LOW-RANK MATRIX RECOVERY WITH PARTIAL SUBSPACE PRIORS: DEGREES OF FREEDOM AND RECOVERY GUARANTEES

Anonymous authors
Paper under double-blind review

ABSTRACT

Matrix recovery with prior information, such as subspaces derived from historical data or auxiliary features, has wide applications in recommender systems, signal processing, and multi-label learning. A central challenge is to quantify the information-theoretic measurement savings offered by prior information, and to verify whether these savings are achievable by recovery methods. To address this challenge, we characterize the intrinsic degrees of freedom (DoF) of low-rank matrices under column-subspace priors using quotient-space and Grassmann-manifold geometry. Specifically, we show that for a rank-r matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ with a known r_1-dimensional subspace of its column space, its DoF is equal to $nr + (m-r)(r-r_1)$. This provides a theoretical lower bound on the number of scalar measurements required to guarantee exact recovery. To verify that this lower bound is attainable, we develop a DoF-optimal recovery method (DORM), which relies on a designed two-stage measurement scheme and achieves the exact recovery using a number of measurements equal to the DoF. For general linear measurements, we propose a projected matrix recovery (PMR) framework, which uses subspace priors to reduce the original problem to standard rank-$(r-r_1)$ matrix recovery via measurement-space projection. In particular, under Gaussian measurements, we establish that PMR achieves exact recovery with high probability from $nr_1 + \mathcal{O}((r-r_1)(m+n-r))$ measurements, matching the DoF up to a constant factor. Synthetic and real-world experiments demonstrate measurement savings of the proposed methods, consistent with our theoretical analysis.

1 INTRODUCTION

Low-rank matrix recovery has emerged as a fundamental problem in machine learning and signal processing, with applications including recommender systems, image inpainting, multi-label learning, and wireless channel estimation (Radhakrishnan et al., 2025; Yang et al., 2024; Ramlatchan et al., 2018; Miao & Kou, 2022; Xia et al., 2023). The core principle underlying low-rank recovery is that many real-world matrices exhibit a low-rank structure (Nguyen et al., 2019). This structural property enables accurate reconstruction from partial observations, reducing the number of required measurements. The low-rank matrix recovery problem is usually constructed as follows

$$\min_{\mathbf{Z}\in\mathbb{R}^{m\times n}} \operatorname{rank}(\mathbf{Z}) \quad \text{subject to} \quad \mathbf{y} = \mathcal{A}(\mathbf{Z}), \tag{1}$$

where $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^K$ is a linear measurement operator. Traditional methods have been developed to solve the matrix recovery problem, such as nuclear-norm minimization (Recht et al., 2010), alternating minimization (Jain et al., 2013; Hardt, 2014), and singular-value thresholding (Cai et al., 2010). Under suitable assumptions, rank-r matrices can be recovered with sample complexity on the order of $r(m+n)$ up to logarithmic factors (Recht et al., 2010).

Beyond low rank, many applications provide additional prior information that can be used to improve matrix recovery (Zhao & Guo, 2017; Rao & Lau, 2015; So & Ye, 2007; Ardakani et al., 2024; Natarajan & Dhillon, 2014). Among these types of prior information, subspace priors are particularly common. In recommendation systems and the Netflix problem, the target matrix records users' ratings of items. The subspace prior represents known patterns in user attributes, item features,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

or users' historical ratings (Zhao & Guo, 2017; Ardakani et al., 2024). In massive multiple-input multiple-output (MIMO) channel estimation, the target matrix records the channels between the transmitter and receiver antennas. Signal directions that remain nearly unchanged over time, such as the line-of-sight path, can provide a subspace prior for estimating the current channel (Rao & Lau, 2015). In sensor-network localization, the target matrix describes the positions or pairwise distances of sensors. Known anchor positions and previously measured network geometry can provide prior subspace information for locating the remaining sensors (So & Ye, 2007). In disease–gene association prediction, the target matrix records the associations between diseases and genes. Known disease characteristics, gene functions, and previously verified associations can provide subspace priors for predicting unknown associations (Natarajan & Dhillon, 2014).

Many studies have analyzed matrix recovery with prior subspace information (Xu et al., 2013; Chiang et al., 2015; Eftekhari et al., 2018; Bayat & Daei, 2020; López, 2023; Chang et al., 2025). These works show that suitable prior subspaces or feature spaces can reduce the required number of observations. For example, prior subspaces containing the true row and column spaces can accelerate recovery (Xu et al., 2013); noisy or inexact subspaces have been studied in matrix completion (Chiang et al., 2015); weighted nuclear-norm methods exploit principal-angle information (Eftekhari et al., 2018; Bayat & Daei, 2020); and more recent analyses cover structured prior subspaces and partial subspace information (López, 2023; Chang et al., 2025). A more detailed discussion of related work is provided in Appendix A.

Although these studies (Xu et al., 2013; Chiang et al., 2015; Eftekhari et al., 2018; Bayat & Daei, 2020; López, 2023; Chang et al., 2025) have demonstrated the benefits of incorporating prior information into matrix recovery, the number of measurements depends on the recovery algorithm. A more fundamental question is therefore: what is the minimum number of measurements required for matrix recovery with subspace prior information? This is an information-theoretic question about the fundamental measurement limit, independent of any particular recovery algorithm. This question is not only theoretically interesting, but also practical. For example, using fewer measurements means we can collect less data, which directly lowers acquisition costs and reduces storage needs.

The degrees of freedom (DoF) provide a theoretical lower bound on the number of scalar measurements. For example, for matrix recovery without prior information, the DoF of a rank-r matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ is $r(m+n-r)$, as shown in Recht et al. (2010). The work in Recht et al. (2010) established that, under certain conditions, such matrices can be recovered from approximately $\mathcal{O}(r(m+n))$ measurements, which is consistent with DoF. However, to the best of our knowledge, no existing work has analyzed the DoF of matrices under prior information. To fill this gap, we establish a theoretical framework to quantify how prior information reduces DoF, providing an algorithm-independent lower bound on the minimum number of measurements required for matrix recovery. Moreover, we propose a DoF-optimal recovery method (DORM), which attains this bound under its designed measurements. Our work establishes the theoretical framework connecting prior information to measurement complexity in matrix recovery, providing fundamental insights and algorithms for applications where subspace priors are naturally available.

Our main contributions are as follows.

- We characterize the DoF of low-rank matrices under complete and partial column subspace priors using quotient-space and Grassmann-manifold geometry. When the complete column space of a rank-r matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ is known, we show its DoF reduces to nr. When only an r_1-dimensional subspace ($r_1 < r$) of its column space is known, we show that the DoF becomes $nr + (m-r)(r-r_1)$. Compared with the complete-subspace case, the additional term $(m-r)(r-r_1)$ quantifies the missing subspace information.

- We propose DORM, a concrete DoF-optimal recovery method for matrices with complete or partial column space priors. Under its designed measurement schemes and the stated nondegeneracy condition, DORM exactly recovers the target matrix using a number of scalar measurements equal to its DoF, proving that the DoF lower bound is attainable.

- Under Gaussian measurements, we propose projected matrix recovery (PMR) for matrix recovery with subspace priors, without DORM's specialized measurement design. The proposed PMR uses the available subspace prior information to construct a projection in measurement space, transforming the original problem into a standard reduced-rank matrix

recovery problem. We show that PMR achieves exact recovery with high probability using $nr_1 + \mathcal{O}((r - r_1)(m + n - r))$ measurements, matching the DoF up to a constant factor.

2 DEGREES OF FREEDOM OF LOW-RANK MATRICES AND SUBSPACES

The DoF count provides a fundamental benchmark for the number of measurements required to determine a structured matrix. For a rank-r matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$, prior work has established that $\mathrm{DoF}(\mathbf{M}) = (m + n - r)r$ (Recht et al., 2010). Since each scalar linear measurement provides at most one independent constraint, the number K of measurements must satisfy

$$K \geq \mathrm{DoF}(\mathbf{M}) = (m + n - r)r. \tag{2}$$

This inequality provides a theoretical lower bound on the number of measurements, which means that any algorithm requires at least $(m + n - r)r$ scalar linear measurements to guarantee exact recovery of a rank-r matrix in $\mathbb{R}^{m\times n}$. For example, Recht et al. (2010) establish nuclear-norm recovery with $\mathcal{O}(r(m + n)\log(mn))$ suitable random linear measurements.

To determine the DoF of matrices with column-subspace priors, we first characterize the DoF of a subspace. Analogously to the matrix DoF, this is the minimum number of independent scalar measurements required to determine the subspace. Geometrically, an r-dimensional subspace $\mathcal{U} \subset \mathbb{R}^m$ is a point on the Grassmann manifold $\mathrm{Gr}(r, m)$. Any such subspace is spanned by a basis matrix $\mathbf{U} \in \mathbb{R}^{m\times r}$ with mr parameters. Since $\mathbf{U}$ and $\mathbf{UT}$ span the same subspace for any invertible $\mathbf{T} \in \mathbb{R}^{r\times r}$, removing this change-of-basis ambiguity yields the quotient

$$\mathrm{Gr}(r, m) \simeq \{\mathbf{U} \in \mathbb{R}^{m\times r} : \mathrm{rank}(\mathbf{U}) = r\}/\mathrm{GL}(r),$$

where $\mathrm{GL}(r)$ denotes the general linear group of invertible $r \times r$ real matrices. The dimension of this quotient space is given by the following standard result.

Proposition 2.1 (Degrees of freedom of a subspace)**.** *Let $\mathcal{U} \subseteq \mathbb{R}^m$ be an r-dimensional subspace, and let $\mathrm{Gr}(r, m)$ denote the Grassmannian of all such subspaces. Then*

$$\mathrm{DoF}(\mathcal{U}) := \dim\big(\mathrm{Gr}(r, m)\big) = mr - r^2 = (m - r)r. \tag{3}$$

Therefore, by Proposition 2.1, determining an unknown r-dimensional subspace requires at least $(m - r)r$ independent scalar constraints from a DoF perspective.

3 DEGREES OF FREEDOM UNDER COLUMN-SUBSPACE PRIOR INFORMATION

In this section, we characterize the DoF of a matrix under column-subspace priors. We consider both complete knowledge of the column space (Xu et al., 2013; Zhao & Guo, 2017) and partial knowledge of the column space (Rao & Lau, 2015; So & Ye, 2007; Natarajan & Dhillon, 2014).

3.1 DEGREES OF FREEDOM OF LOW-RANK MATRIX UNDER COMPLETE SUBSPACE INFORMATION

Suppose that $\mathbf{M} \in \mathbb{R}^{m\times n}$ has rank r, and let $\mathcal{U} = \mathrm{col}(\mathbf{M})$ denote its r-dimensional column subspace, where $\mathrm{col}(\cdot)$ denotes the span of the columns of a matrix. Complete subspace information means $\mathcal{U}$ is known, or equivalently, that a basis matrix $\mathbf{U} \in \mathbb{R}^{m\times r}$ spanning $\mathcal{U}$ is available, so that

$$\mathbf{M} = \mathbf{UR}, \tag{4}$$

where $\mathbf{R} \in \mathbb{R}^{r\times n}$ is the coefficient matrix associated with the basis matrix $\mathbf{U}$. Intuitively, once a basis $\mathbf{U}$ for the known column space is fixed, $\mathbf{M}$ is uniquely represented by its coefficient matrix $\mathbf{R}$ in (4). Since $\mathbf{R}$ contains nr unknown entries, this suggests that the DoF of $\mathbf{M}$ is nr. The following theorem rigorously establishes this intuitive count under complete column subspace information.

Theorem 3.1. *When the column space of a rank-r matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ is known to be $\mathcal{U}$, the* DoF *of* $\mathbf{M}$ *is given by*

$$\mathrm{DoF}\big(\mathbf{M} \mid \mathrm{col}(\mathbf{M}) = \mathcal{U}\big) = nr. \tag{5}$$

Remark 3.1. *As illustrated in Theorem 3.1, the* DoF *of* $\mathbf{M}$ *reduces to nr when its column space is known. Moreover, the difference between the original* DoF *of* $\mathbf{M}$ *and its* DoF *under this condition equals the* DoF *of its column subspace, namely $(m - r)r$ in Proposition 2.1.*

Theorem 3.1 establishes a fundamental relationship between structural prior knowledge and DoF in low-rank matrix recovery. It states that when the column space of a matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ of rank r is known, the number of independent parameters needed to determine $\mathbf{M}$ decreases from $(m+n-r)r$ in (2) to nr in (5), which is consistent with the result in Xu et al. (2013). This reduction shows how prior information can simplify the sample complexity of low-rank matrix recovery.

3.2 Degrees of freedom under partial subspace information

We now relax the assumption of complete column-space knowledge. Suppose that an r_1-dimensional subspace $\mathcal{U}_1 \subseteq \mathrm{col}(\mathbf{M})$ is known, with $r_1 < r$. Representing this known subspace by an orthonormal basis matrix $\mathbf{U}_1 \in \mathbb{R}^{m\times r_1}$, we can decompose the target matrix as

$$\mathbf{M} = \mathbf{U}_1\mathbf{R}_1 + \mathbf{U}_2\mathbf{R}_2, \tag{6}$$

where $\mathbf{U}_2 \in \mathbb{R}^{m\times(r-r_1)}$ spans the unknown orthogonal completion of $\mathcal{U}_1$ within $\mathrm{col}(\mathbf{M})$, so that

$$\mathcal{U}_1 \oplus \mathrm{col}(\mathbf{U}_2) = \mathrm{col}(\mathbf{M}), \tag{7}$$

and $\mathbf{R}_1 \in \mathbb{R}^{r_1\times n}$ and $\mathbf{R}_2 \in \mathbb{R}^{(r-r_1)\times n}$ are the corresponding coefficient matrices. As indicated by (7), knowing $\mathcal{U}_1$ does not fully specify $\mathcal{U} = \mathrm{col}(\mathbf{M})$. We denote the remaining degrees of freedom by $\mathrm{DoF}(\mathcal{U} \mid \mathcal{U}_1 \subseteq \mathcal{U})$. The following lemma quantifies these remaining degrees of freedom, which is a first step before determining the DoF of $\mathbf{M}$.

Lemma 3.1. *For an r-dimensional subspace $\mathcal{U} \subseteq \mathbb{R}^m$, when a partial r_1-dimensional subspace $\mathcal{U}_1 \subseteq \mathcal{U}$ is known, the* DoF *of the remaining subspace is given by*

$$\mathrm{DoF}(\mathcal{U} \mid \mathcal{U}_1 \subseteq \mathcal{U}) = (m-r)(r-r_1). \tag{8}$$

The result in Lemma 3.1 indicates that when a partial subspace $\mathcal{U}_1 \subseteq \mathcal{U}$ is known, the remaining DoF of $\mathcal{U}$ is reduced proportionally to the dimension difference between $\mathcal{U}_1$ and $\mathcal{U}$, i.e., $\mathrm{DoF}(\mathcal{U} \mid \mathcal{U}_1 \subseteq \mathcal{U}) \propto (r-r_1)$. Based on Lemma 3.1, we now present the main result regarding the DoF of a low-rank matrix under partial column-space information, as stated in Theorem 3.2.

Theorem 3.2. *For a rank-r matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$, if an r_1-dimensional subspace $\mathcal{U}_1 \subseteq \mathrm{col}(\mathbf{M})$ is known ($r_1 \le r$), then the* DoF *of $\mathbf{M}$ is*

$$\mathrm{DoF}\big(\mathbf{M} \mid \mathcal{U}_1 \subseteq \mathrm{col}(\mathbf{M})\big) = nr + (m-r)(r-r_1). \tag{9}$$

The two parts in (9) can be understood through a two-step interpretation. First, extending the known subspace $\mathcal{U}_1$ to the full column space $\mathcal{U}$ requires $(m-r)(r-r_1)$ degrees of freedom, as characterized by Lemma 3.1. Once $\mathcal{U}$ is fixed, specifying $\mathbf{M}$ requires nr degrees of freedom, as established in Theorem 3.1. Combining these two steps yields the total $nr + (m-r)(r-r_1)$.

Remark 3.2. *The results in Theorem 3.2 provide a unified framework that interpolates between two extreme cases: when $r_1 = 0$ (no prior information), the* DoF *becomes $(m+n-r)r$, which is the original* DoF *of a rank-r matrix; when $r_1 = r$ (complete column space known), the* DoF *becomes nr, consistent with Theorem 3.1. The additional term $(m-r)(r-r_1)$ quantifies the remaining uncertainty after knowing the partial column space.*

We can further extend Theorem 3.2 to the case where both partial column-space and partial row-space information are available. The corresponding results are in Appendix C.1.

4 Matrix recovery algorithms based on degrees of freedom

Based on the DoF analysis in Section 2 and Section 3, we develop recovery methods that exactly recover matrices with subspace priors using a number of measurements equal to their DoF, thereby establishing the achievability of the DoF bounds.

Following the setting of existing works (Candès & Recht, 2009; Recht et al., 2010; Candès & Plan, 2011), we assume that the measurements of the target matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ are linear. The linear measurements $\mathbf{y} \in \mathbb{R}^K$ are defined as

$$\mathbf{y} = \mathcal{A}(\mathbf{M}), \tag{10}$$

where $\mathcal{A}(\cdot) : \mathbb{R}^{m\times n} \to \mathbb{R}^K$ denotes the measurement operator, and the k-th entry is given by

$$[\mathcal{A}(\mathbf{M})]_k = \langle \mathbf{A}_k, \mathbf{M}\rangle, \tag{11}$$

with $\mathbf{A}_k \in \mathbb{R}^{m\times n}$ the k-th measurement matrix.

4.1 MATRIX RECOVERY WITH COMPLETE COLUMN-SPACE PRIOR

For a rank-r matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ with known complete column space, the DoF analysis in Theorem 3.1 shows that the DoF is nr, which implies that the minimum number of measurements required for matrix recovery is nr. In the following, we show that there exists a method based on the linear measurement model in (10) that achieves exact recovery of $\mathbf{M}$ using exactly nr measurements, matching the theoretical DoF under a complete column-space prior.

Given the complete column-space information, the matrix $\mathbf{M}$ can be decomposed as in (4), i.e., $\mathbf{M} = \mathbf{UR}$. By the property of the Kronecker product,

$$\operatorname{vec}(\mathbf{M}) = \operatorname{vec}(\mathbf{UR}) = (\mathbf{I}_n \otimes \mathbf{U}) \operatorname{vec}(\mathbf{R}). \tag{12}$$

Substituting this into (11), the measurement vector $\mathbf{y}$ can be rewritten as

$$\mathbf{y} = \mathbf{A}(\mathbf{I}_n \otimes \mathbf{U}) \operatorname{vec}(\mathbf{R}), \tag{13}$$

where $\mathbf{A} \in \mathbb{R}^{K\times mn}$ has k-th row $\operatorname{vec}(\mathbf{A}_k)^T$ in (11). For convenience, let $\widetilde{\mathbf{A}} := \mathbf{A}(\mathbf{I}_n \otimes \mathbf{U}) \in \mathbb{R}^{K\times nr}$. Then the problem of recovering $\mathbf{M}$ from $\mathbf{y}$ in (10) is transformed into a low-dimensional vector recovery problem:

$$\min_{\mathbf{R}} \left\| \mathbf{y} - \widetilde{\mathbf{A}} \operatorname{vec}(\mathbf{R}) \right\|_2^2. \tag{14}$$

When $\widetilde{\mathbf{A}}$ has full column rank, $\operatorname{vec}(\mathbf{R})$ can be uniquely solved via least squares (LS):

$$\operatorname{vec}(\mathbf{R}) = (\widetilde{\mathbf{A}}^T \widetilde{\mathbf{A}})^{-1} \widetilde{\mathbf{A}}^T \mathbf{y}. \tag{15}$$

Since $\mathbf{R}$ is correctly recovered, the target matrix is perfectly recovered as $\mathbf{M} = \mathbf{UR}$. We refer to this recovery scheme as the LS algorithm.[1]

In particular, when the entries of $\mathbf{A}_k$ are independent and identically distributed (i.i.d.) Gaussian, $\widetilde{\mathbf{A}}$ has full column rank with probability one whenever $K \geq nr$. For other random measurement designs, the full-column-rank condition is likewise typically satisfied. This shows that, given the complete column-space prior, the LS algorithm achieves exact matrix recovery with the minimum number nr of measurements by solving (14), which matches the theoretical DoF in Theorem 3.1. We formalize this guarantee as follows.

Theorem 4.1 (Exact recovery with complete column-space prior)**.** *For a rank-r matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ with known complete column space of dimension r, when the entries of the measurement matrices $\mathbf{A}_k$ in* (11) *are i.i.d. Gaussian, the LS algorithm in* (14) *recovers $\mathbf{M}$ exactly almost surely using exactly nr measurements, matching the DoF in Theorem 3.1.*

4.2 MATRIX RECOVERY WITH PARTIAL COLUMN SPACE PRIOR

For the case with partial column space information, we provide two recovery algorithms based on different measurement matrices. Specifically, the DORM uses the designed measurement matrices, and the PMR method employs the random measurement matrices.

Let the matrix to be recovered be a rank-r matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$, and $\mathcal{U}$ is the r-dimensional column subspace of $\mathbf{M}$. We assume that an r_1-dimensional subspace $\mathcal{U}_1$ is known, where $\mathcal{U}_1 \subseteq \mathcal{U}$ and $r_1 \leq r$. Let $\mathbf{U}_1 \in \mathbb{R}^{m\times r_1}$ be an orthonormal basis of $\mathcal{U}_1$. Then, the target matrix $\mathbf{M}$ can be expressed as

$$\mathbf{M} = \mathbf{U}_1\mathbf{R}_1 + \mathbf{U}_\perp\mathbf{X}, \tag{16}$$

where $[\mathbf{U}_1, \mathbf{U}_\perp]$ is orthogonal and $\mathbf{U}_\perp \in \mathbb{R}^{m\times(m-r_1)}$, while $\mathbf{X} \in \mathbb{R}^{(m-r_1)\times n}$ has rank $r - r_1$. Equivalently, $\mathbf{X} = \mathbf{VR}$ with $\mathbf{V} \in \mathbb{R}^{(m-r_1)\times(r-r_1)}$ orthonormal and $\mathbf{R} \in \mathbb{R}^{(r-r_1)\times n}$.

4.2.1 DORM VIA DESIGNED MEASUREMENT MATRICES

To achieve exact matrix recovery under a partial column space prior using a number of measurements equal to the DoF, the proposed DORM uses designed measurement matrices $\mathbf{A}_k$ in (11). Specifically, we construct two measurement operators $\mathcal{B}$ and $\mathcal{C}$, where $\mathcal{B}(\cdot) : \mathbb{R}^{m\times n} \to \mathbb{R}^{K_1}$ and $\mathcal{C}(\cdot) : \mathbb{R}^{m\times n} \to \mathbb{R}^{K_2}$. According to (10), the measurements are

$$\mathbf{y}_\mathcal{B} = \mathcal{B}(\mathbf{M}), \qquad \mathbf{y}_\mathcal{C} = \mathcal{C}(\mathbf{M}). \tag{17}$$

[1] In practice, (14) can be solved by a stable LS method; see Appendix D.3 for costs.

and the k-th entry is given by

$$[\mathcal{B}(\mathbf{M})]_k = \langle \mathbf{B}_k, \mathbf{M} \rangle, \qquad [\mathcal{C}(\mathbf{M})]_k = \langle \mathbf{C}_k, \mathbf{M} \rangle. \tag{18}$$

where $\mathbf{B}_k \in \mathbb{R}^{m\times n}$ and $\mathbf{C}_k \in \mathbb{R}^{m\times n}$ denote the k-th measurement matrix in $\mathcal{B}$ and $\mathcal{C}$, respectively.

Before discussing how to recover $\mathbf{M}$ from (17), we make the following assumption on $\mathbf{M}$.

Assumption 4.1. *After removing the component in the known subspace $\mathcal{U}_1$, the first $r - r_1$ columns of $\mathbf{M}$ still contain enough independent information to determine the unknown subspace $\mathcal{U}_2$, that is*

$$\operatorname{rank}\big((\mathbf{I}_m - \mathbf{U}_1\mathbf{U}_1^T)\mathbf{M}_1\big) = r - r_1, \qquad \mathbf{M}_1 := \mathbf{M}_{:,1:r-r_1} \in \mathbb{R}^{m\times(r-r_1)}. \tag{19}$$

When $\mathbf{M}$ is generated randomly, this condition holds with probability one, because a random rank-r matrix almost surely has components outside any fixed proper subspace of its column space. Under Assumption 4.1, DORM proceeds in two stages: first recover the missing subspace $\mathcal{U}_2$ from $\mathbf{y}_\mathcal{B}$ in (17), then recover $\mathbf{M}$ under the resulting complete column-space prior from $\mathbf{y}_\mathcal{C}$.

(i) Estimation of the remaining subspace. We use the operator $\mathcal{B}$ in (17) to observe the first $r - r_1$ columns of $\mathbf{M}$. Concretely, we set $\mathbf{B}_k$ in (18) as

$$\mathbf{B}_k = \mathbf{e}_{i_k}\mathbf{e}_{j_k}^T, \tag{20}$$

where $\mathbf{e}_{i_k}$ and $\mathbf{e}_{j_k}$ are the m- and n-dimensional unit vectors, respectively, indexed for $k = 1, \ldots, m(r-r_1)$ by $i_k = \lfloor (k-1)/(r-r_1) \rfloor + 1$ and $j_k = \big((k-1) \bmod (r-r_1)\big) + 1$. Reshaping $\mathbf{y}_\mathcal{B}$ then yields $\mathbf{M}_1 \in \mathbb{R}^{m\times(r-r_1)}$. To extract $\mathcal{U}_2$ from $\mathbf{M}_1$, we project $\mathbf{M}_1$ onto the orthogonal complement of $\mathcal{U}_1$:

$$\widetilde{\mathbf{M}}_1 = \mathbf{P}\mathbf{M}_1 \in \mathbb{R}^{m\times(r-r_1)}, \tag{21}$$

where $\mathbf{P} = \mathbf{I}_m - \mathbf{U}_1\mathbf{U}_1^T \in \mathbb{R}^{m\times m}$. Then the remaining subspace is given by

$$\mathcal{U}_2 = \operatorname{col}(\widetilde{\mathbf{M}}_1), \qquad \widehat{\mathbf{U}}_2 = \operatorname{orth}(\widetilde{\mathbf{M}}_1) \in \mathbb{R}^{m\times(r-r_1)}. \tag{22}$$

Under Assumption 4.1, $\operatorname{col}(\widetilde{\mathbf{M}}_1) = \mathcal{U}_2$, so together with the known prior we obtain the complete column space $\mathcal{U} = \mathcal{U}_1 \oplus \mathcal{U}_2$. This stage uses $K_1 = m(r - r_1)$ measurements.

(ii) Matrix recovery under the complete column space. With $\mathcal{U}$ in hand, the problem reduces to the LS algorithm in Section 4.1. Concretely, we partition $\mathbf{M}$ as

$$\mathbf{M} = [\mathbf{M}_1 \quad \mathbf{M}_2] = [\mathbf{M}_1 \quad \mathbf{U}\mathbf{S}_2], \tag{23}$$

where $\mathbf{M}_2 \in \mathbb{R}^{m\times(n-r+r_1)}$, $\mathbf{S}_2 \in \mathbb{R}^{r\times(n-r+r_1)}$, and $\mathbf{U} = [\mathbf{U}_1, \widehat{\mathbf{U}}_2] \in \mathbb{R}^{m\times r}$ is the orthonormal basis of $\mathcal{U} = \mathcal{U}_1 \oplus \mathcal{U}_2$ from (22). Since $\mathbf{M}_1$ and $\mathbf{U}$ in (23) are already obtained from the first step, it remains only to estimate $\mathbf{S}_2$.

Specifically, to estimate $\mathbf{S}_2$ from $\mathbf{y}_\mathcal{C}$ in (17), we take the measurement matrices of the form

$$\mathbf{C}_k = [\mathbf{0} \quad \mathbf{G}_k], \tag{24}$$

where $\mathbf{0} \in \mathbb{R}^{m\times(r-r_1)}$ and $\mathbf{G}_k \in \mathbb{R}^{m\times(n-r+r_1)}$ has i.i.d. Gaussian entries. Combining (17), (23) and (24), estimating $\mathbf{S}_2$ reduces to

$$\min_{\mathbf{S}_2} \|\mathbf{y}_\mathcal{C} - \widetilde{\mathbf{C}} \operatorname{vec}(\mathbf{S}_2)\|_2^2, \tag{25}$$

where $\widetilde{\mathbf{C}} := \mathbf{G}(\mathbf{I}_{n-r+r_1} \otimes \mathbf{U}) \in \mathbb{R}^{K_2\times r(n-r+r_1)}$, and $\mathbf{G} \in \mathbb{R}^{K_2\times m(n-r+r_1)}$ has k-th row $\operatorname{vec}(\mathbf{G}_k)^T$. By Theorem 4.1, when the entries of $\mathbf{G}_k$ are i.i.d. Gaussian, (25) recovers $\mathbf{S}_2$ (and hence $\mathbf{M}_2$) almost surely with $K_2 = nr - r(r - r_1)$ measurements. The procedure of DORM is summarized in Algorithm 1, and we formalize the guarantee as follows.

Theorem 4.2 (Exact recovery with partial column space prior)**.** *Let $\mathbf{M} \in \mathbb{R}^{m\times n}$ be a rank-r matrix with column space $\mathcal{U} = \operatorname{col}(\mathbf{M})$, and suppose that an r_1-dimensional subspace $\mathcal{U}_1 \subseteq \mathcal{U}$ is known a priori. Under Assumption 4.1 and with the measurement matrices specified in* (20) *and* (24)*, DORM in Algorithm 1 recovers $\mathbf{M}$ exactly almost surely using*

$$K = nr + (m - r)(r - r_1) \tag{26}$$

measurements, matching the DoF in Theorem 3.2.

Remark 4.1. *In particular, the required measurements of DORM are consistent with DoF in two boundary cases. When $r_1 = r$, the full column space is known, so DORM reduces to step (ii) only, namely matrix recovery under the complete column space, which coincides with the LS algorithm in Section 4.1 and the requirement is therefore nr. When $r_1 = 0$, the requirement becomes $nr + (m - r)r = (m + n - r)r$, which matches the DoF of $\mathbf{M}$ without prior information.*

Algorithm 1 DORM

Input: prior basis $\mathbf{U}_1$; measurements $\mathbf{y}_{\mathcal{B}}$, $\mathbf{y}_{\mathcal{C}}$ in (17)

1: Reshape $\mathbf{y}_{\mathcal{B}}$ to obtain $\mathbf{M}_1 \in \mathbb{R}^{m\times(r-r_1)}$

2: Set $\mathbf{P} = \mathbf{I}_m - \mathbf{U}_1\mathbf{U}_1^T$ and $\widetilde{\mathbf{M}}_1 = \mathbf{P}\mathbf{M}_1$ as in (21)

3: Compute $\widehat{\mathbf{U}}_2 = \mathrm{orth}(\widetilde{\mathbf{M}}_1)$ and set $\mathbf{U} = [\mathbf{U}_1, \widehat{\mathbf{U}}_2]$

4: Solve (25) for $\widehat{\mathbf{S}}_2 \in \mathbb{R}^{r\times(n-r+r_1)}$ and set $\widehat{\mathbf{M}}_2 = \mathbf{U}\widehat{\mathbf{S}}_2 \in \mathbb{R}^{m\times(n-r+r_1)}$

Output: $\widehat{\mathbf{M}} = [\mathbf{M}_1, \widehat{\mathbf{M}}_2]$

4.2.2 PMR VIA RANDOM MEASUREMENT MATRICES

In contrast to DORM in Section 4.2.1, which attains the DoF-optimal measurement count through a designed two-stage operator, we next consider the more common setting of unstructured measurements. Intuitively, PMR uses the partial column-space prior to remove the known-subspace contribution through a projection in measurement space.

Here, we consider Gaussian measurements with $[\mathbf{A}_k]_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/K)$.[2] For a rank-$r$ matrix with a known r_1-dimensional partial column space $\mathcal{U}_1$, we adopt the orthogonal decomposition in (16), $\mathbf{M} = \mathbf{U}_1\mathbf{R}_1 + \mathbf{U}_\perp\mathbf{X}$. We focus on the nontrivial case $0 \le r_1 < r$. Using this decomposition and the linearity of the measurement operator in (10), we express the measurements in matrix form as follows. Let $\mathbf{A} \in \mathbb{R}^{K\times mn}$ have k-th row $\mathrm{vec}(\mathbf{A}_k)^T$, as in (13), and define

$$\mathbf{G}_1 := \mathbf{A}(\mathbf{I}_n \otimes \mathbf{U}_1) \in \mathbb{R}^{K\times nr_1}, \qquad \mathbf{G}_2 := \mathbf{A}(\mathbf{I}_n \otimes \mathbf{U}_\perp) \in \mathbb{R}^{K\times(m-r_1)n}.$$

The measurements in (10) are then given by

$$\mathbf{y} = \mathbf{G}_1\,\mathrm{vec}(\mathbf{R}_1) + \mathbf{G}_2\,\mathrm{vec}(\mathbf{X}). \tag{27}$$

Here $\mathbf{G}_1$ and $\mathbf{G}_2$ are known, while $\mathbf{R}_1$ and $\mathbf{X}$ are unknown. To recover $\mathbf{X}$ independently of $\mathbf{R}_1$, we eliminate the first term by projecting the measurements onto the orthogonal complement of $\mathrm{col}(\mathbf{G}_1)$.

Under Gaussian measurements, $\mathbf{G}_1$ has full column rank almost surely for $K \ge nr_1$. Choose $\mathbf{Q}_2 \in \mathbb{R}^{K\times(K-nr_1)}$ whose columns form an orthonormal basis for the orthogonal complement of $\mathrm{col}(\mathbf{G}_1)$. Since $\mathbf{Q}_2^T\mathbf{G}_1 = \mathbf{0}$, multiplying (27) by $\mathbf{Q}_2^T$ removes the term involving $\mathbf{R}_1$ and yields

$$\widetilde{\mathbf{y}} := \mathbf{Q}_2^T\mathbf{y} = \mathbf{Q}_2^T\mathbf{G}_2\,\mathrm{vec}(\mathbf{X}) =: \mathcal{H}(\mathbf{X}), \tag{28}$$

where $\mathcal{H} : \mathbb{R}^{(m-r_1)\times n} \to \mathbb{R}^{K-nr_1}$ is the effective measurement operator. The projected measurements thus depend only on $\mathbf{X}$, which has rank $r - r_1$ by (16). Recovering $\mathbf{X}$ from (28) is therefore a standard rank-$(r - r_1)$ matrix recovery problem. Meanwhile, the following lemma shows that $\mathcal{H}(\cdot)$ is a Gaussian measurement operator when the entries of $\mathbf{A}_k$ are i.i.d. Gaussian.

Lemma 4.1 (Gaussianity of the effective operator)**.** *Suppose the sensing matrices $\mathbf{A}_k$ in* (11) *have i.i.d. Gaussian entries. Then the matrix representation of the effective operator $\mathcal{H}(\cdot) : \mathbb{R}^{(m-r_1)\times n} \to \mathbb{R}^{K-nr_1}$ in* (28) *also has independent Gaussian entries.*

Therefore, by Lemma 4.1, recovering $\mathbf{X}$ from $\mathcal{H}(\mathbf{X})$ in (28) can be accomplished using existing algorithms such as nuclear-norm minimization (Recht et al., 2010; Candès & Plan, 2011) or alternating least squares (ALS) (Jain et al., 2013; Hardt, 2014). Let $\widehat{\mathbf{X}} \in \mathbb{R}^{(m-r_1)\times n}$ denote the resulting estimate of $\mathbf{X}$. Using this estimate in (27), we recover the coefficients $\mathbf{R}_1$ by least squares:

$$\mathrm{vec}(\widehat{\mathbf{R}}_1) = \mathbf{G}_1^{\dagger}\big(\mathbf{y} - \mathbf{G}_2\,\mathrm{vec}(\widehat{\mathbf{X}})\big), \tag{29}$$

where $\widehat{\mathbf{R}}_1 \in \mathbb{R}^{r_1\times n}$, and the full matrix is assembled as

$$\widehat{\mathbf{M}} = \mathbf{U}_1\widehat{\mathbf{R}}_1 + \mathbf{U}_\perp\widehat{\mathbf{X}} \in \mathbb{R}^{m\times n}. \tag{30}$$

The complete PMR procedure is summarized in Algorithm 2.[3]

[2] The PMR framework also applies to other random measurement ensembles.

[3] When the complete column space is known ($r_1 = r$), the unknown block satisfies $\mathbf{X} = \mathbf{0}$, so PMR reduces to the LS algorithm in Section 4.1.

Algorithm 2 PMR

Input: $\mathbf{U}_1$, Gaussian operator $\mathcal{A}$, observations $\mathbf{y}$, rank r

1: Construct $\mathbf{U}_\perp$, $\mathbf{Q}_2$, $\widetilde{\mathbf{y}}$, and $\mathcal{H}$ from (27)–(28)
2: Recover the rank-$(r-r_1)$ matrix $\widehat{\mathbf{X}}$ from $(\mathcal{H}, \widetilde{\mathbf{y}})$
3: Recover $\mathrm{vec}(\widehat{\mathbf{R}}_1) = \mathbf{G}_1^{\dagger}(\mathbf{y} - \mathbf{G}_2\,\mathrm{vec}(\widehat{\mathbf{X}}))$ in (29)

Output: $\widehat{\mathbf{M}} = \mathbf{U}_1\widehat{\mathbf{R}}_1 + \mathbf{U}_\perp\widehat{\mathbf{X}}$

The following theorem establishes an exact recovery guarantee for PMR under i.i.d. Gaussian measurements and characterizes the required number of measurements.

Theorem 4.3 (Recovery guarantee for PMR under Gaussian sensing)**.** *Let $\mathbf{M} \in \mathbb{R}^{m\times n}$ be a rank-r matrix with a known partial column space of dimension r_1, where $0 \le r_1 < r$. Under i.i.d. $\mathcal{N}(0, 1/K)$ sensing in* (10)*, PMR in Algorithm 2 recovers $\mathbf{M}$ exactly with high probability using*

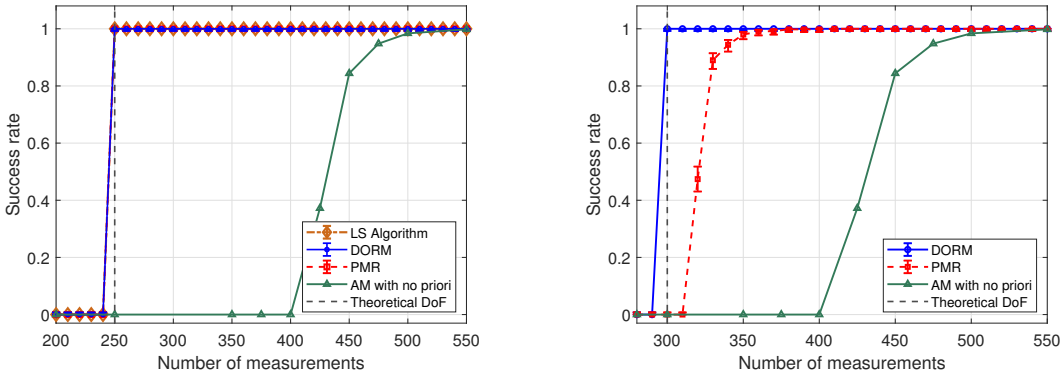
$$K = nr_1 + \mathcal{O}((r-r_1)(m+n-r)) \tag{31}$$

measurements, matching the DoF in Theorem 3.2 up to a constant factor.

In Theorem 4.3, the first term nr_1 accounts for the measurements needed to recover $\mathbf{R}_1$, while the second term $\mathcal{O}((r-r_1)(m+n-r))$ accounts for those needed to recover $\mathbf{X}$. Compared with the theoretical count $\mathrm{DoF}(\mathbf{M} \mid \mathcal{U}_1)$ in Theorem 3.2, the constant-factor overhead in the second term comes from the existing recovery guarantees for the rank-$(r-r_1)$ matrix $\mathbf{X}$.

5 Experiments

We numerically compare the measurement requirements of the LS, DORM, and PMR algorithms with the DoF analysis in Section 3. As a baseline, we also evaluate AM with no prior (Jain et al., 2013). In particular, for PMR, alternating minimization (Jain et al., 2013) is used to recover the low-rank component $\mathbf{X}$. In the synthetic experiments, the target matrix $\mathbf{M} \in \mathbb{R}^{30\times 50}$ is generated randomly with rank $r = 5$. A recovery is considered successful if $\|\widehat{\mathbf{M}} - \mathbf{M}\|_F^2 / \|\mathbf{M}\|_F^2 < 10^{-6}$.



(a) Complete column-space prior ($r_1 = r = 5$).

(b) Partial column space prior ($r_1 = 3$).

Figure 1: Recovery success rate versus measurement count for random rank-5 matrices in $\mathbb{R}^{30\times 50}$ with (a) complete ($r_1 = 5$) and (b) partial ($r_1 = 3$) column-space priors. Vertical dashed lines indicate theoretical DoF; error bars show 95% confidence intervals over 500 trials.

Under complete column subspace information ($r_1 = r = 5$), Figure 1(a) compares the LS algorithm, DORM, and PMR against AM with no prior under i.i.d. Gaussian measurements. We report the success rate over 500 independent trials for each K ranging from 200 to 550. The theoretical DoF in this setting is $nr = 250$, as given in Theorem 3.1. As analyzed in Section 4.2, DORM and PMR both reduce to the LS algorithm when the complete column space is known, explaining their coincident curves. As illustrated in Figure 1(a), all these three methods achieve 100% success at exactly 250 measurements, matching the theoretical DoF in Theorem 3.1, whereas AM with no prior requires substantially more measurements.

Under partial column-subspace information ($r_1 = 3$), Figure 1(b) uses the same success criterion and 500 trials per measurement count. DORM employs the designed two-stage measurement operator in Section 4.2.1, whereas PMR and AM with no prior use i.i.d. Gaussian measurements. The DoF of $\mathbf{M}$ is $nr + (m - r)(r - r_1) = 300$, as given in Theorem 3.2. One can observe in Figure 1(b) that DORM achieves a 100% success rate with exactly 300 measurements, matching the theoretical DoF and Theorem 4.2. In contrast, PMR under Gaussian sensing requires slightly more measurements and attains near-100% success around $K = 350$, consistent with Theorem 4.3. Without prior information, AM attains near-100% success around $K = 500$. These results demonstrate that the partial column space prior reduces the number of measurements required for recovery.

Real-world data. We evaluate DORM and PMR on Jester recommendation data (Goldberg et al., 2001) and USC massive-MIMO data (Choi et al., 2022), using leading left singular vectors of historical observations as subspace priors. The target matrices have sizes 100×1000 and 500×1000, with $(r, r_1) = (15, 10)$ and $(20, 15)$, respectively. We use observed ratings for Jester and linear channel measurements for USC, with equal scalar measurement budgets across methods. Table 1 reports mean absolute error (MAE) for Jester and normalized mean squared error (NMSE) for USC, averaged over ten trials. Both methods consistently outperform AM with no prior at every tested budget, with DORM achieving the lowest errors. These results demonstrate the practical value of historical subspace priors; full protocols and results are provided in Appendix D.4.

Table 1: Recovery errors on real-world data, averaged over ten trials (lower is better).

	Jester (MAE)			USC massive MIMO (NMSE)		
K/DoF	DORM	PMR	AM with no prior	DORM	PMR	AM with no prior
120%	**4.765**	4.953	5.405	**0.0925**	0.1333	1.0563
160%	**3.738**	3.903	4.259	**0.0355**	0.0504	0.3976
200%	**3.338**	3.442	3.688	**0.0263**	0.0315	0.1165

Additional experiments in Appendix D evaluate PMR under different measurement designs and compare its performance with DORM. Meanwhile, we analyze the recovery performance of the proposed methods under measurement noise in Appendix C.2 and report experimental results in Appendix D. In addition, Appendix D.3 provides computational complexity analysis of the proposed methods and runtime experiments for large-scale matrix recovery.

6 Conclusion

In this paper, we quantify how subspace prior information reduces the measurements needed for low-rank matrix recovery and establish when these savings are achievable. For a rank-r matrix in $\mathbb{R}^{m\times n}$ with a known r_1-dimensional subspace of its column space, we characterize the DoF as $nr + (m - r)(r - r_1)$. Moreover, DORM establishes the achievability of this DoF bound under the designed two-stage measurements. For general linear measurements, PMR uses subspace priors to reduce the problem to standard reduced-rank matrix recovery. Under Gaussian measurements, it achieves exact recovery with high probability using measurements on the order of the DoF. Synthetic and real-world experiments support the predicted measurement savings of the subspace priors. Future work will address approximate priors and sharper recovery guarantees under noise.

Artificial intelligence use statement

Generative AI was used to assist with language polishing, including correcting grammatical errors and improving the clarity and readability of the manuscript, as well as to provide coding assistance for the experiments. The authors carefully reviewed all AI-assisted text to ensure that it accurately reflected the intended meaning and checked the AI-assisted code for correctness. The authors take full responsibility for the final content of the paper.

Reproducibility statement

Complete proofs of the theoretical results presented in this paper are provided in the appendix. The appendix also provides a detailed description of the experimental setup, including data preprocessing, measurement allocation, stopping rules, and evaluation metrics. The complete code will be made publicly available after the review process concludes.

References

H. S. F. Ardakani, S. Daei, A. Amini, M. Skoglund, and G. Fodor. Subspace-informed matrix completion, 2024. arXiv:2405.07890 [eess].

Siavash Bayat and Sajad Daei. An optimal hybrid nuclear norm regularization for matrix sensing with subspace prior information. *IEEE Access*, 8:130937–130946, 2020.

Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

Jian-Feng Cai, Tong Wu, and Ruizhe Xia. Fast non-convex matrix sensing with optimal sample complexity. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pp. 497–520. PMLR, 2025.

E. J. Candes and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010.

Emmanuel J Candès and Yaniv Plan. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory*, 57(4):2342–2359, 2011. doi: 10.1109/TIT.2011.2111771.

Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

Gengshuo Chang, Wei Zhang, and Lehan Zhang. Matrix completion with incomplete side information via orthogonal complement projection. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 7393–7416. PMLR, 2025.

Kai-Yang Chiang, Cho-Jui Hsieh, and Inderjit S. Dhillon. Matrix completion with noisy side information. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

Thomas Choi, Jorge Gomez-Ponce, Colton Bullard, Issei Kanno, Masaaki Ito, Takeo Ohseki, Kosuke Yamazaki, and Andreas F. Molisch. Using a drone sounder to measure channels for cell-free massive MIMO systems. In *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2506–2511. IEEE, 2022. doi: 10.1109/WCNC51071.2022.9771649.

A. Eftekhari, D. Yang, and M. B. Wakin. Weighted matrix completion and recovery with prior subspace information. *IEEE Transactions on Information Theory*, 64(6):4044–4071, 2018.

Ken Goldberg, Theresa Roeder, Dhruv Gupta, and Chris Perkins. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4:133–151, 2001. doi: 10.1023/A:1011419012209.

Moritz Hardt. Understanding alternating minimization for matrix completion. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 651–660. IEEE, 2014.

Prateek Jain, Raghu Meka, and Inderjit S. Dhillon. Guaranteed rank minimization via singular value projection. In *Advances in Neural Information Processing Systems*, volume 23, 2010.

Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 665–674. ACM, 2013.

O. López. Near-optimal weighted matrix completion. *J. Mach. Learn. Res.*, 24(283):1–40, 2023.

J. Miao and K. I. Kou. Color image recovery using low-rank quaternion matrix completion algorithm. *IEEE Transactions on Image Processing*, 31:190–201, 2022.

Nagarajan Natarajan and Inderjit S. Dhillon. Inductive matrix completion for predicting gene-disease associations. *Bioinformatics*, 30(12):i60–i68, 2014.

L. T. Nguyen, J. Kim, and B. Shim. Low-rank matrix completion: A contemporary survey. *IEEE Access*, 7: 94215–94237, 2019.

A. Radhakrishnan, M. Belkin, and D. Drusvyatskiy. Linear recursive feature machines provably recover low-rank matrices. *Proceedings of the National Academy of Sciences*, 122(13):e2411325122, 2025. doi: 10.1073/pnas.2411325122.

A. Ramlatchan, M. Yang, Q. Liu, M. Li, J. Wang, and Y. Li. A survey of matrix completion methods for recommendation systems. *Big Data Mining and Analytics*, 1(4):308–323, 2018.

X. Rao and V. K. N. Lau. Compressive sensing with prior support quality information and application to massive MIMO channel estimation with temporal correlation. *IEEE Transactions on Signal Processing*, 63(18): 4914–4924, 2015.

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.

Anthony Man-Cho So and Yinyu Ye. Theory of semidefinite programming for sensor network localization. *Mathematical Programming*, 109(2-3):367–384, 2007.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2018. doi: 10.1017/9781108231596.

Per-Ake Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12(1):99–111, 1972. doi: 10.1007/BF01932678.

Yuelong Xia, Mingjing Tang, and Pei Wang. Bi-directional matrix completion for highly incomplete multi-label learning via co-embedding predictive side information. *Applied Intelligence*, 53(23):28074–28098, 2023.

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 (NeurIPS 2013)*, volume 26. Curran Associates, Inc., 2013.

Z. Yang, S. Ma, Y. Liu, H. Zhang, and X. Lyu. DOA estimation of coherent sources via low-rank matrix decomposition. *IEEE Wireless Communications Letters*, 13(11):3049–3053, 2024.

Feipeng Zhao and Yuhong Guo. Learning discriminative recommendation systems with side information. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, volume 2017, pp. 3469–3475, 2017.

A Related Work

Firstly, Candès & Recht (2009) provided the rigorous theoretical guarantee for the low-rank matrix completion problem, proposing a convex relaxation method that uses the nuclear norm as a matrix rank function in (1). It was shown that if the number of measurements K obeys

$$K \geq Cn^{1.2} r \log n, \tag{32}$$

then, with very high probability, most $n \times n$ matrices of rank r can be perfectly recovered. Recht et al. (2010) generalized this result to $m \times n$ matrices, establishing an upper bound on the measurement complexity of

$$\mathcal{O}\big(r(m+n)\log(mn)\big). \tag{33}$$

In addition, Candes & Plan (2010) extended this noise-free framework to accommodate noisy measurements, with measurement complexity

$$\mathcal{O}\big(r(m+n)\log^2(mn)\big) \tag{34}$$

for robust matrix completion.

Beyond this setting without prior information, matrix completion with side information has been extensively studied. Leveraging prior knowledge about the row and column subspaces of the target matrix can reduce measurement complexity and improve recovery accuracy. In early research (Xu et al., 2013), the prior information is represented as subspace information and constructed as two matrices $\mathbf{A}$ and $\mathbf{B}$ that

$$\begin{aligned} \mathrm{col}(\mathbf{M}) &\subseteq \mathrm{col}(\mathbf{A}), \\ \mathrm{col}(\mathbf{M}^\top) &\subseteq \mathrm{col}(\mathbf{B}). \end{aligned} \tag{35}$$

That work proposed an inductive matrix completion (IMC) framework, and the problem can be formulated as

$$\min_{\mathbf{X} \in \mathbb{R}^{r_a \times r_b}} ||\mathbf{X}||_{\mathrm{tr}} \quad \text{s.t.} \quad \mathcal{R}_\Omega(\mathbf{A}\mathbf{X}\mathbf{B}^\top) = \mathcal{R}_\Omega(\mathbf{M}). \tag{36}$$

Under this strong assumption (35), they showed that the required number of measurements reduces from $\mathcal{O}\big(r(m+n)\ln^2 n\big)$ to

$$\mathcal{O}\big(r(r_a+r_b)\ln n\big), \tag{37}$$

where r_a, r_b are the dimensions of the side information. However, real-world side information is rarely perfect. Chiang et al. (2015) addressed this limitation by introducing a "dirty" statistical model that decomposes the target matrix into a feature-predictable component and a residual term recovered purely from observations: $\mathbf{M} = \mathbf{A}\mathbf{Z}\mathbf{B}^\top + \mathbf{N}$. Their approach provides theoretical guarantees that measurement complexity can still be reduced whenever the side information is sufficiently informative, even if noisy.

Subsequent work (Eftekhari et al., 2018; Bayat & Daei, 2020; López, 2023) adopted a more geometric perspective by modeling prior knowledge as approximate subspaces $\tilde{\mathcal{U}}$, $\tilde{\mathcal{V}}$ close to the true column and row spaces. Eftekhari et al. (2018) introduced principal angles between the true and prior subspaces and constructed two weighted projection operators

$$\mathbf{Q}_{\widetilde{\mathcal{U}}_r, w} := w \cdot \mathbf{P}_{\widetilde{\mathcal{U}}_r} + \mathbf{P}_{\widetilde{\mathcal{U}}_r^\perp} \in \mathbb{R}^{m \times m}, \tag{38}$$

where $w \in (0, 1]$, and $\mathbf{P}_{\widetilde{\mathcal{U}}_r}, \mathbf{P}_{\widetilde{\mathcal{U}}_r^\perp} \in \mathbb{R}^{m \times m}$ are the orthogonal projections onto the prior column subspace $\widetilde{\mathcal{U}}_r \subseteq \mathbb{R}^m$ and its orthogonal complement, respectively. Likewise, define $\mathbf{Q}_{\widetilde{\mathcal{V}}_r, w} \in \mathbb{R}^{n \times n}$ and the problem can be modeled as

$$\min_{\mathbf{X} \in \mathbb{R}^{m \times n}} \left\| \mathbf{Q}_{\widetilde{\mathcal{U}}_r, w} \mathbf{X} \mathbf{Q}_{\widetilde{\mathcal{V}}_r, w} \right\|_* \quad \text{s.t.} \quad \mathcal{R}_\Omega(\mathbf{X}) = \mathcal{R}_\Omega(\mathbf{M}). \tag{39}$$

For square matrices, sufficiently accurate subspace priors and suitable weights can reduce the sample complexity to $\mathcal{O}(rn \log n)$, assuming bounded coherence parameters. For matrix sensing, Bayat & Daei (2020) proposed a weighted hybrid nuclear norm over four subspace blocks, with weights chosen through a statistical-dimension analysis of reconstruction error.

Furthermore, López (2023) unifies prior subspace information into a single matrix subspace $\mathbf{T}$ and proposes a general weighted matrix completion framework. In the real-valued setting considered here, $\mathbf{T} \subseteq \mathbb{R}^{m\times n}$, and the estimator is

$$\begin{aligned} \widehat{\mathbf{M}}^{\omega} \in & \underset{\mathbf{X}\in\mathbb{R}^{m\times n}}{\arg\min} \left\|\omega\mathcal{P}_{\mathbf{T}}(\mathbf{X}) + \mathcal{P}_{\mathbf{T}^{\perp}}(\mathbf{X})\right\|_{*} \\ & \text{s.t.} \quad \|P_{\Omega}(\mathbf{X}) - \mathbf{y}_{\Omega}\|_2 \le \eta. \end{aligned} \tag{40}$$

where $\omega \in [0,1]$, and $\mathcal{P}_{\mathbf{T}}$ and $\mathcal{P}_{\mathbf{T}^{\perp}}$ are the orthogonal projections onto $\mathbf{T}$ and its orthogonal complement, respectively. Here, $P_{\Omega} : \mathbb{R}^{m\times n} \to \mathbb{R}^{|\Omega|}$ extracts the entries indexed by Ω into a vector. The observations satisfy $\mathbf{y}_{\Omega} = P_{\Omega}(\mathbf{M}) + \mathbf{e}_{\Omega}$, where $\mathbf{e}_{\Omega} \in \mathbb{R}^{|\Omega|}$ is measurement noise with $\|\mathbf{e}_{\Omega}\|_2 \le \eta$. Define $N := \max\{m,n\}$ and

$$\rho := \max_{\mathbf{X}\in\mathbf{T}} \operatorname{rank}(\mathbf{X}), \qquad M_1 := \frac{mn}{\rho} \max_{\substack{\mathbf{X}\in\mathbf{T} \\ \|\mathbf{X}\|_F\le 1}} \max_{i,j} |X_{ij}|^2, \tag{41}$$

where M_1 is the joint incoherence parameter. Under the sampling and weight conditions of the theorem, a sample size of order $\mathcal{O}(M_1\rho\log^3 N)$ suffices for the recovery guarantees. The recovery error depends on measurement noise, low-rank approximation error, and prior-subspace mismatch.

Although these methods accommodate imperfect priors, their recovery guarantees depend on the accuracy and structure of the supplied subspace information.

To address the realistic setting where the side information is incomplete or only partially informative, Chang et al. (2025) propose the Orthogonal Complement Matrix Completion (OCMC) model. Their key insight is that both the component of $\mathbf{M}$ lying in the subspace spanned by $\mathbf{A}$ and $\mathbf{B}$, and its orthogonal complement, exhibit low-rank structure. Accordingly, they regularize both parts via a hybrid nuclear norm objective:

$$\begin{aligned} & \min_{\mathbf{X}} \|\mathbf{X}\|_* + \lambda \left\|\mathrm{P}_{\mathbf{A}^{\perp}\mathbf{B}^{\perp}}(\mathbf{X})\right\|_* \\ & \text{s.t.} \quad \mathcal{R}_{\Omega}(\mathbf{X}) = \mathcal{R}_{\Omega}(\mathbf{M}), \end{aligned} \tag{42}$$

where $\mathrm{P}_{\mathbf{A}^{\perp}\mathbf{B}^{\perp}}$ denotes the projection onto the orthogonal complement of the prior subspaces. This formulation enables robust recovery even when the residual norm $\|\mathrm{P}_{\mathbf{A}^{\perp}\mathbf{B}^{\perp}}(\mathbf{M})\|_F$ is large—a common scenario in real-world datasets such as MovieLens. Empirically, OCMC achieves accurate completion with only 25–40% measurements, depending on the completeness of the prior matrices, offering a practical and theoretically grounded solution for imperfect side information.

In summary, the works above show that subspace or feature-space priors can reduce the number of measurements relative to standard low-rank recovery. Their sample complexities, however, depend on the recovery method and on how the prior is modeled, which may be an exact enclosing subspace, an approximate subspace, or only partial side information. They do not characterize the intrinsic degrees of freedom of a rank-r matrix under a column-subspace prior, nor whether this DoF lower bound for the number of measurements is attainable. We address this gap by deriving an algorithm-independent DoF formula under complete and partial column-space priors. In addition, we develop recovery methods that achieve, or nearly achieve, the corresponding measurement bound.

B Proofs of Main Results

This section provides proofs of the DoF formulas and recovery guarantees stated in the main text.

B.1 Proof of Proposition 2.1

Proof. Any r-dimensional subspace $\mathcal{U} \subset \mathbb{R}^m$ can be represented as the column space of a full column-rank matrix

$$\mathbf{U} \in \mathbb{R}^{m\times r}, \qquad \operatorname{rank}(\mathbf{U}) = r,$$

that is, $\mathcal{U} = \operatorname{col}(\mathbf{U})$. Such a representation is not unique: for any invertible matrix

$$\mathbf{R} \in \mathrm{GL}(r) \triangleq \{\mathbf{R} \in \mathbb{R}^{r\times r} : \det(\mathbf{R}) \neq 0\},$$

it holds that

$$\mathrm{col}(\mathbf{U}) = \mathrm{col}(\mathbf{UR}).$$

Hence, all matrices representing the same subspace form an equivalence class

$$[\mathbf{U}] = \{\mathbf{UR} : \mathbf{R} \in \mathrm{GL}(r)\}.$$

As a result, the Grassmann manifold can be identified with the quotient space

$$\mathrm{Gr}(r, m) = \{\mathbf{U} \in \mathbb{R}^{m\times r} : \mathrm{rank}(\mathbf{U}) = r\}/\mathrm{GL}(r).$$

The set $\{\mathbf{U} \in \mathbb{R}^{m\times r} : \mathrm{rank}(\mathbf{U}) = r\}$ is an open subset of $\mathbb{R}^{m\times r}$ and therefore has dimension mr. The general linear group $\mathrm{GL}(r)$ is an open subset of $\mathbb{R}^{r\times r}$ and has dimension r^2. Since the right action of $\mathrm{GL}(r)$ on the set of full column-rank matrices is smooth and free, the quotient $\mathrm{Gr}(r, m)$ is a smooth manifold whose dimension equals the difference between the dimension of the numerator space and that of the equivalence group, namely,

$$\dim \mathrm{Gr}(r, m) = mr - r^2 = r(m - r).$$

Therefore, for any $\mathcal{U} \in \mathrm{Gr}(r, m)$, the DoF are given by

$$\mathrm{DoF}(\mathcal{U}) = r(m - r),$$

which concludes the proof. □

B.2 Proof of Theorem 3.1

We provide two methods to establish the result in Theorem 3.1.

Proof. Method 1: Let $\mathbf{M} \in \mathbb{R}^{m\times n}$ be a matrix with $\mathrm{rank}(\mathbf{M}) = r$. Denote its column space by

$$\mathcal{U} \triangleq \mathrm{col}(\mathbf{M}) \subset \mathbb{R}^m,$$

and assume that this r-dimensional subspace $\mathcal{U}$ is known. Let $\mathbf{U} \in \mathbb{R}^{m\times r}$ be a known basis matrix of $\mathcal{U}$ (so $\mathrm{col}(\mathbf{U}) = \mathcal{U}$ and $\mathbf{U}$ has full column rank).

We write $\mathbf{M} = (\mathbf{m}_1, \ldots, \mathbf{m}_n)$ to denote the columns of $\mathbf{M}$. Since each column $\mathbf{m}_i \in \mathcal{U} = \mathrm{col}(\mathbf{U})$, there exists a coefficient vector $\mathbf{k}_i \in \mathbb{R}^r$ such that

$$\mathbf{m}_i = \mathbf{U}\mathbf{k}_i, \qquad i = 1, \ldots, n.$$

Stacking these coefficient vectors into $\mathbf{R} \triangleq (\mathbf{k}_1, \ldots, \mathbf{k}_n) \in \mathbb{R}^{r\times n}$, we obtain the factorization

$$\mathbf{M} = \mathbf{UR}.$$

Because $\mathbf{U}$ is known, specifying $\mathbf{M}$ is equivalent to specifying $\mathbf{R}$. The matrix $\mathbf{R}$ contains nr scalar entries, hence contributes nr DoF. Therefore, the DoF of $\mathbf{M}$ under the known column-space prior is nr.

Method 2: According to (2), the DoF of the set of rank-r matrices in $\mathbb{R}^{m\times n}$ is

$$\mathrm{DoF}(\mathbf{M}) = r(m + n - r).$$

Meanwhile, according to Proposition 2.1, the DoF of an r-dimensional subspace $\mathcal{U} \subset \mathbb{R}^m$ (i.e., the Grassmann manifold $\mathrm{Gr}(r, m)$) is

$$\mathrm{DoF}(\mathcal{U}) = r(m - r).$$

When $\mathrm{col}(\mathbf{M}) = \mathcal{U}$ is known, the uncertainty associated with choosing the column subspace is removed. Hence the remaining DoF equals the total rank-r DoF minus the DoF of the column subspace:

$$\begin{aligned}\mathrm{DoF}(\mathbf{M} \mid \mathrm{col}(\mathbf{M}) = \mathcal{U}) &= \mathrm{DoF}(\mathbf{M}) - \mathrm{DoF}(\mathcal{U}) \\ &= r(m + n - r) - r(m - r) \\ &= nr.\end{aligned}$$

This proves the claim. □

B.3 Proof of Lemma 3.1

Proof. Since $\mathcal{U}$ must contain the known subspace $\mathcal{U}_1$, determining $\mathcal{U}$ is equivalent to selecting a complementary subspace of dimension $r - r_1$ within the orthogonal complement $\mathcal{U}_1^\perp$.

The dimension of the orthogonal complement $\mathcal{U}_1^\perp$ is $m - r_1$. Thus, the problem reduces to choosing an $(r - r_1)$-dimensional subspace in an $(m - r_1)$-dimensional space. The DoF are given by the dimension of the Grassmannian $\mathrm{Gr}(r - r_1, m - r_1)$:

$$\begin{aligned}\mathrm{DoF}(\mathcal{U} \mid \mathcal{U}_1 \subseteq \mathcal{U}) &= \dim \mathrm{Gr}(r - r_1, m - r_1) \\ &= (r - r_1)\left[(m - r_1) - (r - r_1)\right] \\ &= (r - r_1)(m - r). \end{aligned} \tag{43}$$

This concludes the proof. □

B.4 Proof of Theorem 3.2

Proof. We combine Theorem 3.1 and Lemma 3.1. Fix an r_1-dimensional subspace $\mathcal{U}_1$ and consider rank-r matrices $\mathbf{M}$ with $\mathcal{U}_1 \subseteq \mathrm{col}(\mathbf{M})$. Specifying such a matrix requires first determining the full column space $\mathcal{U} = \mathrm{col}(\mathbf{M})$, and then determining the entries of $\mathbf{M}$ within $\mathcal{U}$.

Since $\mathcal{U}_1$ is fixed and must be contained in $\mathcal{U}$, the DoF required to specify $\mathcal{U}$ are given in Lemma 3.1,

$$\mathrm{DoF}(\mathcal{U} \mid \mathcal{U}_1 \subseteq \mathcal{U}) = (m - r)(r - r_1). \tag{44}$$

Once the subspace $\mathcal{U}$ is fixed, the DoF of $\mathbf{M}$ is given based on Theorem 3.1:

$$\mathrm{DoF}(\mathbf{M} \mid \mathrm{col}(\mathbf{M}) = \mathcal{U}) = nr. \tag{45}$$

Then, by the chain rule of DoF (additivity of manifold dimensions), the total DoF count is the sum of the costs of these two steps:

$$\begin{aligned}\mathrm{DoF}\left(\mathbf{M} \mid \mathcal{U}_1 \subseteq \mathrm{col}(\mathbf{M})\right) &= \mathrm{DoF}(\mathbf{M} \mid \mathrm{col}(\mathbf{M}) = \mathcal{U}) + \mathrm{DoF}(\mathcal{U} \mid \mathcal{U}_1 \subseteq \mathcal{U}) \\ &= nr + (m - r)(r - r_1). \end{aligned} \tag{46}$$

This completes the proof.

□

B.5 Proof of Theorem 4.1

Proof. Let $\mathbf{U} \in \mathbb{R}^{m \times r}$ be the known orthonormal basis and write $\mathbf{M} = \mathbf{U}\mathbf{R}$. Vectorization gives

$$\mathrm{vec}(\mathbf{M}) = (\mathbf{I}_n \otimes \mathbf{U})\,\mathrm{vec}(\mathbf{R}).$$

The columns of $\mathbf{I}_n \otimes \mathbf{U}$ are orthonormal because

$$(\mathbf{I}_n \otimes \mathbf{U})^T(\mathbf{I}_n \otimes \mathbf{U}) = \mathbf{I}_n \otimes (\mathbf{U}^T\mathbf{U}) = \mathbf{I}_{nr}.$$

Let the rows of $\mathbf{A} \in \mathbb{R}^{K \times mn}$ be independent $\mathcal{N}(\mathbf{0}, \mathbf{I}_{mn}/K)$ vectors. By Gaussian rotational invariance, the rows of

$$\widetilde{\mathbf{A}} := \mathbf{A}(\mathbf{I}_n \otimes \mathbf{U})$$

are independent $\mathcal{N}(\mathbf{0}, \mathbf{I}_{nr}/K)$ vectors. At $K = nr$, $\widetilde{\mathbf{A}}$ is therefore a square matrix with an absolutely continuous Gaussian distribution. Thus $\widetilde{\mathbf{A}}$ is nonsingular almost surely.

Therefore, when $K = nr$, the reduced system $\mathbf{y} = \widetilde{\mathbf{A}}\,\mathrm{vec}(\mathbf{R})$ admits the unique solution $\mathrm{vec}(\mathbf{R}) = \widetilde{\mathbf{A}}^{-1}\mathbf{y}$ almost surely, and hence $\mathbf{M} = \mathbf{U}\mathbf{R}$ is recovered exactly. Since this measurement count matches the DoF nr in Theorem 3.1, the LS algorithm is DoF-optimal. □

B.6 Proof of Theorem 4.2

Proof. Let $q = r - r_1$. Stage 1 uses $K_1 = mq$ canonical measurements to recover $\mathbf{M}_1 = \mathbf{M}_{:,1:q}$ exactly. By Assumption 4.1, $\mathrm{col}((\mathbf{I}_m - \mathbf{U}_1\mathbf{U}_1^T)\mathbf{M}_1)$ is precisely the q-dimensional missing subspace. Hence $\widehat{\mathbf{U}}_2$ in (22) is exact, and $\mathbf{U} = [\mathbf{U}_1, \widehat{\mathbf{U}}_2]$ spans $\mathrm{col}(\mathbf{M})$.

It remains to recover the coefficient matrix of the unobserved block $\mathbf{M}_2 = \mathbf{U}\mathbf{S}_2$. In (25), the reduced sensing matrix $\widetilde{\mathbf{C}} = \mathbf{G}(\mathbf{I}_{n-q} \otimes \mathbf{U})$ is Gaussian because $\mathbf{U}$ is orthonormal. With $K_2 = r(n-q)$, it is square and nonsingular almost surely; equivalently, this is the complete-column-space case established in Theorem 4.1. Thus Stage 2 recovers $\mathbf{S}_2$, and hence $\mathbf{M}_2$, exactly. Together with the observed block $\mathbf{M}_1$, DORM recovers $\mathbf{M}$ exactly. Finally,

$$K = K_1 + K_2 = mq + r(n-q) = nr + (m-r)q = nr + (m-r)(r-r_1).$$

This equals the DoF in Theorem 3.2. □

B.7 Proof of Lemma 4.1

Proof. By definition, the two measurement blocks satisfy

$$\mathbf{G}_1 = \mathbf{A}(\mathbf{I}_n \otimes \mathbf{U}_1), \qquad \mathbf{G}_2 = \mathbf{A}(\mathbf{I}_n \otimes \mathbf{U}_\perp).$$

Since $[\mathbf{U}_1, \mathbf{U}_\perp]$ is orthogonal, Gaussian rotational invariance implies that $[\mathbf{G}_1, \mathbf{G}_2]$ has i.i.d. $\mathcal{N}(0, 1/K)$ entries. In particular, $\mathbf{G}_1$ and $\mathbf{G}_2$ are independent. Choose $\mathbf{Q}_2$ using only $\mathbf{G}_1$, so it is also independent of $\mathbf{G}_2$. For each column $\mathbf{g}_j$ of $\mathbf{G}_2$, conditioning on $\mathbf{G}_1$ gives

$$\mathbf{Q}_2^T\mathbf{g}_j \mid \mathbf{G}_1 \sim \mathcal{N}\left(\mathbf{0}, \frac{1}{K}\mathbf{Q}_2^T\mathbf{Q}_2\right) = \mathcal{N}\left(\mathbf{0}, \frac{1}{K}\mathbf{I}_{K-nr_1}\right).$$

The projected columns remain conditionally independent, and the diagonal covariance implies independence of their entries. This joint conditional distribution does not depend on $\mathbf{G}_1$, so $\mathbf{Q}_2^T\mathbf{G}_2$ also has i.i.d. $\mathcal{N}(0, 1/K)$ entries unconditionally. Since

$$\mathcal{H}(\mathbf{X}) = \mathbf{Q}_2^T\mathbf{G}_2\,\mathrm{vec}(\mathbf{X}),$$

this proves that the matrix representation of $\mathcal{H}$ has independent Gaussian entries. □

B.8 Proof of Theorem 4.3

Proof. By Lemma 4.1, the effective operator $\mathcal{H}$ is Gaussian. We analyze the recovery of the two unknown blocks separately.

Step 1: Recovery of X. Since $\mathbf{G}_1 \in \mathbb{R}^{K \times nr_1}$ has full column rank almost surely, the projection onto its orthogonal complement leaves $K - nr_1$ effective measurements. By Lemma 4.1, these are standard Gaussian measurements of the $(m - r_1) \times n$ rank-$(r - r_1)$ matrix $\mathbf{X}$. Thus recovering $\mathbf{X}$ is the standard rank-constrained matrix-sensing problem

$$\min_{\mathbf{X} \in \mathbb{R}^{(m-r_1)\times n}} \frac{1}{2}\|\widetilde{\mathbf{y}} - \mathcal{H}(\mathbf{X})\|_2^2 \quad \text{subject to} \quad \mathrm{rank}(\mathbf{X}) \le r - r_1.$$

Classical matrix-sensing solvers—including nuclear-norm minimization, singular value projection, and Riemannian gradient descent—provide recovery guarantees with the common effective sample complexity (Recht et al., 2010; Candès & Plan, 2011; Jain et al., 2010; Cai et al., 2025)

$$K - nr_1 = \mathcal{O}((r - r_1)(m + n - r_1)).$$

Since $r - r_1 \le \min\{m - r_1, n\}$, this is equivalently $K - nr_1 = \mathcal{O}((r - r_1)(m + n - r))$.

Step 2: Recovery of $\mathbf{R}_1$. Once $\mathbf{X}$ is known, subtracting its contribution from the original measurements gives

$$\mathbf{y} - \mathbf{G}_2\,\mathrm{vec}(\mathbf{X}) = \mathbf{G}_1\,\mathrm{vec}(\mathbf{R}_1).$$

The matrix $\mathbf{R}_1 \in \mathbb{R}^{r_1 \times n}$ has nr_1 unknown entries. Since $\mathbf{G}_1$ is a $K \times nr_1$ Gaussian matrix, it has full column rank almost surely whenever $K \ge nr_1$, and hence $\mathbf{R}_1$ is the unique LS solution. This condition is automatically satisfied by the Step-1 budget. Consequently, the same K original measurements provide $K - nr_1$ projected measurements for recovering $\mathbf{X}$ and retain the nr_1-dimensional component needed to recover $\mathbf{R}_1$; no additional samples are collected in Step 2. Therefore the total sample complexity is

$$K = nr_1 + \mathcal{O}((r - r_1)(m + n - r)).$$

Substitution into (16) then recovers $\mathbf{M}$, proving (31). □

C Extensions and Discussion

In this section, we study the DoF when both column- and row-space priors are available. We then analyze how measurement noise affects recovery accuracy. We also introduce an iterative refinement of DORM to improve recovery from noisy measurements.

C.1 Extension to partial column and row priors

In this subsection, we extend the DoF analysis in Theorem 3.2 to the case where partial information about both the column and row spaces is available. The following theorem gives the resulting DoF.

Theorem C.1. *Let $\mathbf{M} \in \mathbb{R}^{m\times n}$ have rank r. Suppose an r_1-dimensional subspace $\mathcal{U}_1 \subseteq \mathrm{col}(\mathbf{M})$ and an s_1-dimensional subspace $\mathcal{V}_1 \subseteq \mathrm{row}(\mathbf{M})$ are known, where $0 \le r_1, s_1 \le r$. Then*

$$\mathrm{DoF}(\mathbf{M} \mid \mathcal{U}_1 \subseteq \mathrm{col}(\mathbf{M}),\ \mathcal{V}_1 \subseteq \mathrm{row}(\mathbf{M})) = r^2 + (m-r)(r-r_1) + (n-r)(r-s_1). \tag{47}$$

Proof. We first determine the unknown part of the column space. Since $\mathcal{U}_1 \subseteq \mathrm{col}(\mathbf{M})$, the remaining uncertainty of the column space is

$$\mathrm{DoF}(\mathcal{U} \mid \mathcal{U}_1 \subseteq \mathcal{U}) = (m-r)(r-r_1),$$

from Lemma 3.1.

We next determine the unknown part of the row space. By symmetry, because $\mathcal{V}_1 \subseteq \mathrm{row}(\mathbf{M})$, the remaining uncertainty of the row space is

$$\mathrm{DoF}(\mathcal{V} \mid \mathcal{V}_1 \subseteq \mathcal{V}) = (n-r)(r-s_1).$$

Finally, we determine the coordinates under fixed column and row spaces. After both full column and row spaces are fixed, $\mathbf{M}$ can be written as $\mathbf{M} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$, where $\mathbf{U} \in \mathbb{R}^{m\times r}$ and $\mathbf{V} \in \mathbb{R}^{n\times r}$ are fixed basis matrices of the two spaces, and $\mathbf{S} \in \mathbb{R}^{r\times r}$ is free. Hence

$$\mathrm{DoF}(\mathbf{M} \mid \mathrm{col}(\mathbf{M}) = \mathcal{U},\ \mathrm{row}(\mathbf{M}) = \mathcal{V}) = r^2.$$

By additivity of manifold dimensions,

$$\begin{aligned}\mathrm{DoF}(\mathbf{M} \mid \mathcal{U}_1 \subseteq \mathrm{col}(\mathbf{M}),\ \mathcal{V}_1 \subseteq \mathrm{row}(\mathbf{M})) &= \mathrm{DoF}(\mathbf{M} \mid \mathrm{col}(\mathbf{M}) = \mathcal{U},\ \mathrm{row}(\mathbf{M}) = \mathcal{V}) \\ &\quad + \mathrm{DoF}(\mathcal{U} \mid \mathcal{U}_1 \subseteq \mathcal{U}) + \mathrm{DoF}(\mathcal{V} \mid \mathcal{V}_1 \subseteq \mathcal{V}) \\ &= r^2 + (m-r)(r-r_1) + (n-r)(r-s_1).\end{aligned}$$

This completes the proof. □

C.2 Noise stability under measurement noise

In this subsection, we analyze the recovery errors of the LS algorithm in Section 4.1, DORM in Section 4.2.1, and PMR in Section 4.2.2 under noisy measurements.

C.2.1 Complete column-space prior

In this part, we study the LS estimator from Section 4.1 under measurement noise. Recall that, in the noiseless setting, the complete-prior measurements reduce to (13), namely $\mathbf{y} = \widetilde{\mathbf{A}}\,\mathrm{vec}(\mathbf{R})$ with $\widetilde{\mathbf{A}} = \mathbf{A}(\mathbf{I}_n \otimes \mathbf{U})$. With additive noise $\boldsymbol{\xi} \in \mathbb{R}^K$, the corresponding noisy model is

$$\mathbf{y} = \widetilde{\mathbf{A}}\,\mathrm{vec}(\mathbf{R}) + \boldsymbol{\xi}, \qquad \widetilde{\mathbf{A}} = \mathbf{A}(\mathbf{I}_n \otimes \mathbf{U}). \tag{48}$$

Applying the same LS formula as in (15) to (48) yields $\mathrm{vec}(\widehat{\mathbf{R}}) = (\widetilde{\mathbf{A}}^T\widetilde{\mathbf{A}})^{-1}\widetilde{\mathbf{A}}^T\mathbf{y}$, where $\widehat{\mathbf{R}} \in \mathbb{R}^{r\times n}$. The following theorem bounds the Frobenius error of the resulting estimate $\widehat{\mathbf{M}} = \mathbf{U}\widehat{\mathbf{R}}$.

Theorem C.2 (Complete-prior noise stability)**.** *Consider the noisy model* (48) *and the LS estimate $\widehat{\mathbf{R}}$ above, and set $\widehat{\mathbf{M}} = \mathbf{U}\widehat{\mathbf{R}}$. Let $p = nr$ and, for $t \ge 0$, define*

$$\alpha_K(p,t) := 1 - \sqrt{p/K} - t/\sqrt{K}. \tag{49}$$

Suppose the entries of the sensing matrices are independent $\mathcal{N}(0, 1/K)$ variables and $\alpha_K(p,t) > 0$. Then, with probability at least $1 - e^{-t^2/2}$ over the sensing operator,

$$\|\widehat{\mathbf{M}} - \mathbf{M}\|_F \le \frac{\|\boldsymbol{\xi}\|_2}{\alpha_K(p,t)}. \tag{50}$$

Proof. As in the proof of Theorem 4.1, $\widetilde{\mathbf{A}}$ is a $K \times p$ matrix with independent $\mathcal{N}(0, 1/K)$ entries. The standard Gaussian smallest-singular-value bound (Vershynin, 2018) gives

$$\sigma_{\min}(\widetilde{\mathbf{A}}) \ge \alpha_K(p,t)$$

with the stated probability. Least-squares perturbation and the orthonormality of $\mathbf{U}$ then give

$$\|\widehat{\mathbf{M}} - \mathbf{M}\|_F = \|\widehat{\mathbf{R}} - \mathbf{R}\|_F \le \|\widetilde{\mathbf{A}}^\dagger\|_2 \|\boldsymbol{\xi}\|_2,$$

which proves (50). □

Compared with the noiseless result in Theorem 4.1, where $K = nr$ already yields exact recovery, the noisy bound requires $\alpha_K(p,t) > 0$ and thus $K > nr$. In other words, under measurement noise, Theorem C.2 provides a uniform error bound only when $K > nr$.

C.2.2 DORM WITH A PARTIAL COLUMN SPACE PRIOR

In this part, we analyze the recovery error of DORM when the observations are noisy. In Stage 1, the block $\mathbf{M}_1 \in \mathbb{R}^{m\times(r-r_1)}$ is observed with additive noise $\mathbf{E}_1 \in \mathbb{R}^{m\times(r-r_1)}$, so the observed block is

$$\overline{\mathbf{M}}_1 = \mathbf{M}_1 + \mathbf{E}_1. \tag{51}$$

Recall from (21) that $\mathbf{P} = \mathbf{I}_m - \mathbf{U}_1\mathbf{U}_1^T$ projects onto $\mathrm{col}(\mathbf{U}_1)^\perp$. Following (22), DORM estimates the missing subspace basis from the noisy projected block as

$$\widehat{\mathbf{U}}_2 = \mathrm{orth}(\mathbf{P}\overline{\mathbf{M}}_1) \in \mathbb{R}^{m\times(r-r_1)}, \tag{52}$$

using its leading $r - r_1$ left singular vectors. Because of the noise, $\mathrm{col}(\widehat{\mathbf{U}}_2)$ generally differs from the true subspace $\mathcal{U}_2$. For convenience, we let γ be the $(r - r_1)$-th largest singular value of the projected block $\mathbf{PM}_1$, and write

$$\gamma := \sigma_{r-r_1}(\mathbf{PM}_1), \qquad \varepsilon_{\mathrm{B}} := \|\mathbf{PE}_1\|_2, \qquad \eta_{\mathrm{B}} := \frac{\varepsilon_{\mathrm{B}}}{\gamma - \varepsilon_{\mathrm{B}}}. \tag{53}$$

Here, $\gamma > 0$ quantifies the nondegeneracy condition in Assumption 4.1, ε_{B} is the spectral norm of the projected Stage-1 perturbation, and η_{B} is the resulting subspace-perturbation factor.

Lemma C.1 (DORM subspace estimation error)**.** *Consider Stage 1 of DORM under the noisy observation model* (51)*. Let $\mathbf{U}_2$ be an orthonormal basis for the true missing subspace $\mathcal{U}_2$, and let $\widehat{\mathbf{U}}_2$ be DORM's estimate in* (52)*. Define the corresponding orthogonal projectors by*

$$\mathbf{P}_2 := \mathbf{U}_2\mathbf{U}_2^T, \qquad \widehat{\mathbf{P}}_2 := \widehat{\mathbf{U}}_2\widehat{\mathbf{U}}_2^T. \tag{54}$$

With γ, ε_{B}, and η_{B} defined in (53)*, if $\varepsilon_{\mathrm{B}} < \gamma$, then*

$$\|\widehat{\mathbf{P}}_2 - \mathbf{P}_2\|_2 \le \eta_{\mathrm{B}}. \tag{55}$$

Proof. Wedin's perturbation theorem (Wedin, 1972), applied to $\mathbf{P}\overline{\mathbf{M}}_1 = \mathbf{PM}_1 + \mathbf{PE}_1$, gives

$$\|\widehat{\mathbf{P}}_2 - \mathbf{P}_2\|_2 \le \frac{\varepsilon_{\mathrm{B}}}{\gamma - \varepsilon_{\mathrm{B}}} = \eta_{\mathrm{B}}.$$

This concludes the proof. □

Consequently, the estimated full basis $\widehat{\mathbf{U}} = [\mathbf{U}_1, \widehat{\mathbf{U}}_2] \in \mathbb{R}^{m\times r}$ generally spans a column space different from that of $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2] \in \mathbb{R}^{m\times r}$. The following corollary quantifies this full column-space error, which propagates to the matrix recovery in step (ii).

Corollary C.1 (DORM full column-space estimation error)**.** *Under the conditions of Lemma C.1, let* $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$ *and* $\widehat{\mathbf{U}} = [\mathbf{U}_1, \widehat{\mathbf{U}}_2]$*. Then the full column-space error satisfies*

$$\|\widehat{\mathbf{U}}\widehat{\mathbf{U}}^T - \mathbf{U}\mathbf{U}^T\|_2 = \|\widehat{\mathbf{P}}_2 - \mathbf{P}_2\|_2 \le \eta_{\mathrm{B}}. \tag{56}$$

Proof. Since the known basis $\mathbf{U}_1$ is unchanged, the full column-space projectors satisfy

$$\widehat{\mathbf{U}}\widehat{\mathbf{U}}^T - \mathbf{U}\mathbf{U}^T = \widehat{\mathbf{P}}_2 - \mathbf{P}_2.$$

The conclusion follows from Lemma C.1. □

We next use this subspace error bound to analyze DORM's matrix recovery error following step (ii) in Section 4.2.1. With the measurement matrices $\mathbf{C}_k = [\mathbf{0}, \mathbf{G}_k]$ in (24), the noisy observations satisfy

$$\mathbf{y}_{\mathcal{C}} = \mathcal{C}(\mathbf{M}) + \boldsymbol{\xi}_2 = \mathbf{G}\,\mathrm{vec}(\mathbf{M}_2) + \boldsymbol{\xi}_2, \tag{57}$$

where $\mathbf{G}$ has k-th row $\mathrm{vec}(\mathbf{G}_k)^T$, as in (25), and $\boldsymbol{\xi}_2 \in \mathbb{R}^{K_2}$ is the measurement noise. Assume that the entries of $\mathbf{G}_k$ are i.i.d. $\mathcal{N}(0, 1/K_2)$ and that the Stage-2 operator is independent of Stage 1. Using the estimated basis $\widehat{\mathbf{U}}$, DORM solves the same LS problem as in (25), with

$$\widetilde{\mathbf{C}} := \mathbf{G}(\mathbf{I}_{n-r+r_1} \otimes \widehat{\mathbf{U}}) \in \mathbb{R}^{K_2 \times r(n-r+r_1)}, \qquad \widehat{\mathbf{M}}_2 = \widehat{\mathbf{U}}\widehat{\mathbf{S}}_2. \tag{58}$$

Here $\widehat{\mathbf{S}}_2$ denotes the LS estimate, and DORM returns $\widehat{\mathbf{M}} = [\overline{\mathbf{M}}_1, \widehat{\mathbf{M}}_2]$. The following theorem bounds DORM's recovery error by combining the Stage-1 subspace estimation error with the Stage-2 LS perturbation.

Theorem C.3 (DORM noise stability)**.** *Consider DORM under the partial column-space prior in Section 4.2.1 and the noisy observations in* (51) *and* (57)*. Assume that the matrices* $\mathbf{G}_k$ *in* (24) *have i.i.d.* $\mathcal{N}(0, 1/K_2)$ *entries. Suppose that* $\gamma > 0$ *and, for some* $\delta_{\mathrm{B}} \in [0, 1]$*,*

$$\mathbb{P}(\varepsilon_{\mathrm{B}} < \gamma) \ge 1 - \delta_{\mathrm{B}},$$

where ε_{B} *is defined in* (53)*. Let* $\widehat{\mathbf{M}} = [\overline{\mathbf{M}}_1, \widehat{\mathbf{M}}_2]$ *be the DORM estimate, where* $\overline{\mathbf{M}}_1$ *is from* (51) *and* $\widehat{\mathbf{M}}_2$ *is from* (58)*. Then, for any* $t \ge 0$ *such that* $\alpha_2 := 1 - \sqrt{p_2/K_2} - t/\sqrt{K_2} > 0$*, with* $p_2 = r(n - r + r_1)$*, with probability at least* $1 - \delta_{\mathrm{B}} - 2e^{-t^2/2}$ *over the Stage-1 perturbation and the Stage-2 operator, this estimate satisfies*

$$\begin{aligned}\|\widehat{\mathbf{M}} - \mathbf{M}\|_F \le \|\mathbf{E}_1\|_F + \frac{\|\boldsymbol{\xi}_2\|_2}{\alpha_2} \\ + \left(1 + \frac{\beta_2}{\alpha_2}\right)\eta_{\mathrm{B}}\|\mathbf{P}\mathbf{M}_2\|_F,\end{aligned} \tag{59}$$

where $\mathbf{P}$ *is defined in* (21)*,* γ*,* ε_{B}*, and* η_{B} *in* (53)*, and* $\beta_2 := 1 + t/\sqrt{K_2}$*.*

Proof. Consider the Stage-1 event $\mathcal{E}_{\mathrm{B}} := \{\varepsilon_{\mathrm{B}} < \gamma\}$. On this event, Lemma C.1 and Corollary C.1 give the subspace error bounds in (55) and (56).

To account for the error in $\widehat{\mathbf{U}}$, define the projected coefficients $\mathbf{S}_{2,0} \in \mathbb{R}^{r \times (n-r+r_1)}$ and the orthogonal residual $\mathbf{E}_\perp \in \mathbb{R}^{m \times (n-r+r_1)}$ by

$$\mathbf{S}_{2,0} := \widehat{\mathbf{U}}^T\mathbf{M}_2, \qquad \mathbf{E}_\perp := (\mathbf{I}_m - \widehat{\mathbf{U}}\widehat{\mathbf{U}}^T)\mathbf{M}_2. \tag{60}$$

Thus $\mathbf{M}_2 = \widehat{\mathbf{U}}\mathbf{S}_{2,0} + \mathbf{E}_\perp$, and the Stage-2 observation model in (57) becomes

$$\mathbf{y}_{\mathcal{C}} = \widetilde{\mathbf{C}}\,\mathrm{vec}(\mathbf{S}_{2,0}) + \mathbf{G}\,\mathrm{vec}(\mathbf{E}_\perp) + \boldsymbol{\xi}_2. \tag{61}$$

Using $\widehat{\mathbf{M}}_2 = \widehat{\mathbf{U}}\widehat{\mathbf{S}}_2$, where $\widehat{\mathbf{S}}_2$ is given in (25), and the decomposition in (60), we obtain

$$\begin{aligned}\|\widehat{\mathbf{M}}_2 - \mathbf{M}_2\|_F &= \|\widehat{\mathbf{U}}\widehat{\mathbf{S}}_2 - (\widehat{\mathbf{U}}\mathbf{S}_{2,0} + \mathbf{E}_\perp)\|_F \\ &= \|\widehat{\mathbf{U}}(\widehat{\mathbf{S}}_2 - \mathbf{S}_{2,0}) - \mathbf{E}_\perp\|_F \\ &\le \|\widehat{\mathbf{U}}(\widehat{\mathbf{S}}_2 - \mathbf{S}_{2,0})\|_F + \|\mathbf{E}_\perp\|_F \\ &= \|\widehat{\mathbf{S}}_2 - \mathbf{S}_{2,0}\|_F + \|\mathbf{E}_\perp\|_F,\end{aligned} \tag{62}$$

where the last equality uses the orthonormality of $\widehat{\mathbf{U}}$. In the following, we bound the two terms on the right-hand side of (62) separately.

To bound $\|\mathbf{E}_\perp\|_F$, note that the supplied basis $\mathbf{U}_1$ is included unchanged in $\widehat{\mathbf{U}} = [\mathbf{U}_1, \widehat{\mathbf{U}}_2]$, so the residual in (60) can be written as $\mathbf{E}_\perp = (\mathbf{P}_2 - \widehat{\mathbf{P}}_2)\mathbf{P}\mathbf{M}_2$. Consequently, Lemma C.1 yields

$$\|\mathbf{E}_\perp\|_F \le \eta_{\mathrm{B}} \|\mathbf{P}\mathbf{M}_2\|_F. \tag{63}$$

We next bound $\|\widehat{\mathbf{S}}_2 - \mathbf{S}_{2,0}\|_F$ in (62). The LS solution in (25) satisfies $\mathrm{vec}(\widehat{\mathbf{S}}_2) = \widetilde{\mathbf{C}}^\dagger \mathbf{y}_{\mathcal{C}}$. Together with (61), this gives

$$\mathrm{vec}(\widehat{\mathbf{S}}_2 - \mathbf{S}_{2,0}) = \widetilde{\mathbf{C}}^\dagger \big(\mathbf{G}\,\mathrm{vec}(\mathbf{E}_\perp) + \boldsymbol{\xi}_2\big).$$

Conditional on the Stage-1 observations, $\widehat{\mathbf{U}}$ is fixed and has orthonormal columns. Thus, as in the analysis of (25), $\widetilde{\mathbf{C}} = \mathbf{G}(\mathbf{I}_{n-r+r_1} \otimes \widehat{\mathbf{U}})$ has independent $\mathcal{N}(0, 1/K_2)$ entries. The Gaussian smallest-singular-value bound (Vershynin, 2018) therefore yields an event $\mathcal{E}_\sigma$ of probability at least $1 - e^{-t^2/2}$ on which

$$\sigma_{\min}(\widetilde{\mathbf{C}}) \ge \alpha_2. \tag{64}$$

In addition, $\mathbf{E}_\perp$ is fixed conditional on Stage 1 and the Stage-2 operator is independent, so there is a second event $\mathcal{E}_G$ of probability at least $1 - e^{-t^2/2}$ on which

$$\|\mathbf{G}\,\mathrm{vec}(\mathbf{E}_\perp)\|_2 \le \beta_2 \|\mathbf{E}_\perp\|_F. \tag{65}$$

Combining (62) with $\|\widetilde{\mathbf{C}}^\dagger\|_2 \le 1/\alpha_2$ from (64) and (65) gives

$$\begin{aligned}
\|\widehat{\mathbf{M}}_2 - \mathbf{M}_2\|_F &\le \|\widehat{\mathbf{S}}_2 - \mathbf{S}_{2,0}\|_F + \|\mathbf{E}_\perp\|_F \\
&= \Big\|\widetilde{\mathbf{C}}^\dagger \big(\mathbf{G}\,\mathrm{vec}(\mathbf{E}_\perp) + \boldsymbol{\xi}_2\big)\Big\|_2 + \|\mathbf{E}_\perp\|_F \\
&\le \|\widetilde{\mathbf{C}}^\dagger\|_2 \big(\|\mathbf{G}\,\mathrm{vec}(\mathbf{E}_\perp)\|_2 + \|\boldsymbol{\xi}_2\|_2\big) + \|\mathbf{E}_\perp\|_F \\
&\le \|\mathbf{E}_\perp\|_F + \frac{\|\mathbf{G}\,\mathrm{vec}(\mathbf{E}_\perp)\|_2 + \|\boldsymbol{\xi}_2\|_2}{\alpha_2} \\
&\le \left(1 + \frac{\beta_2}{\alpha_2}\right) \|\mathbf{E}_\perp\|_F + \frac{\|\boldsymbol{\xi}_2\|_2}{\alpha_2},
\end{aligned} \tag{66}$$

where the last two inequalities hold with probability at least $1 - 2e^{-t^2/2}$, namely on the event $\mathcal{E}_\sigma \cap \mathcal{E}_G$ defined through (64) and (65). Finally, substituting (63) into (66) and adding the Stage-1 block error $\|\mathbf{E}_1\|_F$, we obtain on $\mathcal{E}_{\mathrm{B}} \cap \mathcal{E}_\sigma \cap \mathcal{E}_G$ the recovery error bound

$$\begin{aligned}
\|\widehat{\mathbf{M}} - \mathbf{M}\|_F &\le \|\mathbf{E}_1\|_F + \|\widehat{\mathbf{M}}_2 - \mathbf{M}_2\|_F \\
&\le \|\mathbf{E}_1\|_F + \frac{\|\boldsymbol{\xi}_2\|_2}{\alpha_2} + \left(1 + \frac{\beta_2}{\alpha_2}\right) \eta_{\mathrm{B}} \|\mathbf{P}\mathbf{M}_2\|_F.
\end{aligned} \tag{67}$$

This is precisely (59). Consequently, $\mathbb{P}(\mathcal{E}_{\mathrm{B}}^c) \le \delta_{\mathrm{B}}$, while $\mathbb{P}(\mathcal{E}_\sigma^c) \le e^{-t^2/2}$ and $\mathbb{P}(\mathcal{E}_G^c) \le e^{-t^2/2}$. A union bound therefore gives the joint success probability $1 - \delta_{\mathrm{B}} - 2e^{-t^2/2}$. □

One can observe from Theorem C.3 that the three terms in (59) correspond, respectively, to the direct Stage-1 block error, amplification of the Stage-2 measurement noise by the reduced LS system, and propagation of the Stage-1 subspace error.

C.2.3 PMR WITH A PARTIAL COLUMN SPACE PRIOR

In this part, we analyze the recovery error of PMR under noisy measurements. Recall the noiseless model in (27). Adding measurement noise $\boldsymbol{\xi} \in \mathbb{R}^K$, we obtain

$$\mathbf{y} = \mathbf{G}_1\,\mathrm{vec}(\mathbf{R}_1) + \mathbf{G}_2\,\mathrm{vec}(\mathbf{X}) + \boldsymbol{\xi}. \tag{68}$$

With the orthonormal matrix $\mathbf{Q}_2$ in (28), projecting (68) gives the noisy effective observations

$$\widetilde{\mathbf{y}} = \mathbf{Q}_2^T \mathbf{y} = \mathcal{H}(\mathbf{X}) + \mathbf{Q}_2^T \boldsymbol{\xi}. \tag{69}$$

Here $\mathcal{H}$ is the same effective operator as in (28), namely $\mathcal{H}(\mathbf{X}) = \mathbf{Q}_2^T \mathbf{G}_2 \operatorname{vec}(\mathbf{X})$. Assume that a low-rank solver applied to (69) returns an estimate $\widehat{\mathbf{X}}$ satisfying

$$\|\widehat{\mathbf{X}} - \mathbf{X}\|_F \le C_{\mathcal{H}} \|\mathbf{Q}_2^T \boldsymbol{\xi}\|_2 \tag{70}$$

with probability at least $1 - \delta_{\mathcal{H}}$. Such stability guarantees are available for standard noisy low-rank recovery methods, including nuclear-norm minimization (Candes & Plan, 2010; Recht et al., 2010; Candès & Plan, 2011) and alternating minimization (Jain et al., 2013; Hardt, 2014). With $\widehat{\mathbf{X}}$ in hand, the LS refit (29) produces $\widehat{\mathbf{R}}_1$, and the final PMR estimate is given by (30). The following theorem provides the PMR recovery error bound under the noisy measurements (68).

Theorem C.4 (PMR noise stability)**.** *Let $\mathbf{Q}_1 \in \mathbb{R}^{K \times nr_1}$ have orthonormal columns spanning $\operatorname{col}(\mathbf{G}_1)$. For $t \ge 0$, define*

$$\alpha_{\mathrm{P}} := 1 - \sqrt{\frac{nr_1}{K}} - \frac{t}{\sqrt{K}}, \qquad \rho_{\mathrm{P}} := \sqrt{\frac{nr_1}{K}} + \frac{t}{\sqrt{K}}. \tag{71}$$

If $\alpha_{\mathrm{P}} > 0$, then the final PMR estimate in (30) *satisfies, with probability at least $1 - \delta_{\mathcal{H}} - 2e^{-t^2/2}$,*

$$\begin{aligned} \|\widehat{\mathbf{M}} - \mathbf{M}\|_F^2 \le{}& C_{\mathcal{H}}^2 \|\mathbf{Q}_2^T \boldsymbol{\xi}\|_2^2 \\ &+ \frac{\left(\|\mathbf{Q}_1^T \boldsymbol{\xi}\|_2 + \rho_{\mathrm{P}} C_{\mathcal{H}} \|\mathbf{Q}_2^T \boldsymbol{\xi}\|_2\right)^2}{\alpha_{\mathrm{P}}^2}. \end{aligned} \tag{72}$$

Proof. Let $\boldsymbol{\Delta}_X = \operatorname{vec}(\widehat{\mathbf{X}} - \mathbf{X}) \in \mathbb{R}^{(m-r_1)n}$. Substituting (68) into the refit (29) gives

$$\operatorname{vec}(\widehat{\mathbf{R}}_1 - \mathbf{R}_1) = \mathbf{G}_1^{\dagger}(\boldsymbol{\xi} - \mathbf{G}_2 \boldsymbol{\Delta}_X).$$

The Gaussian smallest-singular-value bound gives $\sigma_{\min}(\mathbf{G}_1) \ge \alpha_{\mathrm{P}}$ except on an event of probability $e^{-t^2/2}$. Moreover, conditional on the projected problem and hence on $\boldsymbol{\Delta}_X$, the independence of the orthogonal Gaussian blocks $\mathbf{Q}_1^T \mathbf{G}_2$ and $\mathbf{Q}_2^T \mathbf{G}_2$ yields

$$\|\mathbf{Q}_1^T \mathbf{G}_2 \boldsymbol{\Delta}_X\|_2 \le \rho_{\mathrm{P}} \|\boldsymbol{\Delta}_X\|_2$$

except on another event of probability $e^{-t^2/2}$. Therefore

$$\|\widehat{\mathbf{R}}_1 - \mathbf{R}_1\|_F \le \frac{\|\mathbf{Q}_1^T \boldsymbol{\xi}\|_2 + \rho_{\mathrm{P}} \|\widehat{\mathbf{X}} - \mathbf{X}\|_F}{\alpha_{\mathrm{P}}}.$$

Combining this inequality with (70), the orthogonality of $\mathbf{U}_1$ and $\mathbf{U}_\perp$, and a union bound proves (72). □

The bound separates the two sources of PMR error. The first is the stability of the chosen low-rank solver on the projected problem, represented by $C_{\mathcal{H}}$. The second is the LS refit of $\mathbf{R}_1$, whose noise amplification is controlled by α_{P}^{-1}. An error in $\widehat{\mathbf{X}}$ also propagates into this refit through ρ_{P}. Hence oversampling improves both the conditioning of the projected recovery problem and that of the refit. Unlike DORM, PMR has no Stage-1 subspace term when the supplied prior $\mathbf{U}_1$ is exact; prior mismatch is a separate source of model error not covered by Theorem C.4.

C.3 Iterative refinement of DORM under measurement noise

In this subsection, we introduce an iterative refinement of DORM that mitigates sequential error propagation under measurement noise. The construction in Section 4.2.1 first recovers the missing subspace from $\mathcal{B}$ and then the remaining coefficients from $\mathcal{C}$, so a first-stage error is carried into the second stage. To mitigate this, we replace the sequential two-stage estimation with an alternating optimization procedure.

Suppose the noisy observations corresponding to (17) are

$$\mathbf{y}_{\mathcal{B}} = \mathcal{B}(\mathbf{M}) + \boldsymbol{\varepsilon}_{\mathcal{B}}, \qquad \mathbf{y}_{\mathcal{C}} = \mathcal{C}(\mathbf{M}) + \boldsymbol{\varepsilon}_{\mathcal{C}}, \tag{73}$$

Algorithm 3 DORM with joint alternating refinement (DORM-Refine)

Input: Exact prior basis $\mathbf{U}_1$; operators $\mathcal{B}$ and $\mathcal{C}$; observations $\mathbf{y}_\mathcal{B}$ and $\mathbf{y}_\mathcal{C}$; rank r; maximum iterations H; tolerance τ

1: Form $\mathcal{D}$ and $\mathbf{z}$ by (74)
2: Initialize $\mathbf{U}_2^{(0)}$ and $\mathbf{U}^{(0)}$ by (77)
3: **for** $h = 1$ to H **do**
4: Fix $\mathbf{U}^{(h-1)}$ and update $\mathbf{R}^{(h)}$ by (78)
5: Partition $\mathbf{R}^{(h)}$ into $\mathbf{R}_1^{(h)}$ and $\mathbf{R}_2^{(h)}$
6: Fix $\mathbf{R}^{(h)}$ and update $\mathbf{U}_2^{(h)}$ by (79)
7: Set $\mathbf{U}^{(h)} = [\mathbf{U}_1, \mathbf{U}_2^{(h)}]$ and $\widehat{\mathbf{M}}^{(h)} = \mathbf{U}^{(h)}\mathbf{R}^{(h)}$
8: Stop if the relative decrease of (76) is below τ
9: **end for**
10: With the final basis $\mathbf{U}^{(h)}$, refit the coefficients once to obtain $\mathbf{R}^{\mathrm{out}}$

Output: $\widehat{\mathbf{M}} = \mathbf{U}^{(h)}\mathbf{R}^{\mathrm{out}}$

where $\mathbf{y}_\mathcal{B}, \boldsymbol{\varepsilon}_\mathcal{B} \in \mathbb{R}^{K_1}$ and $\mathbf{y}_\mathcal{C}, \boldsymbol{\varepsilon}_\mathcal{C} \in \mathbb{R}^{K_2}$. Define the weighted stacked operator $\mathcal{D} : \mathbb{R}^{m\times n} \to \mathbb{R}^{K_1+K_2}$ and stacked data $\mathbf{z} \in \mathbb{R}^{K_1+K_2}$ by

$$\mathcal{D}(\mathbf{X}) := \begin{bmatrix} w_\mathcal{B}\mathcal{B}(\mathbf{X}) \\ w_\mathcal{C}\mathcal{C}(\mathbf{X}) \end{bmatrix}, \qquad \mathbf{z} := \begin{bmatrix} w_\mathcal{B}\mathbf{y}_\mathcal{B} \\ w_\mathcal{C}\mathbf{y}_\mathcal{C} \end{bmatrix}. \tag{74}$$

The positive weights $w_\mathcal{B}$ and $w_\mathcal{C}$ may be chosen as the inverse noise standard deviations of the corresponding stages. When the two noise levels are unknown, we simply set both weights to one. In either case, refinement reuses the original $K_1 + K_2$ DORM measurements and requires no additional samples.

Let $q := r - r_1$ be the dimension of the missing subspace, with orthonormal basis $\mathbf{U}_2 \in \mathbb{R}^{m\times q}$. Define the complete column-space basis $\mathbf{U} := [\mathbf{U}_1, \mathbf{U}_2] \in \mathbb{R}^{m\times r}$ and partition the coefficient matrix as $\mathbf{R} := [\mathbf{R}_1^T, \mathbf{R}_2^T]^T \in \mathbb{R}^{r\times n}$, where $\mathbf{R}_1 \in \mathbb{R}^{r_1\times n}$ and $\mathbf{R}_2 \in \mathbb{R}^{q\times n}$. Under the exact prior model, the true target matrix satisfies

$$\mathbf{M} = \mathbf{U}\mathbf{R} = \mathbf{U}_1\mathbf{R}_1 + \mathbf{U}_2\mathbf{R}_2. \tag{75}$$

The unknown block $\mathbf{U}_2$ satisfies $\mathbf{U}_2^T\mathbf{U}_2 = \mathbf{I}_q$ and $\mathbf{U}_1^T\mathbf{U}_2 = \mathbf{0}$, so $\mathbf{U}$ is an orthonormal basis for the full column space. The matrix $\mathbf{R}$ collects the coordinates of all n target columns in this basis. Its two row blocks are denoted by $\mathbf{R}_1$ and $\mathbf{R}_2$.

Using the representation (75), the refinement fits all stacked observations through the joint objective

$$F(\mathbf{U}, \mathbf{R}) := \|\mathbf{z} - \mathcal{D}(\mathbf{U}\mathbf{R})\|_2^2 + \lambda_R\|\mathbf{R}\|_F^2. \tag{76}$$

Here $\mathbf{U}\mathbf{R}$ is the current estimate of $\mathbf{M}$. We minimize (76) over $\mathbf{R}$ and the admissible choices of $\mathbf{U}_2$, while keeping $\mathbf{U}_1$ fixed. The optional parameter $\lambda_R \geq 0$ regularizes the coefficient update and can be chosen positive when the joint LS system is poorly conditioned. Starting from the DORM estimate, each refinement iteration performs two block updates: it first fixes $\mathbf{U}_2$ and refits $\mathbf{R}$, and then fixes $\mathbf{R}$ and updates $\mathbf{U}_2$. These updates are repeated until the objective stabilizes. The resulting procedure, termed DORM-Refine, is summarized in Algorithm 3, and the detailed steps are given below.

Initialization from DORM. We initialize the missing subspace from Stage 1 of DORM in Algorithm 1. The designed operator $\mathcal{B}$ has $K_1 = mq$ outputs. Reshape $\mathbf{y}_\mathcal{B}$ into the noisy block $\overline{\mathbf{M}}_1 \in \mathbb{R}^{m\times q}$, form $\mathbf{P} = \mathbf{I}_m - \mathbf{U}_1\mathbf{U}_1^T$, and let $\mathbf{U}_2^{(0)} \in \mathbb{R}^{m\times q}$ contain the leading q left singular vectors of $\mathbf{P}\overline{\mathbf{M}}_1$. Set

$$\mathbf{U}^{(0)} := \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2^{(0)} \end{bmatrix} \in \mathbb{R}^{m\times r}. \tag{77}$$

Subsequent updates start from this basis and do not require the Stage 2 DORM estimate.

Coefficient update. At iteration h, fix the current full basis $\mathbf{U}^{(h-1)} = [\mathbf{U}_1, \mathbf{U}_2^{(h-1)}] \in \mathbb{R}^{m\times r}$. The remaining unknown $\mathbf{R} \in \mathbb{R}^{r\times n}$ contains the coordinates of the n target columns in this basis and is estimated from all stacked observations:

$$\mathbf{R}^{(h)} := \operatorname*{arg\,min}_{\mathbf{R}\in\mathbb{R}^{r\times n}} \left\|\mathbf{z} - \mathcal{D}\Big(\mathbf{U}^{(h-1)}\mathbf{R}\Big)\right\|_2^2 + \lambda_R\|\mathbf{R}\|_F^2. \tag{78}$$

Because $\mathbf{U}^{(h-1)}$ is fixed and $\mathcal{D}$ is linear, (78) is a linear LS problem in the rn entries of $\mathbf{R}$. Partition its solution based on (75) as $\mathbf{R}^{(h)} = [(\mathbf{R}_1^{(h)})^T, (\mathbf{R}_2^{(h)})^T]^T$, where $\mathbf{R}_1^{(h)} \in \mathbb{R}^{r_1 \times n}$ and $\mathbf{R}_2^{(h)} \in \mathbb{R}^{q \times n}$.

Missing-subspace update. Next fix $\mathbf{R}_1^{(h)}$ and $\mathbf{R}_2^{(h)}$, keep the supplied basis $\mathbf{U}_1$ unchanged, and update only the missing directions through

$$\mathbf{U}_2^{(h)} \in \underset{\mathbf{U}_2 \in \mathbb{R}^{m \times q}}{\arg\min} \left\| \mathbf{z} - \mathcal{D}\Big(\mathbf{U}_1 \mathbf{R}_1^{(h)} + \mathbf{U}_2 \mathbf{R}_2^{(h)}\Big) \right\|_2^2, \quad \text{s.t. } \mathbf{U}_2^T \mathbf{U}_2 = \mathbf{I}_q, \qquad \mathbf{U}_1^T \mathbf{U}_2 = \mathbf{0}. \tag{79}$$

Then set $\mathbf{U}^{(h)} = [\mathbf{U}_1, \mathbf{U}_2^{(h)}] \in \mathbb{R}^{m \times r}$ and $\widehat{\mathbf{M}}^{(h)} = \mathbf{U}^{(h)} \mathbf{R}^{(h)}$. The subproblem in (79) is constrained LS on a Stiefel manifold. It may be solved directly by a manifold LS method, or approximately by projection onto $\mathrm{span}(\mathbf{U}_1)^\perp$ and orthonormalization.

Remark C.1. *The original DORM procedure in Section 4.2.1 corresponds to a sequential minimization of* (76)*: the $\mathcal{B}$ measurements first determine $\mathbf{U}$, after which the $\mathcal{C}$ measurements determine $\mathbf{R}$. In the noiseless setting, this sequential scheme already recovers $\mathbf{M}$ exactly at the DoF. Under measurement noise, however, an error in $\mathbf{U}$ is carried into the subsequent estimate of $\mathbf{R}$. DORM-Refine instead alternates between $\mathbf{U}$ and $\mathbf{R}$ in* (76)*, using both observation sets at every iteration, and thereby obtains stronger robustness.*

Empirical comparison under measurement noise. We compare direct DORM and DORM-Refine using the partial-prior noise setting of Appendix D: $(m, n, r) = (30, 50, 5)$, $r_1 = 3$, and independent Gaussian noise with standard deviation $\sigma = 10^{-3}$ in both stages. The resulting DoF is 300. At each measurement count, we perform 500 matched trials in which the two methods share the target matrix, both measurement operators, and both noise realizations. DORM-Refine uses $\lambda_R = 0$, at most 10 alternating iterations, and the stopping tolerance 10^{-7}. The curves in Figure C1 report the median relative Frobenius error; the error bars show the interquartile range.

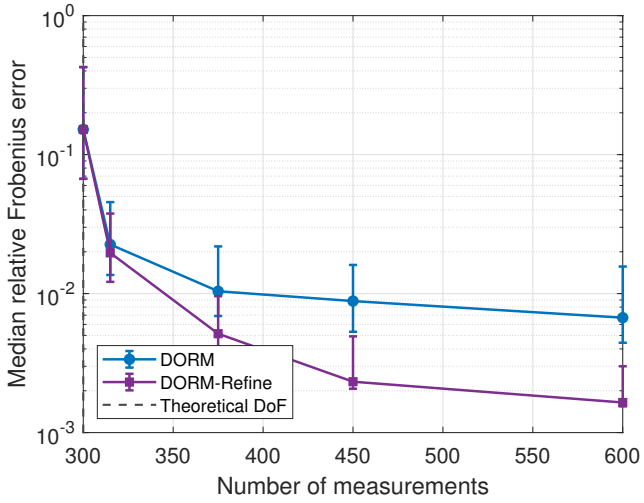


Figure C1: Recovery error of direct DORM and DORM-Refine under additive Gaussian measurement noise. Both methods use the same measurements and noise realization in every trial.

Figure C1 shows that both methods improve as the number of measurements increases, while DORM-Refine consistently attains a lower error. Thus, the alternating updates of $\mathbf{U}$ and $\mathbf{R}$ yield better recovery accuracy under measurement noise.

D Additional Experiments

In this section, we first evaluate PMR under different measurement designs and compare its performance with DORM. We then examine how measurement noise affects recovery accuracy. Next, we

analyze computational complexity and evaluate runtime for large-scale matrix recovery. We also provide the full experimental protocols and results for the Jester and USC massive-MIMO datasets.

D.1 PMR UNDER DORM MEASUREMENTS

In this subsection, we examine how PMR performs under different linear measurement designs. The proposed PMR in section 4.2.2 can be applied with arbitrary linear operators. Here we equip PMR with DORM's two-stage measurements and compare it with DORM. As a benchmark, we also evaluate PMR under ordinary i.i.d. Gaussian measurements. In the experiment, we use the same noiseless partial-prior setting as in Figure 1b, namely $(m, n, r, r_1) = (30, 50, 5, 3)$. DORM's two-stage design uses $K_1 = m(r - r_1) = 60$ entry measurements to reveal the first $r - r_1$ columns and $K_2 = K - K_1$ Gaussian measurements of the remaining columns, as in (24). We use the success criterion from Section 5 and report Wilson 95% confidence intervals over 500 trials.

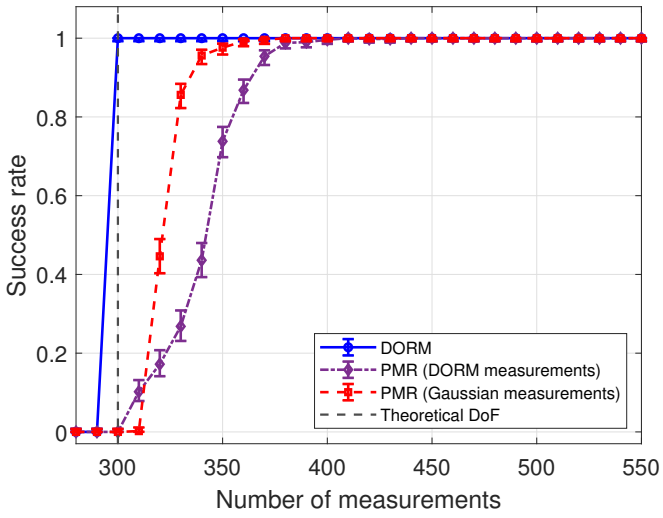


Figure D1: Noiseless recovery under a partial column space prior, comparing DORM with PMR under either DORM's two-stage measurements or i.i.d. Gaussian measurements. Every method uses the same total measurement count K. Only DORM achieves exact recovery at the DoF $K = 300$, whereas PMR requires oversampling under both measurement designs.

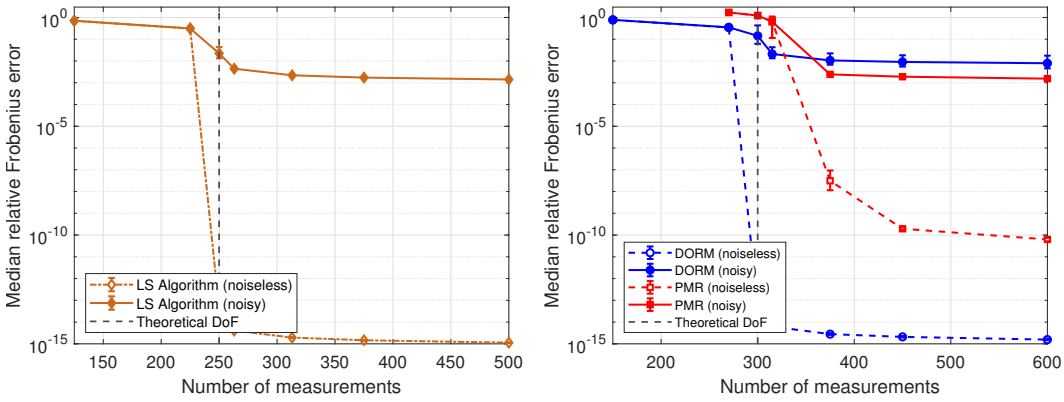
At the DoF count $K = 300$, DORM achieves a 100% success rate, whereas PMR fails under both measurement designs. PMR approaches 100% success only at $K = 350$ under i.i.d. Gaussian measurements (97.6%) and at $K = 380$ under DORM's measurements (98.8%). Thus, DORM's DoF-optimal recovery relies on the joint design of its measurement operator and recovery procedure.

D.2 MEASUREMENT NOISE

In this subsection, we empirically examine how additive measurement noise affects recovery under complete and partial column space priors. In each setting, the noise model and notation are chosen to match the stability analysis in Appendix C.2.

For the complete column-space prior, we use $(m, n, r) = (30, 50, 5)$, for which the complete-prior DoF is $nr = 250$. The LS algorithm is applied to the noisy reduced system in (48). Specifically, the entries of $\boldsymbol{\xi} \in \mathbb{R}^K$ are drawn independently as $\xi_k \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 10^{-3}$. For each measurement count, we conduct 500 matched trials: the noiseless and noisy reconstructions use the same target matrix and reduced Gaussian operator. We report the median relative Frobenius error, with the interquartile range shown by the error bars.

At $K = 250$, the median LS error increases from 4.17×10^{-14} without noise to 2.24×10^{-2} with noise. Increasing the measurement count to $K = 500$ reduces the noisy error to 1.41×10^{-3}. This behavior agrees with Theorem C.2: the square reduced system at the DoF can amplify noise, whereas oversampling improves its conditioning.



(a) Complete column-space prior.

(b) Partial column space prior.

Figure D2: Recovery under additive Gaussian measurement noise. The complete prior uses the LS algorithm, while the partial prior compares DORM and PMR. Curves show median relative Frobenius error, and error bars show the interquartile range over 500 matched trials.

We next consider the partial column space prior. We again use $(m, n, r) = (30, 50, 5)$ and set $r_1 = 3$, giving the partial-prior DoF $nr + (m - r)(r - r_1) = 300$. For DORM, the Stage-1 observation is $\overline{\mathbf{M}}_1 = \mathbf{M}_1 + \mathbf{E}_1$, and the Stage-2 observations follow (57). The entries of $\mathbf{E}_1$ and $\boldsymbol{\xi}_2$ are mutually independent and distributed as $\mathcal{N}(0, \sigma^2)$ with the same noise level $\sigma = 10^{-3}$. For PMR, we use the noisy Gaussian model in (68), with independent $\xi_k \sim \mathcal{N}(0, \sigma^2)$ and $\sigma = 10^{-3}$; the noisy measurements are projected as in (69), and alternating minimization is used to recover $\mathbf{X}$. For each method and measurement count, we conduct 500 matched noiseless–noisy trials using the same target matrix and sensing operator, and report the median relative Frobenius error and its interquartile range.

At $K = 300$, DORM attains a median noisy error of 0.143, which decreases to 7.87×10^{-3} at $K = 600$. PMR requires more than the DoF measurement count, consistent with Theorem 4.3: its median noisy error is 2.42×10^{-3} at $K = 375$ and 1.54×10^{-3} at $K = 600$.

These results show that exact recovery at the noiseless DoF is replaced by stable estimation under noise. At the DoF boundary, the reduced system is square and can amplify measurement noise; oversampling improves its conditioning and lowers the reconstruction error. Consequently, a noisy recovery threshold must also specify the noise level, target accuracy, and desired success probability rather than relying on the noiseless DoF alone.

D.3 Computational complexity and scaling

In this subsection, we study computational complexity and scaling. We separate the analysis according to whether the complete or only a partial column space prior is available. All experiments use Gaussian measurements and report relative Frobenius error. Random measurement generation is excluded from the recovery time.

D.3.1 Complete column-space prior

In this part, we consider the complete column-space prior. When $r_1 = r$, the unknown block $\mathbf{X}$ in PMR is zero, and both PMR and DORM reduce to the LS algorithm in Section 4.1. Hence all three methods solve the same reduced system (14), with nr unknowns. A dense direct solve costs

$$\mathcal{O}\big(Kn^2r^2 + n^3r^3\big)\,,$$

which becomes $\mathcal{O}(n^3r^3)$ at the exact-DoF count $K = nr$. A matrix-free gradient step costs $\mathcal{O}(Knr)$ because it requires one multiplication by the $K \times nr$ reduced operator and one by its adjoint.

To evaluate computational scaling, we generated square rank-5 matrices of size up to 1000×1000 and used the exact-DoF measurement count $K = nr = 5n$. For each size, we solved the same reduced LS problem using either a direct solver or Nesterov accelerated gradient (AG) with a fixed budget of 1000 iterations. The results are reported in Table 2.

Table 2: Complete-prior scaling at the exact DoF. LS, DORM, and PMR are the same reduced LS problem. AG uses 1000 iterations.

Size	K	Direct (s)	Direct error	AG (s)	AG error
100×100	500	0.005	5.34×10^{-13}	0.023	0.088
400×400	2000	0.054	3.62×10^{-12}	0.217	0.022
1000×1000	5000	0.317	2.42×10^{-11}	6.359	0.046

These results show that complete-prior recovery remains fast and accurate at large scales: even for a 1000×1000 matrix, the direct solution in (15) takes only 0.317 seconds and achieves an error of 2.42×10^{-11}. Although the gradient method has lower per-iteration complexity, its 1000-iteration run is slower and less accurate than the direct solution at the tested scales.

D.3.2 Partial column space prior

In this part, we analyze the computational complexity and scaling of DORM and PMR under a partial column space prior.

For DORM, Stage 1 extracts the $(r - r_1)$-dimensional missing subspace from an $m \times (r - r_1)$ block in $\mathcal{O}(mr(r - r_1))$ time. Stage 2 has $r(n - r + r_1)$ unknown coefficients. Its dense LS cost is

$$\mathcal{O}\big(K_2[r(n - r + r_1)]^2 + [r(n - r + r_1)]^3\big)\,,$$

which reduces to $\mathcal{O}([r(n - r + r_1)]^3)$ when $K_2 = r(n - r + r_1)$. A gradient step costs $\mathcal{O}(r^2(n - r + r_1)^2)$ at this allocation.

For PMR, the complexity of Stage 2, which recovers the rank-$(r - r_1)$ matrix $\mathbf{X} \in \mathbb{R}^{(m-r_1)\times n}$, depends on the low-rank recovery algorithm. We use alternating minimization as an example, consistent with our experiments; other recovery algorithms may have different complexities. Writing $\mathbf{X} = \mathbf{VR}$, the coefficient update has $n(r-r_1)$ unknowns and the subspace update has $(m-r_1)(r-r_1)$ unknowns. If both LS subproblems are solved directly, one iteration for recovering $\mathbf{X}$ costs

$$\begin{aligned}\mathcal{O}\Big(&(K - nr_1)(m - r_1)n(r - r_1)\\ &+ (K - nr_1)\big([n(r - r_1)]^2 + [(m - r_1)(r - r_1)]^2\big)\\ &+ [n(r - r_1)]^3 + [(m - r_1)(r - r_1)]^3\Big),\end{aligned}$$

where the first term forms the two alternating minimization design matrices up to a constant factor. A thin QR factorization of $\mathbf{G}_1 \in \mathbb{R}^{K\times nr_1}$ costs $\mathcal{O}(K(nr_1)^2)$, and the final recovery of $\mathbf{R}_1$ is one LS refit with nr_1 unknowns. For T iterations used to recover $\mathbf{X}$, the cost above is multiplied by T.

To evaluate computational scaling under a partial column space prior, we test both DORM and PMR on square matrices of size up to 1000×1000. We set $r = 5$ and $r_1 = 2$. DORM uses the exact conditional DoF $nr + (m - r)(r - r_1)$, whereas PMR uses 1.5 times this value and at most 100 iterations to recover $\mathbf{X}$. For PMR, we directly simulate the Gaussian effective model in Lemma 4.1 using a batched implementation on a graphics processing unit (GPU). Its reported times are medians over three fixed-seed trials and include projection, recovery of $\mathbf{X}$, the $\mathbf{R}_1$ refit, and final reconstruction.

These results show that partial-prior recovery remains fast and accurate at large scales. At the exact DoF, DORM recovers a 1000×1000 matrix in 0.318 seconds with an error of 1.03×10^{-11}. PMR also scales to 1000×1000 under unstructured Gaussian sensing, taking 190.781 seconds and attaining an error of 1.65×10^{-6} at the 100-iteration limit. Its higher runtime is consistent with the two LS updates performed in every iteration for recovering $\mathbf{X}$. Across all tested sizes, PMR achieves relative errors below 1.90×10^{-6}.

D.4 Real-world data protocols

In this subsection, we evaluate the performance of the proposed DORM and PMR on the real-world data, i.e., Jester recommendation and University of Southern California (USC) massive-MIMO data. The experimental protocols and supplementary results are presented below.

Table 3: Partial-prior scaling. DORM uses the exact DoF; PMR uses $1.5\times$ the DoF. Iterations denotes the number of iterations used to recover $\mathbf{X}$.

Method	Size	K	K/DoF	Iterations	Time (s)	Error
DORM	100×100	785	1.0	–	0.002	7.47×10^{-14}
DORM	400×400	3185	1.0	–	0.052	2.94×10^{-12}
DORM	1000×1000	7985	1.0	–	0.318	1.03×10^{-11}
PMR	100×100	1178	1.5	53	0.937	1.89×10^{-6}
PMR	400×400	4778	1.5	59	9.263	1.46×10^{-6}
PMR	1000×1000	11978	1.5	100	190.781	1.65×10^{-6}

D.4.1 JESTER RECOMMENDATION DATA

Data and historical subspace. Jester is a recommendation dataset comprising approximately 4.1 million ratings of 100 jokes on a scale from -10 to 10 (Goldberg et al., 2001). We select two disjoint groups of 1,000 users each from the 14,116 users who rated all 100 jokes. These groups form the historical matrix $\mathbf{H} \in \mathbb{R}^{100\times 1000}$ and the target matrix $\mathbf{M} \in \mathbb{R}^{100\times 1000}$, respectively, with rows corresponding to jokes and columns to users. We observe that both $\mathbf{H}$ and $\mathbf{M}$ exhibit approximately low-rank structure and adopt a rank-15 model for each. Considering that the historical data provide prior information about the column subspace of the target matrix $\mathbf{M}$, we take the subspace spanned by the leading ten left singular vectors of $\mathbf{H}$ as this prior, with basis $\mathbf{U}_1 \in \mathbb{R}^{100\times 10}$ and dimension $r_1 = 10$. We use this prior subspace information to recover $\mathbf{M}$.

Sampling and baselines. DORM first fully observes the first $r - r_1$ columns of $\mathbf{M}$, corresponding to the first five target users. It distributes the remaining observation budget as evenly as possible across the other users and samples ratings uniformly without replacement for each user. We use these rating samples in Stage 2 instead of the Gaussians in Section 4.2.1 to match the recommender setting; DORM remains applicable under this measurement scheme. PMR observes the same total number of ratings as DORM, but spreads them across all target users. Specifically, for each user, it selects a random subset of jokes uniformly without replacement and observes the corresponding ratings. The AM with no prior is given the identical set of user–joke ratings as PMR, but does not use the historical subspace $\mathbf{U}_1$. We denote the number of observed ratings in the target matrix $\mathbf{M}$ by K. The theoretical DoF under the partial-subspace prior is $\mathrm{DoF} = nr + (m - r)(r - r_1) = 15{,}425$.

Evaluation and results. For all three methods, we clip the entries of the final recovered matrix to the known rating range $[-10, 10]$ before evaluation. We report the mean absolute error (MAE), $\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}|\widehat{M}_{ij} - M_{ij}|$, averaged over all ten trials. Table 4 reports the results.

Table 4: Jester: MAE averaged over ten paired trials with $r_1 = 10$ and final entrywise clipping to $[-10, 10]$ (lower is better).

K/DoF	DORM	PMR	AM with no prior
120.00%	4.765	4.953	5.405
160.00%	3.738	3.903	4.259
200.00%	3.338	3.442	3.688

As shown in Table 4, the MAE values of all three methods decrease as the observation budget increases. DORM achieves the lowest error at every budget, followed by PMR and AM with no prior. PMR consistently outperforms AM and approaches DORM's accuracy. These results support the usefulness of the historical rating patterns encoded in $\mathbf{U}_1$ for recovering new users' ratings.

D.4.2 USC MASSIVE-MIMO DATA

Data and historical subspace. The USC WiDeS drone dataset contains wireless channel measurements collected along drone trajectories (Choi et al., 2022), capturing channel variation across

positions and frequencies. We use the trajectory segment identified as H3_70m_R3 in this dataset. At each drone position, we retain the same 500 uniformly selected frequency points out of 2,301 and take the magnitudes of the complex channel coefficients. Rows index frequencies and columns index drone positions. Positions 1–492 form the historical matrix $\mathbf{H} \in \mathbb{R}^{500\times 492}$, and positions 493–1,492 form the target matrix $\mathbf{M} \in \mathbb{R}^{500\times 1000}$. Since nearby drone positions share similar propagation conditions, we adopt a rank-20 model for both $\mathbf{H}$ and $\mathbf{M}$. In particular, for the prior information, we construct $\mathbf{U}_1 \in \mathbb{R}^{500\times 15}$ from the leading 15 left singular vectors of $\mathbf{H}$, representing dominant frequency profiles along the preceding trajectory. These vectors span the known $r_1 = 15$-dimensional prior subspace. Therefore, our task is to use this prior subspace information to recover $\mathbf{M}$, estimating the remaining five-dimensional subspace from the observations.

Measurements and baselines. Each method receives a budget of K scalar observations. DORM first observes all entries in five randomly selected columns of $\mathbf{M}$, using $5m$ observations. It then takes Gaussian linear measurements of the remaining 995 columns of $\mathbf{M}$, allocating the remaining $K - 5m$ observations as evenly as possible among them. PMR and AM with no prior instead allocate all K Gaussian linear measurements among all 1,000 columns. We express each budget as K/DoF, where $\mathrm{DoF} = nr + (m - r)(r - r_1) = 22{,}400$ for all methods.

Evaluation and results. We evaluate the normalized mean squared error (NMSE),

$$\mathrm{NMSE} = \frac{\|\widehat{\mathbf{M}} - \mathbf{M}\|_F^2}{\|\mathbf{M}\|_F^2},$$

using the complete target matrix. We report the arithmetic mean of the per-trial NMSE over ten paired trials in table 5.

Table 5: USC massive MIMO: mean NMSE over ten paired trials at equal scalar observation budgets (lower is better).

K/DoF	DORM	PMR	AM with no prior
120.00%	0.0925	0.1333	1.0563
160.00%	0.0355	0.0504	0.3976
200.00%	0.0263	0.0315	0.1165

As shown in Table 5, the recovery errors of all three methods decrease as the observation budget increases. At every tested budget, DORM achieves the lowest mean NMSE, followed by PMR and AM with no prior. These results support the usefulness of the historical prior $\mathbf{U}_1$. The frequency profiles along the preceding trajectory help recover subsequent channel magnitudes.