($e > 0$) is a straightforward generalization of Theorem 7 in [3], matching [10, Proposition 2] nearly verbatim. Such an argument would lead us to:

$$\begin{aligned} \|\widehat{M} - M\|_F &\lesssim \frac{\sqrt{h}}{l} \left\| Q_{\tilde{U}_r, \lambda} M_{r^+} Q_{\tilde{V}_r, \rho} \right\|_* + \frac{e\sqrt{nh}^{\frac{3}{2}}}{l} \\ &\leq \frac{\sqrt{h}}{l} \left\| Q_{\tilde{U}_r, \lambda} \right\| \cdot \|M_{r^+}\|_* \cdot \left\| Q_{\tilde{V}_r, \rho} \right\| + \frac{e\sqrt{nh}^{\frac{3}{2}}}{l} \\ &= \frac{\sqrt{h}}{l} \|M_{r^+}\|_* + \frac{e\sqrt{nh}^{\frac{3}{2}}}{l}, \quad (\text{see (38)}) \end{aligned}$$

where, in the first line, $\{p_{ij}\} \subset [l, h]$ and e is the noise level. The second inequality uses the fact that $\|AB\|_* \leq \|A\| \cdot \|B\|_*$ for conforming matrices A, B . We therefore arrive at the following result.

Theorem 11. (Leveraged matrix completion with prior knowledge) For integer r and matrix $M \in \mathbb{R}^{n \times n}$, let $M_r \in \mathbb{R}^{n \times n}$ be a rank- r truncation of M , and let $M_{r^+} = M - M_r$ be the residual. Let $\{\mu_i, \nu_i\}_{i=1}^n$ be the leverage scores of M_r (see (73) and (74)). Suppose that the r -dimensional subspaces \tilde{U}_r and \tilde{V}_r represent our prior knowledge about the column and row spaces of M_r , respectively, and let

$$u = \angle [U_r, \tilde{U}_r], \quad v = \angle [V_r, \tilde{V}_r],$$

denote the largest principal angles. Take $\{\check{\mu}_i, \check{\nu}_i\}_{i=1}^n$ to be the leverage scores of subspaces $\check{U} = \text{span}([U_r, \tilde{U}_r])$ and $\check{V} = \text{span}([V_r, \tilde{V}_r])$, respectively. Moreover, consider the probabilities $\{p_{ij}\}_{i,j=1}^n \subset [l, h]$ for $0 < l \leq h \leq 1$, and assume that $l^{-1} = 1/l(n)$ is bounded above by a polynomial in n of finite degree. Recalling (8), acquire the (possibly noisy) measurement matrix $Y = \mathcal{R}_p(M + E)$ where $\|\mathcal{R}_p(E)\|_F \leq e$ for noise level $e \geq 0$. Lastly, for $\lambda, \rho \in (0, 1]$, let \bar{M} be a solution of Program (80). Then, it holds that

$$\|\widehat{M} - M\|_F \lesssim \frac{\sqrt{h}}{l} \|M_{r^+}\|_* + \frac{e\sqrt{nh}^{\frac{3}{2}}}{l},$$

except with a probability of $o(n^{-19})$, and provided that

$$\begin{aligned} &\max[\log(\alpha_5 \cdot n), 1] \cdot \frac{(\mu_i + \nu_j)r \log n}{n} \\ &\cdot \max \left[\alpha_6 \left(1 + \max_i \frac{\check{\mu}_i}{\mu_i} + \max_j \frac{\check{\nu}_j}{\nu_j} \right), 1 \right] \lesssim p_{ij} \leq 1, \\ &\alpha_7 \leq \frac{1}{8}, \end{aligned} \quad (115)$$

for all $i, j \in [1 : n]$. Above, we set

$$\begin{aligned} \alpha_5 &= \alpha_5(u, v, \lambda, \rho) \\ &:= \sqrt{\frac{\lambda^4 \cos^2 u + \sin^2 u}{\lambda^2 \cos^2 u + \sin^2 u}} \cdot \sqrt{\frac{\rho^4 \cos^2 v + \sin^2 v}{\rho^2 \cos^2 v + \sin^2 v}}, \end{aligned}$$

$$\begin{aligned} \alpha_6 &= \alpha_6(u, v, \lambda, \rho) \\ &:= \left(\sqrt{\frac{\lambda^2 \cos^2 u + \sin^2 u}{\rho^2 \cos^2 v + \sin^2 v}} + \sqrt{\frac{\rho^2 \cos^2 v + \sin^2 v}{\lambda^2 \cos^2 u + \sin^2 u}} \right) \\ &\cdot \left(\sqrt{\lambda^4 \cos^2 u + \sin^2 u} + \sqrt{\rho^4 \cos^2 v + \sin^2 v} \right), \end{aligned}$$

$$\begin{aligned}\alpha_7 &= \alpha_7(u, v, \lambda, \rho) \\ &:= \frac{3\sqrt{1-\lambda^2}\sin u}{2\sqrt{\lambda^2\cos^2 u + \sin^2 u}} + \frac{3\sqrt{1-\rho^2}\sin v}{2\sqrt{\rho^2\cos^2 v + \sin^2 v}}.\end{aligned}$$

In particular, suppose we replace the leverage scores $\{\mu_i, \nu_i\}_i$ with their upper bound $\eta_r(M)$ in Section VIII (see (75)), i.e., we set $\mu_i = \nu_i = \eta(M_r)$ for all i . Then, all lemmas in Section VIII-B still hold (confer [4]) and so does the rest of the analysis in Section VIII. This leads to Theorem 2 after noting that $\eta(\check{U}\check{V}^*)$ upper bounds $\{\check{\mu}_i, \check{\nu}_i\}_i$ for all i (see (76)).

REFERENCES

- [1] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.
- [2] B. Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12:3413–3430, 2011.
- [3] E. J. Candès and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010.
- [4] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward. Completing any low-rank matrix, provably. *Journal of Machine Learning Research*, 16:2999–3034, 2015.
- [5] N. Rao, H.-F. Yu, P. K. Ravikumar, and I. S. Dhillon. Collaborative filtering with graph information: Consistency and scalable methods. In *Advances in Neural Information Processing Systems*, pages 2098–2106, 2015.
- [6] A. Aravkin, R. Kumar, H. Mansour, B. Recht, and F. J. Herrmann. Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation. *SIAM Journal on Scientific Computing*, 36(5):S237–S266, 2014.
- [7] T. Sun and C.-H. Zhang. A graphical approach to the analysis of matrix completion. *Stochastic Processes and their Applications*, 2016.
- [8] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.
- [9] M. Fazel, E. Candès, B. Recht, and P. Parrilo. Compressed sensing and robust recovery of low rank matrices. In *42nd Asilomar Conference on Signals, Systems and Computers*, pages 1043–1047. IEEE, 2008.
- [10] A. Eftekhari, M. B. Wakin, and R. A. Ward. MC²: A two-phase algorithm for leveraged matrix completion. *arXiv preprint arXiv:1609.01795*, 2016.
- [11] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.
- [12] M. Xu, R. Jin, and Z.-H. Zhou. Speedup matrix completion with side information: Application to multi-label learning. In *Advances in Neural Information Processing Systems*, pages 2301–2309, 2013.
- [13] P. Jain and I. S. Dhillon. Provable inductive matrix completion. *arXiv preprint arXiv:1306.0626*, 2013.
- [14] K.-Y. Chiang, C.-J. Hsieh, and I. S. Dhillon. Matrix completion with noisy side information. In *Advances in Neural Information Processing Systems*, pages 3447–3455, 2015.
- [15] K. Mohan and M. Fazel. Reweighted nuclear norm minimization with application to system identification. In *Proceedings of the 2010 American Control Conference*, pages 2953–2959. IEEE, 2010.
- [16] A. Eftekhari, L. Balzano, and M. B. Wakin. What to expect when you are expecting on the Grassmannian. *arXiv preprint arXiv:1611.07216*, 2016.
- [17] S. Negahban and M. J. Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *Journal of Machine Learning Research*, 13(May):1665–1697, 2012.
- [18] 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*, pages 2056–2064, 2010.
- [19] M. P. Friedlander, H. Mansour, R. Saab, and O. Yilmaz. Recovering compressively sampled signals using partial support information. *IEEE Transactions on Information Theory*, 58(2):1122–1134, 2012.
- [20] H. Mansour and O. Yilmaz. Weighted-l1 minimization with multiple weighting sets. In *SPIE Optical Engineering+ Applications*, pages 813809–813809. International Society for Optics and Photonics, 2011.
- [21] E. J. Candès, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier Analysis and Applications*, 14(5-6):877–905, 2008.

- [22] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2013.
- [23] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, 2011.

APPENDIX A PROOF OF LEMMA 3

Let us focus on the column spaces first. Let $U_r, \tilde{U}_r \in \mathbb{R}^{n \times r}$ be orthonormal bases for the subspaces U_r, \tilde{U}_r , respectively. Without loss of generality, assume that

$$U_r^* \tilde{U}_r = \cos u \in \mathbb{R}^{r \times r}. \quad (116)$$

(Otherwise, take the SVD of $U_r^* \tilde{U}_r$ and redefine U_r and \tilde{U}_r accordingly.) To simplify the exposition, assume also that all $\sin u$ is invertible, namely all principal angles $\{u_i\}_{i=1}^r$ are nonzero. We set

$$U'_r := -(I_n - U_r U_r^*) \tilde{U}_r (\sin(u))^{-1} \in \mathbb{R}^{n \times r}. \quad (117)$$

Then, it is easy to verify that U'_r has orthonormal columns and that $U_r^* \tilde{U}'_r = -\sin u \in \mathbb{R}^{r \times r}$. Furthermore, we take $U''_{n-2r} \in \mathbb{R}^{n \times (n-2r)}$ with orthonormal columns such that

$$\text{span}(U''_{n-2r}) = \text{span}\left(\begin{bmatrix} U_r & U'_r \end{bmatrix}\right)^\perp. \quad (118)$$

Likewise, we set

$$\tilde{U}'_r := (I_n - \tilde{U}_r \tilde{U}_r^*) U_r (\sin(u))^{-1} \in \mathbb{R}^{n \times r},$$

which, we may again verify, has orthonormal columns and satisfies $U_r^* \tilde{U}'_r = \sin u \in \mathbb{R}^{r \times r}$. It is similarly confirmed that $U_r^* \tilde{U}'_r = \cos u \in \mathbb{R}^{r \times r}$. Lastly, we observe that

$$\begin{aligned}\text{span}\left(\begin{bmatrix} U_r & \tilde{U}_r \end{bmatrix}\right) \\ &= \text{span}\left(\begin{bmatrix} U_r & (I_n - U_r U_r^*) \tilde{U}_r \end{bmatrix}\right) \\ &= \text{span}\left(\begin{bmatrix} U_r & -(I_n - U_r U_r^*) \tilde{U}_r (\sin(u))^{-1} \end{bmatrix}\right) \\ &= \text{span}\left(\begin{bmatrix} U_r & U'_r \end{bmatrix}\right). \quad (\text{see (117)})\end{aligned}$$

A similar argument shows that $\text{span}(\begin{bmatrix} U_r & \tilde{U}_r \end{bmatrix}) = \text{span}(\begin{bmatrix} \tilde{U}_r & \tilde{U}'_r \end{bmatrix})$ and, overall, we find that

$$\begin{aligned}\text{span}\left(\begin{bmatrix} U_r & U'_r \end{bmatrix}\right) &= \text{span}\left(\begin{bmatrix} \tilde{U}_r & \tilde{U}'_r \end{bmatrix}\right) \\ &= \text{span}\left(\begin{bmatrix} U_r & \tilde{U}_r \end{bmatrix}\right),\end{aligned}$$

which completes the proof of Lemma 3.

APPENDIX B PROOF OF LEMMA 4

Below, we derive the claimed inequalities directly from the definition of L in (37), one by one:

$$\begin{aligned}\|L_{11}\| &= \|\Delta_L\| \quad (\text{see (37)}) \\ &= \max_i \sqrt{\lambda^2 \cos^2 u_i + \sin^2 u_i} \\ &= \sqrt{\lambda^2 \cos^2 u_1 + \sin^2 u_1}, \quad \left(u_1 = \max_i u_i \text{ and } \lambda \in (0, 1]\right)\end{aligned}$$

$$\begin{aligned}
\|L_{12}\| &\leq \max_i \frac{(1-\lambda^2) \sin u_i \cdot \cos u_i}{\sqrt{\lambda^2 \cos^2 u_i + \sin^2 u_i}} \quad (\text{see (37)}) \\
&\leq \max_i \frac{(1-\lambda^2) \sin u_i}{\sqrt{\lambda^2 \cos^2 u_i + \sin^2 u_i}} \\
&= \frac{(1-\lambda^2) \sin u_1}{\sqrt{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \quad \left(u_1 = \max_i u_i\right) \\
&\leq 1 - \lambda^2,
\end{aligned}$$

$$\begin{aligned}
\|I_r - L_{22}\| &= \|I_r - \lambda \Delta_L^{-1}\| \quad (\text{see (37)}) \\
&= \max_i \left| 1 - \frac{\lambda}{\sqrt{\lambda^2 \cos^2 u_i + \sin^2 u_i}} \right| \\
&= \max_i \left(1 - \frac{\lambda}{\sqrt{\lambda^2 \cos^2 u_i + \sin^2 u_i}} \right) \quad (\lambda \in (0, 1]) \\
&= 1 - \frac{\lambda}{\sqrt{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \quad \left(u_1 = \max_i u_i\right) \\
&\leq \frac{\sqrt{1-\lambda^2} \sin u_1}{\sqrt{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \quad \left(\sqrt{a+b} - \sqrt{a} \leq \sqrt{b}, a, b \geq 0\right) \\
&\leq \sqrt{1-\lambda^2},
\end{aligned}$$

$$\begin{aligned}
&\| [L_{11} \quad L_{12}] \|^2 \\
&= \max_i \left\| \left[\begin{array}{c} \sqrt{\lambda^2 + (1-\lambda^2) \sin^2 u_i} \\ \frac{(1-\lambda^2) \sin u_i \cos u_i}{\sqrt{\lambda^2 + (1-\lambda^2) \sin^2 u_i}} \end{array} \right] \right\|_2^2 \\
&= \max_i \frac{\lambda^4 + (1-\lambda^4) \sin^2 u_i}{\lambda^2 + (1-\lambda^2) \sin^2 u_i} \\
&= \frac{\lambda^4 + (1-\lambda^4) \sin^2 u_1}{\lambda^2 + (1-\lambda^2) \sin^2 u_1} \leq 1. \quad (\lambda \in (0, 1])
\end{aligned}$$

$$\begin{aligned}
&\| [L_{22} - I_r \quad L_{12}] \|^2 \\
&= \max_i \left\| \left[\begin{array}{c} \frac{\lambda}{\sqrt{\lambda^2 + (1-\lambda^2) \sin^2 u_i}} - 1 \\ \frac{(1-\lambda^2) \sin u_i \cos u_i}{\sqrt{\lambda^2 + (1-\lambda^2) \sin^2 u_i}} \end{array} \right] \right\|_2^2 \\
&= \max_i \left[\frac{\left(\lambda - \sqrt{\lambda^2 + (1-\lambda^2) \sin^2 u_i} \right)^2}{\lambda^2 + (1-\lambda^2) \sin^2 u_i} \right. \\
&\quad \left. \frac{(1-\lambda^2)^2 \sin^2 u_i \cos^2 u_i}{\lambda^2 + (1-\lambda^2) \sin^2 u_i} \right] \\
&\leq \max_i \frac{(1-\lambda^2) \sin^2 u_i + (1-\lambda^2)^2 \sin^2 u_i \cos^2 u_i}{\lambda^2 + (1-\lambda^2) \sin^2 u_i} \\
&\leq \max_i \frac{(1-\lambda^2) \sin^2 u_i + (1-\lambda^2) \sin^2 u_i}{\lambda^2 + (1-\lambda^2) \sin^2 u_i} \\
&= \frac{2(1-\lambda^2) \sin^2 u_1}{\lambda^2 + (1-\lambda^2) \sin^2 u_1}.
\end{aligned}$$

To derive the last bound above, we used the inequality $\sqrt{a} - \sqrt{b} \leq \sqrt{a-b}$ for $a \geq b \geq 0$. Similar bounds hold for R and its blocks, R_{11}, R_{12}, R_{22} . This completes the proof of Lemma 4.

APPENDIX C PROOF OF LEMMA 9

With $\bar{Z} = B_L^* Z B_R \in \mathbb{R}^{n \times n}$, note that

$$\begin{aligned}
\|\bar{\mathcal{R}}_p(\bar{Z})\|_F^2 &= \langle \bar{\mathcal{R}}_p(\bar{Z}), \bar{\mathcal{R}}_p(\bar{Z}) \rangle \\
&= \langle \mathcal{R}_p(Z), \mathcal{R}_p(Z) \rangle \quad (\text{see (84)}) \\
&= \|\mathcal{R}_p(Z)\|_F^2, \quad (119)
\end{aligned}$$

as claimed in (88), and (87) is proved similarly. Additionally, if $\{p_{ij}\} \subset [l, h]$ with $0 < l \leq h \leq 1$, we observe that

$$\begin{aligned}
&\langle \bar{Z}, \bar{\mathcal{R}}_p(\bar{Z}) \rangle \\
&= \langle Z, \mathcal{R}_p(Z) \rangle \quad (\text{see (87)}) \\
&= \sum_{i,j} \frac{\epsilon_{ij}}{p_{ij}} |Z[i, j]|^2 \quad (\text{see (8)}) \\
&\leq \sum_{i,j} \frac{\epsilon_{ij}}{p_{ij}^2} |Z[i, j]|^2 \quad (p_{i,j} \leq 1, \forall i, j) \\
&= \langle \mathcal{R}_p(Z), \mathcal{R}_p(Z) \rangle \quad (\epsilon_{ij}^2 = \epsilon_{ij}) \\
&= \langle \bar{\mathcal{R}}_p(\bar{Z}), \bar{\mathcal{R}}_p(\bar{Z}) \rangle \quad (\text{see (119)}) \\
&= \langle \bar{Z}, (\bar{\mathcal{R}}_p \circ \bar{\mathcal{R}}_p)(\bar{Z}) \rangle, \quad (\bar{\mathcal{R}}_p(\cdot) \text{ is self-adjoint})
\end{aligned}$$

for every $\bar{Z} = B_L^* Z B_R \in \mathbb{R}^{n \times n}$, thereby establishing that (89). Also note that

$$\begin{aligned}
\|\bar{\mathcal{R}}_p(\bar{Z})\|_F &= \|\mathcal{R}_p(Z)\|_F \quad (\text{see (119)}) \\
&= \sqrt{\sum_{i,j} \frac{\epsilon_{ij}}{p_{ij}^2} |Z[i, j]|^2} \quad (\text{see (8)}) \\
&\leq \frac{\|Z\|_F}{l}, \quad (p_{ij} \geq l, \forall i, j) \quad (120)
\end{aligned}$$

for any $\bar{Z} = B_L^* Z B_R \in \mathbb{R}^{n \times n}$. This proves (90). Lastly, to verify (91), we write that

$$\begin{aligned}
\|\bar{\mathcal{R}}_p(\bar{Z})\|_F &= \sqrt{\sum_{i,j} \frac{\epsilon_{ij}}{p_{ij}^2} |Z[i, j]|^2} \quad (\text{see (120)}) \\
&\geq h^{-1} \sqrt{\sum_{i,j} \epsilon_{ij} |Z[i, j]|^2} \\
&= h^{-1} \|\mathcal{P}_p(Z)\|_F \quad (\text{see (85)}) \\
&= h^{-1} \|\bar{\mathcal{P}}_p(\bar{Z})\|_F. \quad (\text{similar to (88)}) \quad (121)
\end{aligned}$$

This completes the proof of Lemma 9.

APPENDIX D PROOF OF LEMMA 10 (CONSTRUCTING THE DUAL CERTIFICATE)

Recall (100) and conveniently define

$$S' := B_L \bar{S}' B_R^* \in \mathbb{R}^{n \times n}, \quad \Lambda := B_L \bar{\Lambda} B_R^* \in \mathbb{R}^{n \times n}. \quad (122)$$

Because of (46) and (100), we note that $\bar{S}' \in \bar{T}$ and that consequently

$$S' \in T, \quad (\text{see (122) and (45)}) \quad (123)$$

so that $\mathcal{P}_T(S') = S'$. After making this transformation and recalling (48), (84), and (86), it is easily verified that it suffices to prove that

$$\|(\mathcal{P}_T - \mathcal{P}_T \circ \mathcal{R}_p \circ \mathcal{P}_T)(\cdot)\|_{F \rightarrow F} \leq \frac{1}{2}, \quad (124)$$

with high probability, and to prove the existence of $\Lambda \in \mathbb{R}^{n \times n}$ such that

$$\|S' - \mathcal{P}_T(\Lambda)\|_F \leq \frac{l}{4\sqrt{2}}, \quad (125)$$

$$\|\mathcal{P}_{T^\perp}(\Lambda)\| \leq \frac{1}{2}, \quad (126)$$

$$\Lambda = \mathcal{P}_p(\Lambda). \quad (127)$$

According to Lemma 5, (124) holds if the sampling probabilities $\{p_{ij}\}$ are sufficiently large (see (83)) and except with a probability of at most n^{-20} .

It remains to construct an admissible Λ . To that end, we use the golfing scheme as follows [23], [4]. Instead of $\mathcal{R}_p(\cdot)$, we equivalently measure M through K independent applications of $\mathcal{R}_q(\cdot)$ (with probabilities $\{q_{ij}\}$ and integer K to be set later). Given K , the two measurement schemes are equivalent (in distribution) if

$$(1 - q_{ij})^K = 1 - p_{ij}, \quad i, j \in [1 : n]. \quad (128)$$

We in fact assume that the set of observed entries at hand is generated through K independent applications of $\mathcal{R}_q(\cdot)$ (rather than an application of $\mathcal{R}_p(\cdot)$). Next, we set $\Lambda^0 = 0$, and define Λ^k , $k \in [1 : K]$, as follows:

$$\Lambda^k := \sum_{k=1}^{k'} \mathcal{R}_q(W^{k'-1}), \quad (129)$$

$$W^{k'} := S' - \mathcal{P}_T(\Lambda^{k'}). \quad (130)$$

We then set $\Lambda = \Lambda^K$; it readily follows that Λ satisfies (127), once we recall (85). We turn our attention to verifying (125) next. For every $k \in [1 : K]$, note that

$$\begin{aligned} W^k &= S' - \mathcal{P}_T(\Lambda^k) \\ &= S' - \mathcal{P}_T(\Lambda^{k-1}) - (\mathcal{P}_T \circ \mathcal{R}_q)(W^{k-1}) \quad (\text{see (129)}) \\ &= S' - \mathcal{P}_T(\Lambda^{k-1}) - (\mathcal{P}_T \circ \mathcal{R}_q \circ \mathcal{P}_T)(W^{k-1}) \\ &= W^{k-1} - (\mathcal{P}_T \circ \mathcal{R}_q \circ \mathcal{P}_T)(W^{k-1}) \quad (\text{see (130)}) \\ &= (\mathcal{P}_T - \mathcal{P}_T \circ \mathcal{R}_q \circ \mathcal{P}_T)(W^{k-1}), \quad (W^{k-1} \in \mathbb{T}) \end{aligned} \quad (131)$$

where the third line above uses the fact that $W^{k-1} \in \mathbb{T}$ from (130) and (123). Under Lemma 5, it then follows that

$$\begin{aligned} \|W^k\|_F &\leq \|(\mathcal{P}_T - \mathcal{P}_T \circ \mathcal{R}_q \circ \mathcal{P}_T)(\cdot)\|_{F \rightarrow F} \|W^{k-1}\|_F \\ &\leq \frac{1}{2} \|W^{k-1}\|_F, \end{aligned} \quad (132)$$

as long as

$$\frac{(\mu_i + \nu_j) r \log n}{n} \lesssim q_{ij} \leq 1, \quad \forall i, j \in [1 : n], \quad (\text{see (83)}) \quad (133)$$

and except with a probability of at most n^{-20} . It immediately follows that

$$\begin{aligned} \|W^K\|_F &\leq \|(\mathcal{P}_T - \mathcal{P}_T \circ \mathcal{R}_q \circ \mathcal{P}_T)(\cdot)\|_{F \rightarrow F}^K \cdot \|W^0\|_F \\ &\leq \left(\frac{1}{2}\right)^K \|S'\|_F \\ &= \left(\frac{1}{2}\right)^K \|B_L \bar{S}' B_R^*\|_F \quad (\text{see (122)}) \\ &= \left(\frac{1}{2}\right)^K \|\bar{S}'\|_F \quad (B_L, B_R \text{ are orthonormal bases}) \\ &\leq \left(\frac{1}{2}\right)^K \alpha_5 \sqrt{r}, \quad (\text{see (101)}) \end{aligned} \quad (134)$$

except with a probability of at most Kn^{-20} (invoking the union bound). The second line above uses (132) and then (130) for $k=0$. From (130) and (134), it now follows that

$$\|S' - \mathcal{P}_T(\Lambda)\|_F = \|S' - \mathcal{P}_T(\Lambda^K)\|_F = \|W^K\|_F \leq \frac{l}{4\sqrt{2}},$$

if we take

$$K \gtrsim \max \left[\log \left(\frac{8\alpha_5 \sqrt{r}}{l} \right), 1 \right].$$

Assume that l^{-1} is polynomial in n , namely that $l^{-1} = l^{-1}(n)$ is bounded above by a polynomial in n of finite degree. We therefore established that $\Lambda = \Lambda^K$, as constructed above, satisfies (125) with $K \approx \log(\beta n)$ and except with a probability of at most

$$Kn^{-20} = O(\log(\alpha_5 \cdot n)) \cdot n^{-20} = o(n^{-19}). \quad (135)$$

It remains to verify that Λ also meets the remaining requirements in (126). Introducing a factor $\Delta > 0$ to be set later, we observe that

$$\begin{aligned} \|\mathcal{P}_{T^\perp}(\Lambda)\| &= \|\mathcal{P}_{T^\perp}(\Lambda^K)\| \\ &\leq \sum_{k=1}^K \|(\mathcal{P}_{T^\perp} \circ \mathcal{R}_q)(W^{k-1})\| \quad (\text{see (129)}) \\ &= \sum_{k=1}^K \|(\mathcal{P}_{T^\perp} \circ (\mathcal{I} - \mathcal{R}_q))(W^{k-1})\| \quad (W^{k-1} \in \mathbb{T}) \\ &\leq \sum_{k=1}^K \|(\mathcal{I} - \mathcal{R}_q)(W^{k-1})\| \\ &\leq \frac{1}{\Delta} \left[\sum_{k=1}^K \|W^{k-1}\|_{\mu(\infty)} + \|W^{k-1}\|_{\mu(\infty, 2)} \right], \quad (\text{Lemma 6}) \end{aligned} \quad (136)$$

as long as

$$\frac{\Delta^2(\mu_i + \nu_j) r \log n}{n} \lesssim q_{ij} \leq 1, \quad i, j \in [1 : n], \quad (137)$$

and except for a probability of $Kn^{-20} = o(n^{-19})$ since $l^{-1} = \text{poly}(n)$ (as described in (135)). Consider the weighted infinity norm in the last line of (136). Under Lemma 8, we note that

$$\begin{aligned} & \|W^{k-1}\|_{\mu(\infty)} \\ &= \|(\mathcal{P}_T - \mathcal{P}_T \circ \mathcal{R}_q \circ \mathcal{P}_T)(W^{k-2})\|_{\mu(\infty)} \quad (\text{see (131)}) \\ &\leq \frac{1}{2} \|W^{k-2}\|_{\mu(\infty)} \quad (\text{see Lemma 8}) \\ &\leq \left(\frac{1}{2}\right)^{k-1} \|W^0\|_{\mu(\infty)}, \end{aligned} \quad (138)$$

as long as (133) holds and except for a probability of at most $(k-1)n^{-20} = o(n^{-19})$, since $k \leq K$. Next, we consider the second norm in the last line of (136). Appealing to Lemma 7, we observe that

$$\begin{aligned} & \|W^{k-1}\|_{\mu(\infty,2)} \\ &= \|(\mathcal{P}_T - \mathcal{P}_T \circ \mathcal{R}_q \circ \mathcal{P}_T)(W^{k-2})\|_{\mu(\infty,2)} \quad (\text{see (131)}) \\ &\leq \frac{1}{2} \|W^{k-2}\|_{\mu(\infty)} + \frac{1}{2} \|W^{k-2}\|_{\mu(\infty,2)} \quad (\text{see Lemma 7}) \\ &\leq \left(\frac{1}{2}\right)^{k-1} \|W^0\|_{\mu(\infty)} + \frac{1}{2} \|W^{k-2}\|_{\mu(\infty,2)}, \quad (\text{see (138)}) \end{aligned}$$

as long as (133) holds and except for a probability of at most $kn^{-20} = o(n^{-19})$. From the above recursion, it immediately follows that

$$\begin{aligned} & \|W^{k-1}\|_{\mu(\infty,2)} \\ &\leq (k-1) \left(\frac{1}{2}\right)^{k-1} \|W^0\|_{\mu(\infty)} + \left(\frac{1}{2}\right)^{k-1} \|W^0\|_{\mu(\infty,2)}. \end{aligned} \quad (139)$$

Substituting the last two estimates back into (136), we arrive at

$$\begin{aligned} & \|\mathcal{P}_{T^\perp}(\Lambda^K)\| \\ &\leq \frac{1}{\Delta} \left[\sum_{k=1}^K \|W^{k-1}\|_{\mu(\infty)} + \|W^{k-1}\|_{\mu(\infty,2)} \right] \quad (\text{see (136)}) \\ &\leq \frac{1}{\Delta} \left[\sum_{k=1}^K \left(\frac{1}{2}\right)^{k-1} \|W^0\|_{\mu(\infty)} \right. \\ &\quad \left. + \left((k-1) \left(\frac{1}{2}\right)^{k-1} \|W^0\|_{\mu(\infty)} + \left(\frac{1}{2}\right)^{k-1} \|W^0\|_{\mu(\infty,2)} \right) \right] \\ &\leq \frac{4}{\Delta} \|W^0\|_{\mu(\infty)} + \frac{2}{\Delta} \|W^0\|_{\mu(\infty,2)} \\ &= \frac{4}{\Delta} \|S'\|_{\mu(\infty)} + \frac{2}{\Delta} \|S'\|_{\mu(\infty,2)}. \quad (\text{see (130)}) \end{aligned} \quad (140)$$

The second inequality above uses (138) and (139). Both norms in the last line above are bounded in Appendix E.

Lemma 12. For $S' \in \mathbb{R}^{n \times n}$ defined in (122) (also see (100)), it holds that

$$\begin{aligned} \|S'\|_{\mu(\infty)} &\leq \alpha_6 \left(1 + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} + \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}} \right), \\ \|S'\|_{\mu(\infty,2)} &\leq 2\alpha_6 \left(1 + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} + \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}} \right), \end{aligned}$$

where

$$\begin{aligned} \alpha_6 &= \alpha_6(u_1, v_1, \lambda, \rho) \\ &= \left(\sqrt{\frac{\lambda^2 \cos^2 u_1 + \sin^2 u_1}{\rho^2 \cos^2 v_1 + \sin^2 v_1}} + \sqrt{\frac{\rho^2 \cos^2 v_1 + \sin^2 v_1}{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \right) \\ &\quad \cdot \left(\sqrt{\lambda^4 \cos^2 u_1 + \sin^2 u_1} + \sqrt{\rho^4 \cos^2 v_1 + \sin^2 v_1} \right). \end{aligned} \quad (141)$$

In light of Lemma 12, we accordingly update (140) to read

$$\begin{aligned} & \|\mathcal{P}_{T^\perp}(\Lambda)\| \\ &= \|\mathcal{P}_{T^\perp}(\Lambda^K)\| \\ &\leq \frac{4}{\Delta} \|S'\|_{\mu(\infty)} + \frac{2}{\Delta} \|S'\|_{\mu(\infty,2)} \quad (\text{see (140)}) \\ &\leq \frac{8}{\Delta} \cdot \alpha_6 \left(1 + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} + \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}} \right) \quad (\text{Lemma 12}) \\ &\leq \frac{1}{2}, \end{aligned} \quad (142)$$

where we took

$$\Delta = 16\alpha_6 \left(1 + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} + \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}} \right).$$

After recalling (137), we observe that (142) (equivalently, (126)) holds provided that

$$\alpha_6^2 \left(1 + \max_i \frac{\check{\mu}_i}{\mu_i} + \max_j \frac{\check{\nu}_j}{\nu_j} \right) \frac{(\mu_i + \nu_j) r \log n}{n} \lesssim q_{ij} \leq 1, \quad (143)$$

and except with a probability of $o(n^{-20})$. Overall, we conclude that the dual certificate exists as long as

$$\begin{aligned} & \max \left[\alpha_6^2 \left(1 + \max_i \frac{\check{\mu}_i}{\mu_i} + \max_j \frac{\check{\nu}_j}{\nu_j} \right), 1 \right] \cdot \frac{(\mu_i + \nu_j) r \log n}{n} \\ &\lesssim q_{ij} \leq 1, \quad i, j \in [1 : n], \end{aligned}$$

$$K = \max[\log(\alpha_5 \cdot n), 1].$$

Lastly, recall the relation between $\{q_{ij}\}$, K and $\{p_{ij}\}$ in (128):

$$\begin{aligned} p_{ij} &= 1 - (1 - q_{ij})^K \quad (\text{see (128)}) \\ &\gtrsim K \cdot q_{ij} \\ &= \max[\log(\alpha_5 \cdot n), 1] \cdot \frac{(\mu_i + \nu_j) r \log n}{n} \\ &\quad \cdot \max \left[\alpha_6 \left(1 + \sqrt{\max_i \frac{\check{\mu}_i}{\mu_i}} + \sqrt{\max_j \frac{\check{\nu}_j}{\nu_j}} \right), 1 \right]. \end{aligned}$$

The second line holds if $\{q_{ij}\}$ are sufficiently small, i.e., when n is sufficiently large. This completes the proof of Lemma 10.

APPENDIX E
PROOF OF LEMMA 12

Here, we will estimate the norms of S' (see (122) and (100)). We write that

$$\begin{aligned} & \|S'\|_{\mu(\infty)} \\ &= \left\| \left(\frac{\mu r}{n}\right)^{-\frac{1}{2}} \cdot S' \cdot \left(\frac{\nu r}{n}\right)^{-\frac{1}{2}} \right\|_{\infty} \quad (\text{see (81)}) \\ &= \left\| \left(\frac{\mu r}{n}\right)^{-\frac{1}{2}} \cdot B_L \bar{S}' B_R^* \cdot \left(\frac{\nu r}{n}\right)^{-\frac{1}{2}} \right\|_{\infty} \quad (\text{see (122)}) \\ &= \left\| \left(\frac{\mu r}{n}\right)^{-\frac{1}{2}} \cdot B_L \begin{bmatrix} L_{11} S_{11} R_{11} & L_{11} S_{11} R_{12} \\ L_{12}^* S_{11} R_{11} & 0_r \\ & & 0_{n-2r} \end{bmatrix} \right. \\ & \quad \left. \cdot B_R^* \cdot \left(\frac{\nu r}{n}\right)^{-\frac{1}{2}} \right\|_{\infty}, \end{aligned}$$

where, in the last line, we used (100). It follows that

$$\begin{aligned} & \|S'\|_{\mu(\infty)} \\ &\leq \left\| \left(\frac{\mu r}{n}\right)^{-\frac{1}{2}} \cdot U_r \cdot L_{11} S_{11} R_{11} \cdot V_r^* \cdot \left(\frac{\nu r}{n}\right)^{-\frac{1}{2}} \right\|_{\infty} \\ &+ \left\| \left(\frac{\mu r}{n}\right)^{-\frac{1}{2}} \cdot U_r \cdot L_{11} S_{11} R_{12} \cdot (V_r')^* \cdot \left(\frac{\nu r}{n}\right)^{-\frac{1}{2}} \right\|_{\infty} \\ &+ \left\| \left(\frac{\mu r}{n}\right)^{-\frac{1}{2}} \cdot U_r \cdot L_{12}^* S_{11} R_{11} \cdot V_r^* \cdot \left(\frac{\nu r}{n}\right)^{-\frac{1}{2}} \right\|_{\infty} \quad (\text{see (31)}) \\ &\leq \left\| \left(\frac{\mu r}{n}\right)^{-\frac{1}{2}} U_r \right\|_{2 \rightarrow \infty} \cdot \|L_{11} S_{11} R_{11}\| \cdot \left\| \left(\frac{\nu r}{n}\right)^{-\frac{1}{2}} V_r \right\|_{2 \rightarrow \infty} \\ &+ \left\| \left(\frac{\mu r}{n}\right)^{-\frac{1}{2}} U_r \right\|_{2 \rightarrow \infty} \cdot \|L_{11} S_{11} R_{12}\| \cdot \left\| \left(\frac{\nu r}{n}\right)^{-\frac{1}{2}} V_r' \right\|_{2 \rightarrow \infty} \\ &+ \left\| \left(\frac{\mu r}{n}\right)^{-\frac{1}{2}} U_r \right\|_{2 \rightarrow \infty} \cdot \|L_{12}^* S_{11} R_{11}\| \\ & \quad \cdot \left\| \left(\frac{\nu r}{n}\right)^{-\frac{1}{2}} V_r \right\|_{2 \rightarrow \infty}, \end{aligned}$$

where the last inequality uses the fact that $\|AB\|_{\infty} \leq \|A\|_{2 \rightarrow \infty} \cdot \|B^*\|_{2 \rightarrow \infty}$ for conforming matrices A, B . We continue by writing that

$$\begin{aligned} & \|S'\|_{\mu(\infty)} \\ &\leq \|L_{11}\| \|S_{11}\| \|R_{11}\| + \|L_{11}\| \|S_{11}\| \|R_{12}\| \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}} \\ & \quad + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} \|L_{12}\| \|S_{11}\| \|R_{11}\| \quad (\text{see (78)}) \\ &= \|L_{11}\| \|R_{11}\| + \|L_{11}\| \|R_{12}\| \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}} \\ & \quad + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} \|L_{12}\| \|R_{11}\|, \quad (\text{see (98)}) \end{aligned}$$

which itself leads to

$$\begin{aligned} & \|S'\|_{\mu(\infty)} \\ &\leq \|L_{11}\| \left(\|R_{11}\| + \|R_{12}\| \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}} \right) \\ & \quad + \|R_{11}\| \left(\|L_{11}\| + \|L_{12}\| \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} \right) \\ &\leq \|L_{11}\| \cdot \|[R_{11} \ R_{12}]\| \cdot \left(1 + \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}} \right) \\ & \quad + \|R_{11}\| \cdot \|[L_{11} \ L_{12}]\| \cdot \left(1 + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} \right). \end{aligned}$$

After invoking Lemma 4, we continue simplifying the last line above by writing that

$$\begin{aligned} & \|S'\|_{\mu(\infty)} \\ &\leq \sqrt{\lambda^2 \cos^2 u_1 + \sin^2 u_1} \cdot \sqrt{\frac{\rho^4 \cos^2 v_1 + \sin^2 v_1}{\rho^2 \cos^2 v_1 + \sin^2 v_1}} \\ & \quad \cdot \left(1 + \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}} \right) + \sqrt{\rho^2 \cos^2 v_1 + \sin^2 v_1} \\ & \quad \cdot \sqrt{\frac{\lambda^4 \cos^2 u_1 + \sin^2 u_1}{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \cdot \left(1 + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} \right) \\ &\leq \left(\sqrt{\frac{\lambda^2 \cos^2 u_1 + \sin^2 u_1}{\rho^2 \cos^2 v_1 + \sin^2 v_1}} + \sqrt{\frac{\rho^2 \cos^2 v_1 + \sin^2 v_1}{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \right) \\ & \quad \cdot \left(\sqrt{\rho^4 \cos^2 v_1 + \sin^2 v_1} + \sqrt{\lambda^4 \cos^2 u_1 + \sin^2 u_1} \right) \\ & \quad \cdot \left(1 + \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}} + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} \right) \\ &=: \alpha_6(u_1, v_1, \lambda, \rho) \cdot \left(1 + \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}} + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} \right), \quad (144) \end{aligned}$$

where the last inequality above uses the fact that $ac + bd \leq (a+b)(c+d)$ whenever $a, b, c, d \geq 0$. As for $\|S'\|_{\mu(2, \infty)}$, we begin with writing that

$$\begin{aligned} & \left\| \left(\frac{\mu r}{n}\right)^{-\frac{1}{2}} S' \right\|_{2 \rightarrow \infty} \\ &= \left\| \left(\frac{\mu r}{n}\right)^{-\frac{1}{2}} \cdot B_L \bar{S}' B_R^* \right\|_{2 \rightarrow \infty} \quad (\text{see (122)}) \\ &= \left\| \left(\frac{\mu r}{n}\right)^{-\frac{1}{2}} \cdot B_L \bar{S}' \right\|_{2 \rightarrow \infty}. \quad (\text{rotational invariance}) \end{aligned}$$

It follows that

$$\begin{aligned}
& \left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} S' \right\|_{2 \rightarrow \infty} \\
& \leq \left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} \cdot U_r \cdot L_{11} S_{11} R_{11} \right\|_{2 \rightarrow \infty} \\
& + \left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} \cdot U_r \cdot L_{11} S_{11} R_{12} \right\|_{2 \rightarrow \infty} \\
& + \left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} \cdot U'_r \cdot L_{12}^* S_{11} R_{11} \right\|_{2 \rightarrow \infty} \quad (\text{see (31) and (100)}) \\
& \leq \left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} U_r \right\|_{2 \rightarrow \infty} \|L_{11}\| \|S_{11}\| \|R_{11}\| \\
& + \left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} U_r \right\|_{2 \rightarrow \infty} \|L_{11}\| \|S_{11}\| \|R_{12}\| \\
& + \left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} U'_r \right\|_{2 \rightarrow \infty} \|L_{12}\| \|S_{11}\| \|R_{11}\|,
\end{aligned}$$

where the last inequality applies the fact that $\|AB\|_{2 \rightarrow \infty} \leq \|A\|_{2 \rightarrow \infty} \cdot \|B\|$ for conforming matrices A, B . We continue by simplifying the last inequality and write that

$$\begin{aligned}
& \left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} S' \right\|_{2 \rightarrow \infty} \\
& \leq \|L_{11}\| \|R_{11}\| + \|L_{11}\| \|R_{12}\| + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} \|L_{12}\| \|R_{11}\| \\
& \leq \|L_{11}\| (\|R_{11}\| + \|R_{12}\|) \\
& + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} \|R_{11}\| (\|L_{11}\| + \|L_{12}\|) \\
& \leq 2 \|L_{11}\| \cdot \left\| \begin{bmatrix} R_{11} & R_{12} \end{bmatrix} \right\| \\
& + 2 \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} \|R_{11}\| \cdot \left\| \begin{bmatrix} L_{11} & L_{12} \end{bmatrix} \right\| \\
& \leq 2 \sqrt{\lambda^2 \cos^2 u_1 + \sin^2 u_1} \cdot \sqrt{\frac{\rho^4 \cos^2 v_1 + \sin^2 v_1}{\rho^2 \cos^2 v_1 + \sin^2 v_1}} \\
& + 2 \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} \cdot \sqrt{\rho^2 \cos^2 v_1 + \sin^2 v_1} \\
& \cdot \sqrt{\frac{\lambda^4 \cos^2 u_1 + \sin^2 u_1}{\lambda^2 \cos^2 u_1 + \sin^2 u_1}} \quad (\text{see Lemma 4}) \\
& \leq 2\alpha_6 \cdot \left(1 + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} \right), \quad (\text{see (144)}) \quad (145)
\end{aligned}$$

where the first inequality uses (78) and (98). Similarly, it holds that

$$\left\| \left(\frac{\nu r}{n} \right)^{-\frac{1}{2}} (S')^* \right\|_{2 \rightarrow \infty} \leq 2\alpha_6 \left(1 + \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}} \right), \quad (146)$$

so that, recalling (82), we find that

$$\begin{aligned}
\|S'\|_{\mu(2, \infty)} & = \left\| \left(\frac{\mu r}{n} \right)^{-\frac{1}{2}} S' \right\|_{2 \rightarrow \infty} \vee \left\| \left(\frac{\nu r}{n} \right)^{-\frac{1}{2}} (S')^* \right\|_{2 \rightarrow \infty} \\
& \leq 2\alpha_6 \left(1 + \sqrt{2 \max_i \frac{\check{\mu}_i}{\mu_i}} + \sqrt{2 \max_j \frac{\check{\nu}_j}{\nu_j}} \right), \quad (147)
\end{aligned}$$

which completes the proof of Lemma 12.