Semi-Random Noisy and One-Bit Matrix Completion via Nonconvex Optimization

Xing Gao
University of Illinois Chicago

Binhao Chen
Brown University

Yu Cheng
Brown University

Abstract

We study low-rank matrix completion in the *semi-random model*, where each entry (i, j) is observed independently with an unknown probability $p_{i,j} \geq p$, in contrast to the standard model with a uniform probability p. While prior work has shown that nonconvex approach succeeds in the semi-random model for exact observations (Cheng and Ge, 2018), it remains unclear whether similar guarantees extend to more general observation model, such as noisy or one-bit measurements.

In this paper, we give a unified framework for semi-random matrix recovery applicable to a broad family of observation models. Our approach builds on the preprocessing step of Cheng and Ge (2018) to restore regularity conditions that are violated under adversarial sampling, and leverages the primal-dual framework of Zhang et al. (2018b) to obtain near-optimal recovery guarantees. As concrete corollaries, we show that for both noisy and one-bit matrix completion in the semi-random model, after the preprocessing step, every local minimum of the non-convex objective yields an approximate recovery of the ground-truth matrix.

1 INTRODUCTION

We study the problem of recovering a low-rank matrix $\mathbf{M}^\star$ from incomplete and noisy observations $\mathbf{M}$, where $\mathbf{M}^\star$ has rank at most r and $\mathbf{M}$ is supported on a subset of indices Ω. Matrix completion has been well-studied under the uniform observation model, where each entry of $\mathbf{M}^\star$ is revealed independently with some fixed probability p, known as the sampling rate.

The standard assumption that entries are observed uniformly at random may fail to capture realistic scenarios in which the observation pattern follows mild or adversarial non-uniformity. In collaborative filtering, observations might be concentrated on a few blocks of the matrix because specific groups of users are more active or certain products are more popular.

We focus on a model mis-specification known as the *semi-random model* for matrix recovery, introduced by Cheng and Ge (2018). In this model, the observation probabilities are unknown and non-uniform, but are lower-bounded by some probability p. Alternatively, the semi-random model can be viewed as a two-stage process: first, each entry is observed with probability p; then, additional entries from $\mathbf{M}^\star$ may be revealed adversarially.

Most existing results on matrix completion under semi-random observations focus on real-valued or exact observations from the ground-truth matrix. In practice, observations are often noisy, quantized, or binary, even though the underlying signal is continuous. User ratings may be restricted to integers between 1 and 5, or take binary values such as an upvote or downvote. This motivates the study of more general models, specifically noisy matrix completion and one-bit matrix completion. We define the statistical observation models and their loss functions $\mathcal{F}(\mathbf{X})$ over $\mathbb{R}^{d_1 \times d_2}$ below.

Noisy Matrix Completion. Each observed entry is subject to additive independent and identically distributed (i.i.d.) noise. The observed data matrix takes the form:

$$\forall (i,j) \in \Omega : \mathbf{M}_{ij} = \mathbf{M}^\star_{ij} + \mathbf{N}_{ij},$$

where $\mathbf{N}$ is a noise matrix in which each entry is an i.i.d. sub-exponential random variable, with $\mathbb{E}[\mathbf{N}_{ij}] =$

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

0 and $\mathbb{V}\mathrm{ar}[\mathbf{N}_{ij}] = \frac{\nu^2}{d_1 d_2}$ [1]. We consider the least-squares loss function:

$$\mathcal{F}(\mathbf{X}) = \frac{1}{2p} \sum_{(i,j)\in\Omega} (\mathbf{X}_{ij} - \mathbf{M}_{ij})^2. \tag{1}$$

One-Bit Matrix Completion. Originally studied by Davenport et al. (2014), in this setting observations are quantized to binary values in the following way:

$$\forall (i,j) \in \Omega : \mathbf{M}_{ij} = \begin{cases} +1, & \mathbf{M}^\star_{ij} + \mathbf{N}_{ij} \geq 0 \\ -1, & \mathbf{M}^\star_{ij} + \mathbf{N}_{ij} < 0 \end{cases},$$

where $\mathbf{N}$ is a noise matrix in which each entry is i.i.d. with mean $\mathbb{E}[\mathbf{N}_{ij}] = 0$ and $\mathbb{V}\mathrm{ar}[\mathbf{N}_{ij}] = \frac{\nu^2}{d_1 d_2}$. Equivalently, let f be the cumulative distribution function of $-\mathbf{N}_{ij}$. The observations can be formulated using a probit model:

$$\mathbf{M}_{ij} = \begin{cases} +1, & \text{with probability } f(\mathbf{M}^\star_{ij}), \\ -1, & \text{with probability } 1 - f(\mathbf{M}^\star_{ij}) \end{cases}.$$

We consider the log-likelihood loss function $\mathcal{F}$:

$$\mathcal{F}(\mathbf{X}) = -\frac{1}{pd_1 d_2} \sum_{(i,j)\in\Omega} \Big[\mathbb{1}(\mathbf{M}_{ij} = 1) \cdot \log\big(f(\mathbf{X}_{ij})\big) + \mathbb{1}(\mathbf{M}_{ij} = -1) \cdot \log\big(1 - f(\mathbf{X}_{ij})\big) \Big]. \tag{2}$$

These settings have been studied under the uniform observation model but remain largely unexplored in the semi-random setting[2]. Prior techniques from Cheng and Ge (2018), which rely on the quadratic structure of the noiseless loss, do not apply directly to these settings. The noisy and one-bit models introduce non-quadratic losses and non-linear observations. This leads to the central question of our work:

Can we extend the non-convex optimization framework for noisy and one-bit matrix completion with provable guarantees in the semi-random setting?

To avoid the computationally expensive singular value decomposition steps required by convex relaxation methods, we project the empirical loss $\mathcal{F}(\mathbf{X})$ onto the low-rank manifold using the Burer-Monteiro factorization (Burer and Monteiro, 2003), $\mathbf{X} = \mathbf{U}\mathbf{V}^\top$. This yields the following non-convex optimization problem:

$$\min_{\mathbf{U}\in\mathcal{C}_1, \mathbf{V}\in\mathcal{C}_2} \mathcal{F}(\mathbf{U}\mathbf{V}^\top), \tag{3}$$

where $\mathcal{C}_1 \subset \mathbb{R}^{d_1 \times r}$ and $\mathcal{C}_2 \subset \mathbb{R}^{d_2 \times r}$ are constraint sets that control the scaling and incoherence (see Definition 1) of the factors.

We adopt the primal-dual framework proposed by Zhang et al. (2018b), which provides a global landscape analysis via Lagrangian duality. This framework characterizes the stationary points of the non-convex objective and supports algorithmic extensions to various loss functions, including the least-squares and log-likelihood losses defined above. However, their analysis requires uniform observation models and cannot be applied directly to semi-random inputs. Conversely, to address semi-random inputs, Cheng and Ge (2018) establish a connection to spectral graph theory and design a preprocessing algorithm based on spectral sparsification. This procedure reweights the observed entries to achieve spectral similarity to the uniform observation model, eliminating spurious local minima for the noiseless quadratic loss.

In this work, we unify these two theoretical threads to obtain recovery guarantees for noisy and one-bit matrix completion under the semi-random model. We develop a weighted version of the regularity conditions required in the primal-dual analysis, specialized to the reweighted observations. We prove that the approximation error induced by this spectral preprocessing does not degrade the final recovery guarantee. This allows us to extend the primal-dual paradigm to semi-random models in both noisy and one-bit regimes, solving a problem that remains beyond the scope of either prior method individually.

1.1 Our Contributions

We provide a unified algorithmic framework for a broad class of matrix completion problems, including the noisy and one-bit settings, in the semi-random observation model. Specifically, we have the following results.

Theorem 1 (Informal, see Theorem 3). *Let $\mathbf{M}^\star \in \mathbb{R}^{d\times d}$ be a rank-r, β-incoherent ground-truth matrix with largest singular value σ_1. Suppose the observations follow the semi-random model with reveal probability $p \geq \frac{\mathrm{poly}(r,\log d)}{d}$, and are subject to independent sub-exponential noise with variance ν^2. Let $\epsilon = O\left(\sqrt{\log d / pd}\right)$, for a broad class of loss functions (including both continuous least-squares and discrete log-likelihood losses), with high probability, any local min-*

[1] In this paper, for noisy matrix completion problem, we assume the observation noise $\mathbf{N}_{ij}$ is zero-mean sub-exponential, and $\|\mathbf{N}_{ij}\|_{\psi_1} \leq C(\mathbb{E}[\mathbf{N}_{ij}^2])^{1/2}$ for some constant $C > 0$. This includes standard noise models such as Gaussian, Laplace, and bounded random variables.

[2] The work of Kelner et al. (2024) also studies semi-random (noisy) matrix completion, but under a different observation model than the one considered in this paper. See the related work section for further discussion.

imum of the non-convex objective (4) yields a rank-r factorization $\mathbf{U}\mathbf{V}^\top$ *satisfying the following recovery guarantee:*

$$\|\mathbf{U}\mathbf{V}^\top - \mathbf{M}^\star\|_F^2 \le \epsilon \cdot \mathrm{poly}(\beta, \sigma_1, \nu, r, \log d).$$

This recovers prior guarantees for the noiseless quadratic loss and provides a provable, general-purpose algorithmic framework for matrix completion in the semi-random setting with general observations model.

We outline our main technical contributions below:

- **Spectral similarity-based preprocessing for general losses.** We observe that the spectral sparsification-based reweighting method of Cheng and Ge (2018), originally developed for quadratic loss in the noiseless setting, extends to a much broader family of observation models. By scaling the entries to achieve ϵ-spectral similarity to the uniform distribution, we show this preprocessing remains compatible with non-linear loss functions and robust to random noise.
- **Weighted regularity conditions** To bridge the theoretical gap between the reweighted observations and standard non-convex optimization theory, we introduce weighted versions of the Restricted Strong Convexity (RSC), Restricted Strong Smoothness (RSS) (3), and Deviation conditions (4). We show that these vital regularity conditions hold with high probability after the preprocessing (cf. Lemma 1, Lemma 2), incurring only a polylogarithmic overhead in the error rate compared to the uniform model.
- **Primal-dual landscape under semi-random model.** We revisit the primal-dual framework of Zhang et al. (2018b), which was inherently reliant on the uniform observation assumption. Our analysis synthesizes the global landscape properties of this framework with our newly introduced weighted regularity conditions, proving that the absence of spurious local minima even under semi-random inputs.
- **Provable guarantees for noisy and one-bit matrix completion.** As concrete applications of our unified framework, we explicitly derive the recovery bounds for semi-random noisy matrix completion and one-bit matrix completion (Corollary 1 and 2), which to our knowledge has not previously been studied in the semi-random setting.

1.2 Technical Overview

At a high level, our approach combines two previously disjoint components, global reweighting and primal-dual landscape analysis, into a unified framework for matrix completion under the semi-random observation model. We begin by applying the preprocessing algorithm of Cheng and Ge (2018), which rescales the observed entries to produce a weighted sampling operator that is spectrally similar to the uniform observation model. This allows us to recover key structural properties that typically fail under semi-random inputs. We then incorporate the resulting weights into the empirical loss function to form a weighted objective, proving that it satisfies the necessary regularity conditions. Finally, we apply the primal-dual framework of Zhang et al. (2018b) to analyze the global landscape of the resulting constrained non-convex problem. The resulting algorithmic pipeline applies broadly to continuous and discrete observation models. We elaborate on each component below.

Preprocessing. The primal-dual analysis of Zhang et al. (2018b) requires that the loss function satisfies the Restricted Strong Convexity (RSC), Restricted Strong Smoothness (RSS), and Deviation Conditions. These conditions are no longer guaranteed to hold under the semi-random observation model. To restore them, we apply the preprocessing algorithm of Cheng and Ge (2018). The key insight of Cheng and Ge (2018) is to map matrix indices to edges in a bipartite graph. Given a matrix $\mathbf{X} \in \mathbb{R}^{d_1 \times d_2}$, we define a complete bipartite graph $G = (V_1, V_2, E)$ where the vertex sets $|V_1| = d_1$ and $|V_2| = d_2$ correspond to the row and column indices of $\mathbf{X}$, respectively. Let H denote the semi-random subgraph of G generated by first sampling each edge independently with probability p, followed by an adversarial addition of arbitrary edges. The preprocessing algorithm assigns non-negative weights to the edges of H such that the resulting weighted graph is ϵ-spectrally similar to the complete graph G. We represent these weights as a sparse matrix $\mathbf{W} \in \mathbb{R}^{d_1 \times d_2}$, where $\mathbf{W}_{ij} = w_e$ for the corresponding edge $e = (i, j) \in E(H)$, and zero otherwise. The algorithm in Cheng and Ge (2018) outputs a weight matrix $\mathbf{W}$ achieving ϵ-spectral similarity with $\epsilon = O\left(\sqrt{\frac{\log d}{pd}}\right)$. We include technical preliminaries on this preprocessing procedure in Appendix A, and prove additional properties of $\mathbf{W}$ necessary for our unified analysis, formally stated as Lemmas 6 and 7.

Primal-Dual Framework. Following the framework of Zhang et al. (2018b), the non-convex objective is reformulated by stacking the Burer-Monteiro factors $\mathbf{U} \in \mathbb{R}^{d_1 \times r}$ and $\mathbf{V} \in \mathbb{R}^{d_2 \times r}$ into an augmented variable $\mathbf{Z} = \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \in \mathbb{R}^{(d_1+d_2) \times r}$. To enforce the incoherence conditions on the low-rank factors, inequality constraints $h_i(\mathbf{Z}) \le 0$ are introduced for all $i \in [d_1 + d_2]$,

where:

$$h_i(\mathbf{Z}) = \|\mathbf{Z}_{i,:}\|_2^2 - \alpha_i, \quad \text{with } \alpha_i = \begin{cases} \alpha_1, & \text{if } i \le d_1, \\ \alpha_2, & \text{if } i > d_1. \end{cases}$$

Furthermore, a regularization term $\frac{\gamma}{4}\|\mathbf{U}^\top\mathbf{U} - \mathbf{V}^\top\mathbf{V}\|_F^2$ is added to eliminate the scaling ambiguity between $\mathbf{U}$ and $\mathbf{V}$, yielding the constrained primal objective:

$$\min_{\mathbf{Z}=\begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix}, \mathbf{U}\in\mathbb{R}^{d_1\times r}, \mathbf{V}\in\mathbb{R}^{d_2\times r}} \mathcal{F}(\mathbf{U}\mathbf{V}^\top) + \frac{\gamma}{4}\left\|\mathbf{U}^\top\mathbf{U} - \mathbf{V}^\top\mathbf{V}\right\|_{\mathrm{F}}^2, \quad \text{subject to} \quad h_i(\mathbf{Z}) \le 0 \ \forall\, i \in [d_1 + d_2], \tag{4}$$

where $\mathcal{F}$ is the problem-specific empirical loss function. Zhang et al. (2018b) utilize Lagrangian duality and the KKT conditions to characterize the stationary points, establishing that when $\mathcal{F}$ satisfies the aforementioned regularity conditions, the primal objective (4) exhibits no spurious local minima. We formally restate this landscape property as Theorem 2 in the Preliminaries.

Weighted Loss and Regularity Conditions. Consider a general loss function $\mathcal{F}$ that satisfies the requisite regularity conditions under the uniform observation model. To solve the corresponding matrix completion problem under the semi-random model, we incorporate the pre-processing step generated weight matrix $\mathbf{W}$ directly into the empirical risk formulation. Assuming the unweighted loss decomposes entry-wise as $\mathcal{F}(\mathbf{X}) = \sum_{(i,j)\in\Omega} \frac{1}{p}\mathcal{F}_{ij}(\mathbf{X}_{ij})$, we define the *weighted loss* as:

$$\mathcal{F}_{\mathbf{W}}(\mathbf{X}) = \sum_{(i,j)\in\Omega} \mathbf{W}_{ij}\mathcal{F}_{ij}(\mathbf{X}_{ij}). \tag{5}$$

The challenge is to prove that substituting $\mathcal{F}$ with $\mathcal{F}_{\mathbf{W}}$ in the objective (4) preserves the required geometry. We establish the weighted versions of the RSC, RSS, and Deviation Conditions for $\mathcal{F}_{\mathbf{W}}$ in Section 3, thereby formally unlocking the primal-dual framework for semi-random matrix completion.

1.3 Related Work

Matrix Completion. Matrix completion is a fundamental low-rank matrix recovery problem with applications in collaborative filtering and image restoration (Rennie and Srebro, 2005; Zhang et al., 2013). Early theoretical guarantees established that convex relaxation via nuclear norm minimization achieves a sample complexity of $\Omega(dr)$ (Candes and Recht, 2008; Candès and Tao, 2010; Recht, 2011), solvable via semidefinite programming in time $\widetilde{O}(md^{2.5})$ (Jiang et al., 2020; Huang et al., 2022). To improve computational efficiency, extensive research has analyzed non-convex formulations, demonstrating the convergence of gradient descent under careful or random initialization (Keshavan et al., 2010; Jain et al., 2013; Hardt and Wootters, 2014; Chen and Wainwright, 2015; Zhao et al., 2015; Sun and Luo, 2016; Zheng and Lafferty, 2016; Gu et al., 2016; Tu et al., 2016; Wang et al., 2017; Xu et al., 2017; Zhang et al., 2018a; Chen et al., 2020; Gu et al., 2023; Xu et al., 2023). A complementary line of non-convex research characterizes the global optimization landscape, proving the absence of spurious local minima (De Sa et al., 2015; Ge et al., 2016, 2017; Park et al., 2017; Chen and Li, 2019; Zhu et al., 2018, 2021), with Lagrangian-based frameworks providing a unified landscape analysis (Zhang et al., 2018b; Nie et al., 2018). Recently, Kelner et al. (2023b) achieved near-optimal sample complexity and runtime for the non-convex approach.

One-Bit Matrix Completion. Inspired by one-bit compressive sensing (Boufounos and Baraniuk, 2008), Davenport et al. (2014) formulated the one-bit matrix completion problem and established a minimax error bound of $O(\sqrt{\frac{rd}{m}})$. Subsequent work explored this problem through convex relaxations (Cai and Zhou, 2013; Eamaz et al., 2023, 2024). For non-convex formulations, Ni and Gu (2016) and Zhang et al. (2018b) developed algorithms converging at a linear rate that match the minimax statistical error bound up to logarithmic factors. Other extensions include quantized matrix completion with arbitrary corruption (Lan et al., 2014; Shen et al., 2019) and parameter estimation from quantized heavy-tailed data (Chen et al., 2023).

Non-Uniform Model. Beyond the standard uniform observation assumption, existing literature examines matrix completion under non-uniform observation models. This includes deterministic sampling where the sampling probabilities follow problem-specific distributions (Lee and Shraibman, 2013; Heiman et al., 2014; Bhojanapalli and Jain, 2014; Li et al., 2016; Foucart et al., 2020; Ashraphijuo et al., 2017; Bhojanapalli et al., 2014; Pimentel-Alarcón et al., 2016; Meka et al., 2009; Liu et al., 2019). For non-uniform sampling models where each entry is observed with an independent probability p_{ij} strictly bounded from above and below, Chen and Li (2022, 2024) established recovery guarantees under both Frobenius and entry-wise error metrics.

Semi-Random Model. The semi-random model generalizes non-uniform models by dropping strict distribution assumptions, requiring only that each entry is observed with a probability of at least p, followed by adversarial additions. Originally introduced for combinatorial optimization problems such as graph

coloring, planted clique, and clustering (Blum and Spencer, 1995; Feige and Kilian, 2001; Perry and Wein, 2017; Mathieu and Schudy, 2010; Makarychev et al., 2012), it has been extended to low-rank matrix recovery (Cheng and Ge, 2018), sparse linear regression (Kelner et al., 2023a), and matrix sensing (Gao and Cheng, 2024). Closely related to our setting, Kelner et al. (2024) analyzed semi-random noisy matrix completion under a non-adaptive adversary. In contrast to our global spectral reweighting approach, they employ projected gradient descent with iterative reweighting based on a short-flat decomposition. Their method yields an element-wise ℓ_∞-norm error bound. Our results are complementary: we establish optimal Frobenius-norm recovery guarantees, allow for fully adaptive adversaries, and construct a unified analytical framework that supports non-linear observations such as one-bit matrix completion.

2 PRELIMINARIES

2.1 Notation

We write $[n]$ for the set of integers $\{1, ..., n\}$. We use e_i to denote the i^{th} standard basis vector. For a matrix $\mathbf{A}$, we use $\|\mathbf{A}\|_2$, $\|\mathbf{A}\|_{\text{F}}$, and $\|\mathbf{A}\|_{\max}$ for the spectral norm, Frobenius norm, and the maximum absolute entry of $\mathbf{A}$, respectively. We write $\|\mathbf{A}\|_\infty$ and $\|\mathbf{A}\|_1$ for the maximum ℓ_1 norms of the rows and columns of $\mathbf{A}$, respectively. We write $\mathbf{A}_{i,:}$ for the i^{th} row of $\mathbf{A}$, and $\|\mathbf{A}\|_{2\to\infty}$ for the $2 \to \infty$ induced norm of $\mathbf{A}$ (i.e., the maximum ℓ_2 norm of the rows of $\mathbf{A}$).

For matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{d_1 \times d_2}$, we denote their inner product as $\langle \mathbf{A}, \mathbf{B} \rangle = \text{tr}(\mathbf{A}^\top \mathbf{B}) = \sum_{i,j} \mathbf{A}_{ij}\mathbf{B}_{ij}$. Given a set of entries $\Omega \subseteq [d_1] \times [d_2]$, we define $\langle \mathbf{A}, \mathbf{B} \rangle_\Omega = \sum_{(i,j)\in\Omega} \mathbf{A}_{ij}\mathbf{B}_{ij}$; given a weight matrix $\mathbf{W} \in \mathbb{R}^{d_1 \times d_2}$, we define $\langle \mathbf{A}, \mathbf{B} \rangle_{\mathbf{W}} = \sum_{i,j} \mathbf{W}_{ij}\mathbf{A}_{ij}\mathbf{B}_{ij}$. We write $\|\mathbf{A}\|_\Omega^2$ for $\langle \mathbf{A}, \mathbf{A} \rangle_\Omega$, and $\|\mathbf{A}\|_{\mathbf{W}}^2$ for $\langle \mathbf{A}, \mathbf{A} \rangle_{\mathbf{W}}$.

For a scalar random variable X, we denote its sub-exponential norm (the ψ_1-Orlicz norm) as $\|X\|_{\psi_1} = \inf\{t > 0 : \mathbb{E}[\exp(|X|/t)] \le 2\}$. For a random matrix $\mathbf{X}$, denote $\|\mathbf{X}\|_{\psi_1} = \inf\{t > 0 : \mathbb{E}[\exp(\|\mathbf{X}\|_2/t)] \le 2\}$.

2.2 Low-Rank Matrix Completion

Throughout the paper, we denote the unknown ground-truth matrix as $\mathbf{M}^\star \in \mathbb{R}^{d_1 \times d_2}$ and let $d = \max(d_1, d_2)$. We assume $\mathbf{M}^\star$ has rank $r \ll d$. Let $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0$ be the positive singular values of $\mathbf{M}^\star$. The compact Singular Value Decomposition (SVD) of $\mathbf{M}^\star$ is given by $\mathbf{M}^\star = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}^\top$, where $\widetilde{\mathbf{U}} \in \mathbb{R}^{d_1 \times r}$ and $\widetilde{\mathbf{V}} \in \mathbb{R}^{d_2 \times r}$ are orthogonal matrices, and $\mathbf{\Sigma} = \textbf{diag}(\sigma_1, \ldots, \sigma_r)$. Based on this SVD, we can factorize the ground-truth matrix as $\mathbf{M}^\star = \mathbf{U}^\star \mathbf{V}^{\star\top}$, where the low-rank factors are $\mathbf{U}^\star = \widetilde{\mathbf{U}}\mathbf{\Sigma}^{1/2} \in \mathbb{R}^{d_1 \times r}$ and $\mathbf{V}^\star = \widetilde{\mathbf{V}}\mathbf{\Sigma}^{1/2} \in \mathbb{R}^{d_2 \times r}$.

Recovering a low-rank matrix from randomly sampled observations is impossible if its information is highly concentrated in a few entries (e.g., a matrix that is zero everywhere except for a single entry). To avoid this, it is standard in the literature to assume the ground-truth matrix satisfies the incoherence condition.

Definition 1 (Incoherence). *A rank-r matrix $\mathbf{M}^\star \in \mathbb{R}^{d_1 \times d_2}$ with compact SVD $\mathbf{M}^\star = \widetilde{\mathbf{U}}\mathbf{\Sigma}\widetilde{\mathbf{V}}^\top$ is said to be β-incoherent if*

$$\left\|\widetilde{\mathbf{U}}\right\|_{2\to\infty} \le \sqrt{\frac{\beta r}{d_1}} \quad \text{and} \quad \left\|\widetilde{\mathbf{V}}\right\|_{2\to\infty} \le \sqrt{\frac{\beta r}{d_2}}.$$

That is, the maximum row norms of $\widetilde{\mathbf{U}} \in \mathbb{R}^{d_1 \times r}$ and $\widetilde{\mathbf{V}} \in \mathbb{R}^{d_2 \times r}$ are upper bounded by these quantities.

In the paper, we assume $\mathbf{M}^\star$ is β-incoherent, which implies the maximum row norms of $\mathbf{U}^\star$ and $\mathbf{V}^\star$ are bounded by $\alpha_1 = \sqrt{\frac{\beta r \sigma_1}{d_1}}$ and $\alpha_2 = \sqrt{\frac{\beta r \sigma_1}{d_2}}$, respectively. We define the following constraint sets for the Burer-Monteiro factorization:

$$\mathcal{C}_1 =: \left\{ \mathbf{U} \in \mathbb{R}^{d_1 \times r} \;\middle|\; \|\mathbf{U}\|_{2\to\infty} \le \alpha_1 \right\},$$
$$\mathcal{C}_2 =: \left\{ \mathbf{V} \in \mathbb{R}^{d_2 \times r} \;\middle|\; \|\mathbf{V}\|_{2\to\infty} \le \alpha_2 \right\},$$
$$\mathcal{C} =: \left\{ \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \in \mathbb{R}^{(d_1+d_2)\times r} \;\middle|\; \mathbf{U} \in \mathcal{C}_1, \mathbf{V} \in \mathcal{C}_2 \right\}.$$

By definition, the stacked ground-truth factor $\mathbf{Z}^\star := \begin{pmatrix} \mathbf{U}^\star \\ \mathbf{V}^\star \end{pmatrix}$ belongs to the feasible region $\mathcal{C}$. Furthermore, β-incoherence implies the maximum absolute entry of $\mathbf{M}^\star$ is at most $\alpha := \alpha_1\alpha_2$, i.e., $\|\mathbf{M}^\star\|_{\max} \le \alpha$.

2.3 Semi-Random Matrix Completion

We formally define the semi-random model (Cheng and Ge, 2018) considered in this paper.

Definition 2 (Semi-Random Matrix Completion). *Given a ground-truth matrix $\mathbf{M}^\star \in \mathbb{R}^{d_1 \times d_2}$ and a sampling probability $p \in (0, 1]$, the observed index set Ω is generated via the following two-stage process:*

1. *(Initial uniform sampling) An initial set $\Omega_0 \subseteq [d_1] \times [d_2]$ is generated by including each entry (i, j) independently with probability p.*
2. *(Adversarial addition) An adversary, with full knowledge of the Ω_0 and $\mathbf{M}^\star$, selects an arbitrary set of additional entries Ω_{adv}. The final observed index set is $\Omega = \Omega_0 \cup \Omega_{adv}$.*

The algorithm is then provided with the (noisy or one-bit) observations $\mathbf{M}^\star_{ij}$ for all $(i,j) \in \Omega$, but does not know which entries belong to Ω_0 and which to Ω_{adv}. [3]

2.4 Regularity Conditions on Loss Functions

The quadratic loss function is popular due to the isometric property of its Hessian, which a general loss function is no longer guaranteed to satisfy. A common alternative is to enforce a pair of regularity conditions known as the Restricted Strong Convexity (RSC) and Restricted Strong Smoothness (RSS) Conditions.

Definition 3 (RSC and RSS Conditions (Agarwal et al., 2012))**.** *The loss function $\mathcal{F}$ satisfies the Restricted Strong Convexity Condition with parameter μ, and the Restricted Strong Smoothness Condition with parameter L, if for all matrices $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{d_1 \times d_2}$ with rank at most $6r$:*

$$\begin{aligned}\mathcal{F}(\mathbf{X}_1) \geq \mathcal{F}(\mathbf{X}_2) + \langle \nabla \mathcal{F}(\mathbf{X}_2), \mathbf{X}_1 - \mathbf{X}_2 \rangle \\ + \frac{\mu}{2} \|\mathbf{X}_1 - \mathbf{X}_2\|_{\text{F}}^2, \\ \mathcal{F}(\mathbf{X}_1) \leq \mathcal{F}(\mathbf{X}_2) + \langle \nabla \mathcal{F}(\mathbf{X}_2), \mathbf{X}_1 - \mathbf{X}_2 \rangle \\ + \frac{L}{2} \|\mathbf{X}_1 - \mathbf{X}_2\|_{\text{F}}^2.\end{aligned}$$

To bound the statistical error introduced by observation noise, we require a condition on the gradient of the loss function evaluated at the ground-truth $\mathbf{M}^\star$. Originally introduced for high-dimensional regression (Loh and Wainwright, 2012), Zhang et al. (2018b) adapted the following Deviation Condition for the problem of matrix completion.

Definition 4 (Deviation Condition (Zhang et al., 2018b))**.** *The loss function $\mathcal{F}$ satisfies the Deviation Condition if $\|\nabla \mathcal{F}(\mathbf{M}^\star)\|_2 \leq \delta$ with high probability. The deviation bound δ depends on the sampling rate p and the observation noise ν.*

For the one-bit matrix completion problem, we assume the negative noise $-\mathbf{N}_{ij}$ follows a distribution with cumulative distribution function (CDF) f. Let $\tau = \frac{\nu}{\sqrt{d_1 d_2}}$ denote the entry-wise standard deviation. We replace f with its standardized CDF $g(x)$, i.e., $f(x) = g(x/\tau)$. For example, if the noise is Gaussian, $g(z) = \Phi(z)$ is the standard normal CDF.

We define the dimension-free signal-to-noise ratio $\rho = \frac{\alpha}{\tau} = \frac{\beta r \sigma_1}{\nu}$. Because we assume $\|\mathbf{M}^\star\|_{\max} \leq \alpha$, the arguments evaluated by g are bounded by $|x| \leq \rho$. To ensure the negative log-likelihood has sufficient curvature, we adopt the steepness assumption (Davenport et al., 2014) on the standardized distribution function. We control this steepness via the quantity s_ρ:

$$s_\rho = \sup_{|x| \leq \rho} \frac{|g'(x)|}{g(x)(1-g(x))}, \tag{6}$$

which is a fixed constant given ρ and the noise distribution g.

2.5 Error Guarantees of the Primal-Dual Framework

We restate the main error bound of the primal-dual framework, which bounds the global reconstruction error under the aforementioned regularity conditions.

Theorem 2 (Theorem 3.8 of Zhang et al. (2018b))**.** *Assume the loss function $\mathcal{F}$ satisfies the RSC and RSS (Definition 3) with parameters μ and L such that $L/\mu \in (1, 18/17)$, and the Deviation Condition (Definition 4) with bound δ. For all local minima $\mathbf{Z} = \begin{pmatrix}\mathbf{U} \\ \mathbf{V}\end{pmatrix}$ of the objective function in Equation (4), with high probability, the reconstruction error satisfies*

$$\left\|\mathbf{U}\mathbf{V}^\top - \mathbf{M}^\star\right\|_{\text{F}}^2 \leq \Gamma r \delta^2,$$

where Γ is a constant depending on L/μ. Specifically, $\Gamma = \frac{10}{(10\mu - 9L - \gamma - 3c)c}$ where $c = \frac{18\mu - 17L}{12}$, and the regularization parameter γ in Equation (4) is chosen such that $\mu - L/2 \leq \gamma < \min\{(22\mu - 19L)/4, (3L - 2\mu)/2\}$.

3 OUR RESULTS

3.1 Weighted Regularity Conditions

We first analyze the consequence of spectral preprocessing on the structural conditions required by the primal-dual optimization framework. Consider a matrix completion problem that is well-behaved under the uniform observation model. Specifically, suppose its objective function (4) employs a loss function $\mathcal{F}$ that satisfies both the RSC and RSS Conditions (3) with parameters μ and L, as well as the Deviation Condition (4) with bound $\mathbf{\Delta}$.

For the semi-random model with sampling probability at least p, we first run the preprocessing algorithm to obtain a weight matrix $\mathbf{W}$ achieving ϵ-spectral similarity with $\epsilon = O\left(\sqrt{\frac{\log d}{pd}}\right)$. We then apply these weights to the empirical risk to construct the weighted objective $\mathcal{F}_{\mathbf{W}}$.

To prove that the weighted loss $\mathcal{F}_{\mathbf{W}}$ inherits the necessary regularity properties, we isolate two structural conditions on the base uniform loss $\mathcal{F}$. We state them below, and will subsequently show in Appendix C.1

[3] We assume the noise is generated *after* fixing Ω, i.e., the adversary is oblivious to the noise in each entry.

and C.2 that both the noisy least-squares and one-bit log-likelihood losses satisfy them.

First, to establish the RSC and RSS conditions for the weighted loss, it is sufficient to bound the second-order Taylor expansion of the unweighted loss. We require the uniform loss to admit an entry-wise quadratic bound on its approximation residual, allowing us to absorb the reweighting factors into the curvature bounds.

Condition 1 (Entry-wise Taylor Decomposition)**.** *Suppose the uniform loss function decomposes as* $\mathcal{F}(\mathbf{X}) = \sum_{(i,j)\in\Omega} \frac{1}{p}\mathcal{F}_{ij}(\mathbf{X}_{ij})$ *and satisfies the RSC and RSS conditions with parameters* μ *and* L*. For all feasible matrices* $\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{C}$*, the second-order Taylor expansion residual satisfies:*

$$\begin{aligned}&\mathcal{F}(\mathbf{X}_1) - \mathcal{F}(\mathbf{X}_2) - \langle \nabla\mathcal{F}(\mathbf{X}_2), \mathbf{X}_1 - \mathbf{X}_2\rangle \\ &\quad= \frac{1}{2}\sum_{(i,j)\in\Omega} \frac{1}{p}\mathbf{K}_{ij}(\mathbf{X}_{1,ij} - \mathbf{X}_{2,ij})^2,\end{aligned}$$

for some data-dependent scalars $\mathbf{K}_{ij}$ *such that* $\mu \le \mathbf{K}_{ij} \le L$ *for all* $(i,j) \in \Omega$*.*

Second, to establish the Deviation Condition for $\mathcal{F}_{\mathbf{W}}$, we must control the statistical noise evaluated at the ground truth. Because the preprocessing algorithm alters the observation distribution via $\mathbf{W}$, we require the gradient of the unweighted loss to decouple into independent components. This allows us to apply matrix concentration inequalities to the weighted gradient.

Condition 2 (Gradient Sub-exponentiality)**.** *For the uniform loss function* $\mathcal{F}(\mathbf{X}) = \sum_{(i,j)\in\Omega} \frac{1}{p}\mathcal{F}_{ij}(\mathbf{X}_{ij})$*, its gradient evaluated at the ground-truth matrix* $\mathbf{M}^\star$ *takes the form:*

$$\nabla\mathcal{F}(\mathbf{M}^\star) = \sum_{(i,j)\in\Omega} \frac{1}{p} b_{ij}(\mathbf{M}^\star) e_i e_j^\top,$$

where, conditioned on $\mathbf{M}^\star$*, each* $b_{ij}(\mathbf{M}^\star)$ *is an independent, zero-mean sub-exponential random variable with variance bounded by* $O(\nu^2/d^2)$*, that depends on the distribution of the observation noise* $\mathbf{N}_{ij}$*.*

We now present the core technical lemmas of our unified framework. These lemmas bridge between the preprocessing and the optimization landscape. The first lemma shows that Condition 1 is sufficient to establish the weighted RSC and RSS conditions, with only a slight perturbation to the condition number.

Lemma 1 (Weighted RSC and RSS Conditions)**.** *Suppose the unweighted loss* $\mathcal{F}$ *satisfies Condition 1 with parameters* μ *and* L*. Let* $\mathbf{W}$ *be the weight matrix produced by the preprocessing algorithm achieving* ϵ*-spectral similarity. For any sufficiently small constant* $c \in (0,1)$ *and all matrices* $\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{C}$*, if* $\|\mathbf{X}_1 - \mathbf{X}_2\|_{\mathrm{F}}^2 \ge O(\beta^2 r^2 \sigma_1^2 \epsilon)$*, then the weighted loss* $\mathcal{F}_{\mathbf{W}}$ *satisfies the RSC and RSS conditions with parameters* $\mu_{\mathbf{W}} = (1-c)\mu$ *and* $L_{\mathbf{W}} = (1+c)L$*.*

Proof. Let $\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{C}$ and denote the difference as $\mathbf{\Delta} = \mathbf{X}_1 - \mathbf{X}_2$. By Condition 1, applying the weights $\mathbf{W}_{ij}$ to the empirical loss yields the following second-order expansion for $\mathcal{F}_{\mathbf{W}}$:

$$\begin{aligned}&\mathcal{F}_{\mathbf{W}}(\mathbf{X}_1) - \mathcal{F}_{\mathbf{W}}(\mathbf{X}_2) - \langle \nabla\mathcal{F}_{\mathbf{W}}(\mathbf{X}_2), \mathbf{\Delta}\rangle \\ &= \frac{1}{2}\sum_{(i,j)\in\Omega} \mathbf{K}_{ij}\mathbf{W}_{ij}\mathbf{\Delta}_{ij}^2.\end{aligned}$$

By the assumption that $\mathbf{K}_{ij} \le L$ for all $(i,j) \in \Omega$:

$$\begin{aligned}\frac{1}{2}\sum_{(i,j)\in\Omega} \mathbf{K}_{ij}\mathbf{W}_{ij}\mathbf{\Delta}_{ij}^2 &\le \frac{L}{2}\sum_{(i,j)\in\Omega} \mathbf{W}_{ij}\mathbf{\Delta}_{ij}^2 = \frac{L}{2}\|\mathbf{\Delta}\|_{\mathbf{W}}^2 \\ &\le \frac{L}{2}(1+c)\|\mathbf{\Delta}\|_{\mathrm{F}}^2.\end{aligned}$$

The last inequality follows directly from Lemma 6. Specifically, because $\|\mathbf{\Delta}\|_{\mathrm{F}}^2 \ge O(\beta^2 r^2 \sigma_1^2 \epsilon)$, we have the spectral approximation guarantee $|\,\|\mathbf{\Delta}\|_{\mathbf{W}}^2 - \|\mathbf{\Delta}\|_{\mathrm{F}}^2\,| \le c\,\|\mathbf{\Delta}\|_{\mathrm{F}}^2$ for a small constant c. The lower bound for the RSC condition follows symmetrically. □

The second key lemma bounds the deviation of the weighted gradient. We show that after applying preprocessing, this overhead is controlled.

Lemma 2 (Weighted Deviation Condition)**.** *Suppose the unweighted loss* $\mathcal{F}$ *satisfies Condition 2. Let* $\mathcal{F}_{\mathbf{W}}$ *be the weighted loss weighted by the preprocessing algorithm achieving* ϵ*-spectral similarity. The weighted deviation bound satisfies* $\mathbf{\Delta}_{\mathbf{W}}^2 = \|\nabla\mathcal{F}_{\mathbf{W}}(\mathbf{M}^\star)\|_2^2 \le O(\nu^2\epsilon\log d)$ *with high probability.*

Proof. By Condition 2, we can write the weighted gradient at the ground truth as a sum of independent random matrices,

$$\nabla\mathcal{F}_{\mathbf{W}}(\mathbf{M}^\star) = \sum_{(i,j)\in\Omega} \mathbf{W}_{ij} b_{ij}(\mathbf{M}^\star) e_i e_j^\top =: \sum_{(i,j)\in\Omega} \mathbf{Y}_{ij}.$$

We bound its operator norm using the Matrix Bernstein inequality (Fact 2). Let $d = \max(d_1, d_2)$, we first

compute the required parameter variance σ^2,

$$\begin{aligned}\left\|\sum_{(i,j)\in\Omega}\mathbb{E}\left[\mathbf{Y}_{ij}\mathbf{Y}_{ij}^\top\right]\right\|_2 &\le \left\|\sum_{(i,j)\in\Omega} O\left(\frac{\nu^2}{d^2}\right)\mathbf{W}_{ij}^2 e_i e_i^\top\right\|_2 \\ &= O\left(\frac{\nu^2}{d^2}\right)\max_i\sum_j \mathbf{W}_{ij}^2 \\ &\le O\left(\frac{\nu^2}{d^2}\right)\|\mathbf{W}\|_{\max}\|\mathbf{W}\|_\infty \\ &= O\left(\frac{\nu^2}{d^2}\epsilon\sqrt{d_1 d_2}d_2\right) \le O(\nu^2\epsilon).\end{aligned}$$

The last equality is due to Lemmas 3 and 7. A symmetric calculation yields an almost identical upper bound,

$$\left\|\sum_{(i,j)\in\Omega}\mathbb{E}\left[\mathbf{Y}_{ij}^\top\mathbf{Y}_{ij}\right]\right\|_2 \le O\left(\frac{\nu^2}{d^2}\epsilon\sqrt{d_1 d_2}d_1\right) \le O(\nu^2\epsilon).$$

We next bound the sub-exponential norm parameter R. Note that each summand can be written as

$$\mathbf{Y}_{ij} = b_{ij}(\mathbf{M}^\star)(\mathbf{W}_{ij}e_i e_j^\top).$$

The randomness is from the sub-exponential scalar $b_{ij}(\mathbf{M}^\star)$ while the matrix direction is deterministic. The sub-exponential parameter is therefore bounded by the product of the scalar ψ_1-norm and the operator norm of deterministic matrix,

$$\begin{aligned}R = \|\mathbf{Y}_{ij}\|_{\psi_1} &= \|b_{ij}(\mathbf{M}^\star)\|_{\psi_1}\cdot\left\|\mathbf{W}_{ij}e_i e_j^\top\right\|_2 \\ &\le O\left(\frac{\nu}{d}\right)\cdot\|\mathbf{W}\|_{\max} = O(\nu\epsilon).\end{aligned}$$

We now apply Fact 2 by choosing $t = c\sigma\sqrt{\log d}$ for a sufficiently large constant c. To obtain a high-probability bound of $d^{-\Omega(1)}$, we verify the condition that the sub-Gaussian tail (t^2/σ^2) dominates the sub-exponential tail (t/R), we require $Rt \le O(\sigma^2)$,

$$\frac{Rt}{\sigma^2} = O\left(\frac{\nu\epsilon\cdot\nu\sqrt{\epsilon\log d}}{\nu^2\epsilon}\right) = O(\sqrt{\epsilon\log d}).$$

Recall that $\epsilon = O(\sqrt{\log d/(pd)})$. The domination condition requires $\epsilon\log d \le O(1)$, which is equivalent to $pd \ge \Omega(\log^3 d)$. This is satisfied by our base sampling assumption $p \ge \mathrm{poly}(r,\log d)/d$. The probability of deviation exceeding t is bounded by $2d\cdot d^{-c^2/2} \le d^{-\Omega(1)}$. Plugging in $t^2 = O(\sigma^2\log d)$ yields the desired bound $\mathbf{\Delta}_{\mathbf{W}}^2 \le O(\nu^2\epsilon\log d)$. □

3.2 Error Bounds for Semi-Random Matrix Completion

With the weighted RSC, RSS, and deviation conditions established, we now combine them to prove the global optimality of the primal-dual framework against semi-random adversaries.

Theorem 3 (Semi-Random Error Bound). *Consider a matrix completion problem where the loss $\mathcal{F}$ satisfies the assumptions of Theorem 2, as well as Conditions 1 and 2. Suppose the observation set Ω is generated from the semi-random model with base sampling probability $p \ge \mathrm{poly}(r,\log d)/d$. Let $\mathcal{F}_{\mathbf{W}}$ denote the weighted objective (4) obtained via ϵ-spectral reweighting with $\epsilon = O(\sqrt{\log d/(pd)})$. Then, with high probability, any local minimum $\mathbf{Z} = \begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix}$ of the constrained optimization problem (4) with objective function $\mathcal{F}_{\mathbf{W}}$ satisfies the error bound*

$$\left\|\mathbf{U}\mathbf{V}^\top - \mathbf{M}^\star\right\|_{\mathrm{F}}^2 \le \epsilon\cdot\max\left\{O(\beta^2 r^2\sigma_1^2),\ O(\Gamma r\nu^2\log d)\right\},$$

where $\Gamma > 0$ is a constant depending only on $L_{\mathbf{W}}/\mu_{\mathbf{W}}$.

Proof. If $\left\|\mathbf{U}\mathbf{V}^\top - \mathbf{M}^\star\right\|_{\mathrm{F}}^2 \le O(\beta^2 r^2\sigma_1^2\epsilon)$, the bound holds trivially. Suppose instead $\left\|\mathbf{U}\mathbf{V}^\top - \mathbf{M}^\star\right\|_{\mathrm{F}}^2 > O(\beta^2 r^2\sigma_1^2\epsilon)$. Because the local minimum $\mathbf{Z} \in \mathcal{C}$ and the ground truth $\mathbf{Z}^\star \in \mathcal{C}$, the prerequisite for Lemma 1 is satisfied. This implies that $\mathcal{F}_{\mathbf{W}}$ satisfies the RSC and RSS conditions with condition number $L_{\mathbf{W}}/\mu_{\mathbf{W}} \le \frac{1+c}{1-c}(L/\mu)$. By Lemma 2, the deviation bound satisfies $\mathbf{\Delta}_{\mathbf{W}}^2 \le O(\nu^2\epsilon\log d)$ with high probability. Applying Theorem 2 gives

$$\left\|\mathbf{U}\mathbf{V}^\top - \mathbf{M}^\star\right\|_{\mathrm{F}}^2 \le \Gamma r\mathbf{\Delta}_{\mathbf{W}}^2 \le O(\Gamma r\nu^2\epsilon\log d).$$

The final bound follows by taking the maximum over the two mutually exclusive conditions. □

Remark 1. *The maximum in Theorem 3 separates two sources of error. The first term, $O(\beta^2 r^2\sigma_1^2\epsilon)$, is the error introduced by the spectral preprocessing. This term dominates the recovery error when the observation noise is small. The second term, $O(\Gamma r\nu^2\epsilon\log d)$, is the statistical error by the observation noise. When the noise is large enough, the bound depends on the noise variance ν^2 and is independent of the matrix parameters σ_1 and β.*

4 APPLICATIONS TO NOISY AND 1-BIT MATRIX COMPLETION

We now apply the semi-random error bound (Theorem 3) for two specific observation models: noisy matrix completion and one-bit matrix completion. The uniform observation variants of these problems were analyzed via the primal-dual framework in Zhang et al. (2018b). Here, we state the robust recovery guarantees under the semi-random model as corollaries. Full proofs are deferred to Appendices C.1 and C.2.

Corollary 1 (Semi-Random Noisy Matrix Completion)**.** *Suppose the ground truth matrix* $\mathbf{M}^\star \in \mathbb{R}^{d_1 \times d_2}$ *has rank* r *and is* β*-incoherent. Consider the semi-random observation model with sampling probability at least* $p = \Omega\left(\frac{r^4 \log^3 d}{d}\right)$*. Assume the observation noise* $\mathbf{N}_{ij}$ *is independent and identically distributed with zero mean and variance* $\frac{\nu^2}{d_1 d_2}$*, and satisfies the sub-exponential tail condition. Let* $\mathcal{F}_{\mathbf{W}}$ *be the weighted objective constructed via the preprocessing algorithm. Then, with high probability, any local minimum* $\mathbf{Z} = \begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix}$ *of the constrained optimization problem satisfies*

$$\left\|\mathbf{U}\mathbf{V}^\top - \mathbf{M}^\star\right\|_{\mathrm{F}}^2 \le \max\left\{ O\left(\beta^2\sigma_1^2\sqrt{\frac{r^4 \log d}{pd}}\right), O\left(\nu^2\sqrt{\frac{r^2 \log^3 d}{pd}}\right)\right\}.$$

Corollary 2 (Semi-Random One-Bit Matrix Completion)**.** *Suppose the ground truth matrix* $\mathbf{M}^\star \in \mathbb{R}^{d_1 \times d_2}$ *has rank* r *and is* β*-incoherent. Consider the one-bit matrix completion problem under the semi-random observation model with sampling probability at least* $p = \Omega\left(\frac{r^4 \log^3 d}{d}\right)$*. Assume the observation CDF satisfies the steepness condition (6) with parameter* s_ρ*. Let* $\mathcal{F}_{\mathbf{W}}$ *be the weighted objective constructed via the preprocessing algorithm. Then, with high probability, any local minimum* $\mathbf{Z} = \begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix}$ *of the constrained optimization problem satisfies*

$$\left\|\mathbf{U}\mathbf{V}^\top - \mathbf{M}^\star\right\|_{\mathrm{F}}^2 \le \max\left\{ O\left(\beta^2\sigma_1^2\sqrt{\frac{r^4 \log d}{pd}}\right), O\left(s_\rho^2\sqrt{\frac{r^2 \log^3 d}{pd}}\right)\right\}.$$

Remark 2. *Under the uniform observation model, Zhang et al. (2018b) requires sampling probability* $p = \Omega\left(\frac{r^2 \log d}{d}\right)$ *for both problems, yielding error terms that scale as* $O\left(\frac{r^2 \log d}{pd}\right)$ *and* $O\left(\frac{r \log d}{pd}\right)$*. By comparison, extending these recovery guarantees to the semi-random model incurs an additional factor of* $O(r^2 \log^2 d)$ *in both the sample complexity and the error bounds.*

Acknowledgments

We thank Rong Ge for helpful conversations. Xing Gao was supported in part by NSF Awards ECCS-2217023 and CCF-2307106. Binhao Chen and Yu Cheng were supported in part by NSF Award CCF-2307106.

References

Agarwal, A., Negahban, S., and Wainwright, M. J. (2012). Fast global convergence of gradient methods for high-dimensional statistical recovery. *The Annals of Statistics*, pages 2452–2482.

Ashraphijuo, M., Aggarwal, V., and Wang, X. (2017). On deterministic sampling patterns for robust low-rank matrix completion. *IEEE Signal Processing Letters*, 25(3):343–347.

Bhojanapalli, S. and Jain, P. (2014). Universal matrix completion. In *International Conference on Machine Learning*, pages 1881–1889. PMLR.

Bhojanapalli, S., Jain, P., and Sanghavi, S. (2014). Tighter low-rank approximation via sampling the leveraged element. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, pages 902–920. SIAM.

Blum, A. and Spencer, J. (1995). Coloring random and semi-random k-colorable graphs. *Journal of Algorithms*, 19(2):204–234.

Boufounos, P. T. and Baraniuk, R. G. (2008). 1-bit compressive sensing. In *2008 42nd Annual Conference on Information Sciences and Systems*, pages 16–21. IEEE.

Burer, S. and Monteiro, R. D. (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357.

Cai, T. and Zhou, W.-X. (2013). A max-norm constrained minimization approach to 1-bit matrix completion. *Journal of Machine Learning Research*, 14:3619–3647.

Candes, E. J. and Recht, B. (2008). Exact low-rank matrix completion via convex optimization. In *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, pages 806–812. IEEE.

Candès, E. J. and Tao, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE transactions on information theory*, 56(5):2053–2080.

Chen, J. and Li, X. (2019). Model-free nonconvex matrix completion: Local minima analysis and applications in memory-efficient kernel pca. *Journal of Machine Learning Research*, 20(142):1–39.

Chen, J., Ng, M. K., and Wang, D. (2023). Quantizing heavy-tailed data in statistical estimation:(near) minimax rates, covariate quantization, and uniform recovery. *IEEE Transactions on Information Theory*, 70(3):2003–2038.

Chen, Y., Chi, Y., Fan, J., Ma, C., and Yan, Y. (2020). Noisy matrix completion: Understanding

statistical guarantees for convex relaxation via nonconvex optimization. *SIAM journal on optimization*, 30(4):3098–3121.

Chen, Y. and Li, X. (2022). Determining the number of factors in high-dimensional generalized latent factor models. *Biometrika*, 109(3).

Chen, Y. and Li, X. (2024). A note on entrywise consistency for mixed-data matrix completion. *Journal of Machine Learning Research*, 25(343):1–66.

Chen, Y. and Wainwright, M. J. (2015). Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. *arXiv preprint arXiv:1509.03025*.

Cheng, Y. and Ge, R. (2018). Non-convex matrix completion against a semi-random adversary. In *Conference On Learning Theory*, pages 1362–1394. PMLR.

Davenport, M. A., Plan, Y., Van Den Berg, E., and Wootters, M. (2014). 1-bit matrix completion. *Information and Inference: A Journal of the IMA*, 3(3):189–223.

De Sa, C., Re, C., and Olukotun, K. (2015). Global convergence of stochastic gradient descent for some non-convex matrix problems. In *International conference on machine learning*, pages 2332–2341. PMLR.

Eamaz, A., Yeganegi, F., and Soltanalian, M. (2023). One-bit matrix completion with time-varying sampling thresholds. In *2023 International Conference on Sampling Theory and Applications (SampTA)*, pages 1–5. IEEE.

Eamaz, A., Yeganegi, F., and Soltanalian, M. (2024). Matrix completion from one-bit dither samples. *IEEE Transactions on Signal Processing*.

Feige, U. and Kilian, J. (2001). Heuristics for semi-random graph problems. *Journal of Computer and System Sciences*, 63(4):639–671.

Foucart, S., Needell, D., Pathak, R., Plan, Y., and Wootters, M. (2020). Weighted matrix completion from non-random, non-uniform sampling patterns. *IEEE Transactions on Information Theory*, 67(2):1264–1290.

Gao, X. and Cheng, Y. (2024). Robust matrix sensing in the semi-random model. *Advances in Neural Information Processing Systems*, 36.

Ge, R., Jin, C., and Zheng, Y. (2017). No spurious local minima in nonconvex low rank problems: A unified geometric analysis. In *International Conference on Machine Learning*, pages 1233–1242. PMLR.

Ge, R., Lee, J. D., and Ma, T. (2016). Matrix completion has no spurious local minimum. *Advances in neural information processing systems*, 29.

Gu, Q., Wang, Z. W., and Liu, H. (2016). Low-rank and sparse structure pursuit via alternating minimization. In *Artificial Intelligence and Statistics*, pages 600–609. PMLR.

Gu, Y., Song, Z., Yin, J., and Zhang, L. (2023). Low rank matrix completion via robust alternating minimization in nearly linear time. *arXiv preprint arXiv:2302.11068*.

Hardt, M. and Wootters, M. (2014). Fast matrix completion without the condition number. In *Conference on learning theory*, pages 638–678. PMLR.

Heiman, E., Schechtman, G., and Shraibman, A. (2014). Deterministic algorithms for matrix completion. *Random Structures & Algorithms*, 45(2):306–317.

Huang, B., Jiang, S., Song, Z., Tao, R., and Zhang, R. (2022). Solving sdp faster: A robust ipm framework and efficient implementation. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 233–244. IEEE.

Jain, P., Netrapalli, P., and Sanghavi, S. (2013). Low-rank matrix completion using alternating minimization. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 665–674.

Jiang, H., Kathuria, T., Lee, Y. T., Padmanabhan, S., and Song, Z. (2020). A faster interior point method for semidefinite programming. In *2020 IEEE 61st annual symposium on foundations of computer science (FOCS)*, pages 910–918. IEEE.

Kelner, J., Li, J., Liu, A., Sidford, A., and Tian, K. (2024). Semi-random matrix completion via flow-based adaptive reweighting. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Kelner, J., Li, J., Liu, A. X., Sidford, A., and Tian, K. (2023a). Semi-random sparse recovery in nearly-linear time. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 2352–2398. PMLR.

Kelner, J. A., Li, J., Liu, A., Sidford, A., and Tian, K. (2023b). Matrix completion in almost-verification time. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2102–2128. IEEE.

Keshavan, R. H., Montanari, A., and Oh, S. (2010). Matrix completion from a few entries. *IEEE transactions on information theory*, 56(6):2980–2998.

Lan, A. S., Studer, C., and Baraniuk, R. G. (2014). Matrix recovery from quantized and corrupted measurements. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4973–4977. IEEE.

Lee, T. and Shraibman, A. (2013). Matrix completion from any given set of observations. *Advances in Neural Information Processing Systems*, 26.

Li, Y., Liang, Y., and Risteski, A. (2016). Recovery guarantee of weighted low-rank approximation via alternating minimization. In *International Conference on Machine Learning*, pages 2358–2367. PMLR.

Liu, G., Liu, Q., Yuan, X.-T., and Wang, M. (2019). Matrix completion with deterministic sampling: Theories and methods. *IEEE transactions on pattern analysis and machine intelligence*, 43(2):549–566.

Loh, P.-L. and Wainwright, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of Statistics*, 40(3):1637–1664.

Makarychev, K., Makarychev, Y., and Vijayaraghavan, A. (2012). Approximation algorithms for semi-random partitioning problems. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 367–384.

Mathieu, C. and Schudy, W. (2010). Correlation clustering with noisy input. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 712–728. SIAM.

Meka, R., Jain, P., and Dhillon, I. (2009). Matrix completion from power-law distributed samples. *Advances in neural information processing systems*, 22.

Negahban, S. and Wainwright, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *The Journal of Machine Learning Research*, 13(1):1665–1697.

Ni, R. and Gu, Q. (2016). Optimal statistical and computational rates for one bit matrix completion. In *Artificial Intelligence and Statistics*, pages 426–434. PMLR.

Nie, F., Hu, Z., and Li, X. (2018). Matrix completion based on non-convex low-rank approximation. *IEEE Transactions on Image Processing*, 28(5):2378–2388.

Park, D., Kyrillidis, A., Carmanis, C., and Sanghavi, S. (2017). Non-square matrix sensing without spurious local minima via the burer-monteiro approach. In *Artificial Intelligence and Statistics*, pages 65–74. PMLR.

Perry, A. and Wein, A. S. (2017). A semidefinite program for unbalanced multisection in the stochastic block model. In *2017 International Conference on Sampling Theory and Applications (SampTA)*, pages 64–67. IEEE.

Pimentel-Alarcón, D. L., Boston, N., and Nowak, R. D. (2016). A characterization of deterministic sampling patterns for low-rank matrix completion. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):623–636.

Recht, B. (2011). A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12(12).

Rennie, J. D. and Srebro, N. (2005). Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the 22nd international conference on Machine learning*, pages 713–719.

Shen, J., Awasthi, P., and Li, P. (2019). Robust matrix completion from quantized observations. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 397–407. PMLR.

Sun, R. and Luo, Z.-Q. (2016). Guaranteed matrix completion via non-convex factorization. *IEEE Transactions on Information Theory*, 62(11):6535–6579.

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434.

Tu, S., Boczar, R., Simchowitz, M., Soltanolkotabi, M., and Recht, B. (2016). Low-rank solutions of linear matrix equations via procrustes flow. In *International Conference on Machine Learning*, pages 964–973. PMLR.

Wang, L., Zhang, X., and Gu, Q. (2017). A unified computational and statistical framework for nonconvex low-rank matrix estimation. In *Artificial Intelligence and Statistics*, pages 981–990. PMLR.

Xu, P., Ma, J., and Gu, Q. (2017). Speeding up latent variable gaussian graphical model estimation via nonconvex optimization. *Advances in Neural Information Processing Systems*, 30.

Xu, X., Shen, Y., Chi, Y., and Ma, C. (2023). The power of preconditioning in overparameterized low-rank matrix sensing. In *International Conference on Machine Learning*, pages 38611–38654. PMLR.

Zhang, H., He, W., Zhang, L., Shen, H., and Yuan, Q. (2013). Hyperspectral image restoration using low-rank matrix recovery. *IEEE transactions on geoscience and remote sensing*, 52(8):4729–4743.

Zhang, X., Wang, L., and Gu, Q. (2018a). A unified framework for nonconvex low-rank plus sparse matrix recovery. In *International Conference on Artificial Intelligence and Statistics*, pages 1097–1107. PMLR.

Zhang, X., Wang, L., Yu, Y., and Gu, Q. (2018b). A primal-dual analysis of global optimality in nonconvex low-rank matrix recovery. In *International conference on machine learning*, pages 5862–5871. PMLR.

Zhao, T., Wang, Z., and Liu, H. (2015). A nonconvex optimization framework for low rank matrix estimation. *Advances in Neural Information Processing Systems*, 28.

Zheng, Q. and Lafferty, J. (2016). Convergence analysis for rectangular matrix completion using burer-monteiro factorization and gradient descent. *arXiv preprint arXiv:1605.07051*.

Zhu, Z., Li, Q., Tang, G., and Wakin, M. B. (2018). Global optimality in low-rank matrix optimization. *IEEE Transactions on Signal Processing*, 66(13):3614–3628.

Zhu, Z., Li, Q., Tang, G., and Wakin, M. B. (2021). The global optimization geometry of low-rank matrix optimization. *IEEE Transactions on Information Theory*, 67(2):1308–1331.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   We defined the mathematical setting in Section 1, and stated our assumptions in Section 2 and 3.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Not Applicable]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   We stated our assumptions in Section 2 and 3.
   (b) Complete proofs of all theoretical results. [Yes]
   Proofs are either contained in the main paper, or deferred to the Appendix.
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Not Applicable]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]

(c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]

(d) Information about consent from data providers/curators. [Not Applicable]

(e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

(a) The full text of instructions given to participants and screenshots. [Not Applicable]

(b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

(c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Semi-Random Noisy and One-Bit Matrix Completion via Nonconvex Optimization: Supplementary Materials

A Prior Results on Spectral Pre-processing

We include some relevant results from Cheng and Ge (2018) here. First we need to introduce some additional notation.

For a weighted undirected graph $G = (V, E, w)$ with n vertices and weights $w_e \geq 0$ for each edge $e = (i, j)$, let $D \in \mathbb{R}^{n\times n}$ be a diagonal matrix containing the weighted degree of each vertex, i.e. $D_{ii} = \sum_{(i,j)\in E} w_{(i,j)}$. Let $A \in \mathbb{R}^{n\times n}$ be the adjacency matrix of G, i.e. $A_{ij} = A_{ji} = w_{(i,j)}$. The Laplacian matrix of G is defined as $L =: D - A$. Fix some arbitrary orientation for each edge $e = (i, j) \in E$, we can represent it with vector $b_e \in \mathbb{R}^n$ where $b_e[i] = 1$ and $b_e[j] = -1$, and $L = \sum_{e\in E} w_e b_e b_e^\top$. Symmetric matrix $A \in \mathbb{R}^{d\times d}$ is called positive semidefinite (PSD) if $x^\top Ax \geq 0 \, \forall x \in \mathbb{R}^d$, and we write $A \preccurlyeq B$ if A and B have the same dimension and $B - A$ is PSD.

The preprocessing algorithm assigns weights to the edges of the semi-random graph H so that the resulting weighted graph is ϵ-spectrally-similar to a complete graph G. For sufficiently large p, Cheng and Ge (2018) proves the existence of weights w_e^* with spectral similarity parameter $\epsilon_0 = O\left(\sqrt{\frac{\log d}{pd}}\right)$, so that we can Theorem 4 with L as the Laplacian of G, E as the edge set of H, and $\epsilon = \epsilon_0$.

Theorem 4 (Semi-Random Preprocessing, Theorem 3.1 (Cheng and Ge, 2018)). *Fix $\epsilon \geq 0$, $\epsilon_0 \leq 1/10$, and let L be the Laplacian of a $d_1 \times d_2$ bipartite graph. Given a set of m edges E, let vector $b_e \in \mathbb{R}^{d_1+d_2}$ represent an edge $e \in E$, where $e = (i, j)$ for some $i \in [d_1]$, $j \in [d_2]$, and $b_e[i] = 1, b_e[j] = -1$. Assume there exists weights $w_e^* \geq 0$ such that*

$$(1 - \epsilon_0)L \preccurlyeq \sum_{e\in E} w_e^* b_e b_e^\top \preccurlyeq (1 + \epsilon_0)L.$$

We can find a set of weights $w_e \geq 0$ in time $\widetilde{O}(m/\epsilon^{O(1)})$, such that with high probability,

$$(1 - O(\epsilon_0) - \epsilon)L \preccurlyeq \sum_{e\in E} w_e b_e b_e^\top \preccurlyeq L.$$

Lemma 3 (Spectral properties of the weight matrix, Corollary 3.4 (Cheng and Ge, 2018)). *For a matrix completion problem under the semi-random setup, given $\epsilon > 0$, there exist $p = O\left(\frac{\log d}{d\epsilon^2}\right)$ such that if each entry of $\mathbf{M}^\star$ is observed with probability at least p, then with high probability we can compute a weight matrix $W \in \mathbb{R}^{d_1\times d_2}$ in time $\widetilde{O}(m/\epsilon^{O(1)})$ that achieves ϵ-spectral-similarity, such that W is supported on Ω and $\|W\|_1 \leq d_1, \|W\|_\infty \leq d_2, \|W - J\|_2 = O(\epsilon\sqrt{d_1 d_2})$ where J is the all-ones matrix.*

Lemma 4 (Closeness of Laplacian and adjacency matrix, Lemma 3.6 (Cheng and Ge, 2018)). *Let $L = D - A$ and $\widetilde{L} = \widetilde{D} - \widetilde{A}$ be two graph Laplacians, where D, $\widetilde{D}$ are the degree matrices and A, $\widetilde{A}$ are the adjacency matrices of the graphs. If $(1 - \epsilon)L \preccurlyeq \widetilde{L} \preccurlyeq L$, then we have $(1 - \epsilon)D_{ii} \preccurlyeq \widetilde{D}_{ii} \preccurlyeq D_{ii}$ and $\left\|D^{-1/2}(\widetilde{A} - A)D^{-1/2}\right\|_2 \leq 3\epsilon$.*

Lemma 5 (Preserving the norm via spectral properties, Lemma 4.3 (Cheng and Ge, 2018)). *For any matrices $X \in \mathbb{R}^{d_1\times r}$, $Y \in \mathbb{R}^{d_2\times r}$ and $W \in \mathbb{R}^{d_1\times d_2}$, we have*

$$\left|\left\|XY^\top\right\|_W^2 - \left\|XY^\top\right\|_{\mathrm{F}}^2\right| \leq \|W - J\|_2 \cdot \|X\|_{\mathrm{F}} \cdot \|Y\|_{\mathrm{F}} \cdot \|X\|_{2\to\infty} \cdot \|Y\|_{2\to\infty},$$

where J is the all-ones matrix.

B Properties of the Preprocessing Weight Matrix

This section bounds the deviation of the weighted Frobenius norm and the maximum entry of the weight matrix $\mathbf{W}$ generated by the preprocessing procedure. These bounds are required to determine the parameters for the weighted RSC, RSS, and deviation conditions in Section 3. The proofs rely on prior results from Cheng and Ge (2018), which are reproduced in Appendix A.

Lemma 6 bounds the difference between the weighted and unweighted Frobenius norms for matrices in the constraint set $\mathcal{C}$.

Lemma 6. *Suppose the preprocessing step produces weight matrix W that achieves ϵ-spectral-similarity. For all M_1, M_2 from the constraint set $\mathcal{C}$, either $\left|\|M_1 - M_2\|_W^2 - \|M_1 - M_2\|_{\mathrm{F}}^2\right| \le c \cdot \|M_1 - M_2\|_{\mathrm{F}}^2$ for some small constant c, or $\|M_1 - M_2\|_{\mathrm{F}}^2 \le O(\beta^2 r^2 \sigma_1^2 \epsilon)$.*

Proof. Based on the definition for constraint set $\mathcal{C}$, we can write $M_1 = U_1 V_1^\top$ and $M_2 = U_2 V_2^\top$ for some $U_1, U_2 \in \mathcal{C}_1$ and $V_1, V_2 \in \mathcal{C}_2$, where $\mathcal{C}_1 = \{U \in \mathbb{R}^{d_1 \times r} \mid \|U\|_{2\to\infty} \le \alpha_1\}$, and $\mathcal{C}_2 = \{V \in \mathbb{R}^{d_2 \times r} \mid \|V\|_{2\to\infty} \le \alpha_2\}$. let $X = [U_1, U_2] \in \mathbb{R}^{d_1 \times 2r}$ and $Y = [V_1, -V_2] \in \mathbb{R}^{d_2 \times 2r}$, so that $M_1 - M_2 = XY^\top$.

Following Lemma 3, suppose $\|W - J\|_2 \le C\epsilon\sqrt{d_1 d_2}$ for some constant C. We have the following inequality:

$$\begin{aligned}
\left|\left\|XY^\top\right\|_W^2 - \left\|XY^\top\right\|_{\mathrm{F}}^2\right| &\le \|W - J\|_2 \cdot \|X\|_{\mathrm{F}} \cdot \|Y\|_{\mathrm{F}} \cdot \|X\|_{2\to\infty} \cdot \|Y\|_{2\to\infty} \\
&\le \|W - J\|_2 \cdot \sqrt{d_1}\|X\|_{2\to\infty} \cdot \sqrt{d_2}\|Y\|_{2\to\infty} \cdot \|X\|_{2\to\infty} \cdot \|Y\|_{2\to\infty} \\
&= \sqrt{d_1 d_2} \cdot \|W - J\|_2 \cdot \|X\|_{2\to\infty}^2 \cdot \|Y\|_{2\to\infty}^2 \\
&\le C\epsilon d_1 d_2 \alpha_1^2 \alpha_2^2 \\
&\le C\epsilon d^2 \alpha^2.
\end{aligned}$$

The first inequality is an application of Lemma 5, the second inequality comes the shape of X and Y, the third inequality is due to the bound on $\|W - J\|_2$, and the last inequality comes from $d = \max(d_1, d_2)$ and the definition of $\mathcal{C}_1$ and $\mathcal{C}_2$.

We consider two separate cases:

Case 1: if $\left\|XY^\top\right\|_{\mathrm{F}}^2 \ge \frac{C}{c}\epsilon d^2 \alpha^2$, then:

$$\left|\left\|XY^\top\right\|_W^2 - \left\|XY^\top\right\|_{\mathrm{F}}^2\right| \le C\epsilon d^2 \alpha^2 \le c \cdot \left\|XY^\top\right\|_{\mathrm{F}}^2.$$

Case 2: otherwise $\left\|XY^\top\right\|_{\mathrm{F}}^2 \le \frac{C}{c}\epsilon d^2 \alpha^2$, then:

$$\left\|XY^\top\right\|_{\mathrm{F}}^2 \le \frac{C}{c}\epsilon d^2 \frac{\beta^2 r^2 \sigma_1^2}{d_1 d_2} = O\left(\beta^2 r^2 \sigma_1^2 \epsilon\right). \qquad \square$$

Lemma 7. *Suppose the preprocessing step produces weight matrix W that achieves ϵ-spectral-similarity. Then we have $\|W\|_{\max} = O(\epsilon\sqrt{d_1 d_2})$.*

Proof. Let G denote a $d_1 \times d_2$ complete bipartite graph (corresponding to the full matrix), and H denote the semi-random graph generated by including each edge of G with probability at least p, i.e., edges in H correspond to the the observed indices in Ω. Apply weights W on H to get weighted graph $\widetilde{H}$. Let $L = D - A$ be the Laplacian of G, where D is the degree matrix and A is the adjacency matrix. Similarly write $\widetilde{L} = \widetilde{D} - \widetilde{A}$ for $\widetilde{H}$. Let $J \in \mathbb{R}^{d_1 \times d_2}$ be the all ones matrix, $I_1 \in \mathbb{R}^{d_1 \times d_1}, I_2 \in \mathbb{R}^{d_2 \times d_2}$ be identity matrices of corresponding dimensions, we have:

$$L = \begin{pmatrix} d_2 I_1 & -J \\ -J^\top & d_1 I_2 \end{pmatrix} \qquad \widetilde{L} = \begin{pmatrix} \widetilde{D_1} & -W \\ -W^\top & \widetilde{D_2} \end{pmatrix} \qquad \text{assuming } \widetilde{D} = \begin{pmatrix} \widetilde{D_1} & 0 \\ 0 & \widetilde{D_2} \end{pmatrix}$$

Following Theorem 4, $(1-\epsilon)L \preccurlyeq \widetilde{L} \preccurlyeq L$. For all $x \in \mathbb{R}^{(d_1+d_2)}$:

$$(1-\epsilon)x^\top Lx \le x^\top \widetilde{L}x \tag{7}$$

Consider $x = [\sqrt{d_1}e_i; \sqrt{d_2}e_j]$ where $e_i \in \mathbb{R}^{d_1}, e_j \in \mathbb{R}^{d_2}$ are standard basis vectors. We have the following quadratic forms:

$$x^\top Lx = d_1d_2I_{1,jj} + d_2d_1I_{2,kk} - 2\sqrt{d_1d_2}J_{ij}$$
$$x^\top \widetilde{L}x = d_1\widetilde{D_1}_{jj} + d_2\widetilde{D_2}_{kk} - 2\sqrt{d_1d_2}W_{ij}$$

By Lemma 4, $\widetilde{D_1}_{jj} \le d_2$ and $\widetilde{D_2}_{kk} \le d_1$, so the quadratic forms simplify to:

$$x^\top Lx = 2d_1d_2 - 2\sqrt{d_1d_2} \tag{8}$$
$$x^\top \widetilde{L}x \le 2d_1d_2 - 2\sqrt{d_1d_2}W_{ij} \tag{9}$$

Combining (7) (8) and (9), we have:

$$\begin{aligned} 2d_1d_2 - 2\epsilon d_1d_2 - 2\sqrt{d_1d_2} + 2\epsilon\sqrt{d_1d_2} &\le 2d_1d_2 - 2\sqrt{d_1d_2}W_{ij} \\ \implies W_{ij} &\le \epsilon\sqrt{d_1d_2} + 1 - \epsilon \\ &= O(\epsilon\sqrt{d_1d_2}) \text{ since } \epsilon d = \omega(1) \end{aligned}$$

□

C Omitted Proofs From Section 4

C.1 Proof of Corollary 1

Corollary 1 (Semi-Random Noisy Matrix Completion)**.** *Suppose the ground truth matrix $\mathbf{M}^\star \in \mathbb{R}^{d_1\times d_2}$ has rank r and is β-incoherent. Consider the semi-random observation model with sampling probability at least $p = \Omega\left(\frac{r^4\log^3 d}{d}\right)$. Assume the observation noise $\mathbf{N}_{ij}$ is independent and identically distributed with zero mean and variance $\frac{\nu^2}{d_1d_2}$, and satisfies the sub-exponential tail condition. Let $\mathcal{F}_{\mathbf{W}}$ be the weighted objective constructed via the preprocessing algorithm. Then, with high probability, any local minimum $\mathbf{Z} = \begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix}$ of the constrained optimization problem satisfies*

$$\left\|\mathbf{U}\mathbf{V}^\top - \mathbf{M}^\star\right\|_{\mathrm{F}}^2 \le \max\left\{O\left(\beta^2\sigma_1^2\sqrt{\frac{r^4\log d}{pd}}\right), O\left(\nu^2\sqrt{\frac{r^2\log^3 d}{pd}}\right)\right\}.$$

Proof of Corollary 1. To apply the semi-random error bound in Theorem 3, it suffices to verify that the uniform loss function $\mathcal{F}$ given in (1) satisfies Condition 1 and Condition 2.

Recall the uniform loss function (1) and its gradient

$$\begin{aligned} \nabla\mathcal{F}(\mathbf{X}) &= \sum_{(i,j)\in\Omega}\frac{1}{p}(\mathbf{X}_{ij} - \mathbf{M}_{ij})e_ie_j^\top \\ &= \sum_{(i,j)\in\Omega}\frac{1}{p}b_{ij}(\mathbf{X})e_ie_j^\top. \end{aligned}$$

Evaluating at the ground truth $\mathbf{M}^\star$, we have $b_{ij}(\mathbf{M}^\star) = \mathbf{N}_{ij}$. By assumption, the noise entries $\mathbf{N}_{ij}$ are independent, zero-mean, and sub-exponential with variance $\frac{\nu^2}{d_1d_2}$, which satisfies Condition 2 with variance parameter $O(\nu^2/d^2)$.

The second-order Taylor expansion residual of $\mathcal{F}$ is:

$$\begin{aligned}
&\mathcal{F}(\mathbf{X}_1) - \mathcal{F}(\mathbf{X}_2) - \langle \nabla\mathcal{F}(\mathbf{X}_2), \mathbf{X}_1 - \mathbf{X}_2 \rangle \\
&= \frac{1}{2} \sum_{(i,j)\in\Omega} \frac{1}{p} \left[(\mathbf{X}_{1,ij} - \mathbf{M}_{ij})^2 - (\mathbf{X}_{2,ij} - \mathbf{M}_{ij})^2 \right] - \sum_{(i,j)\in\Omega} \frac{1}{p} (\mathbf{X}_{2,ij} - \mathbf{M}_{ij})(\mathbf{X}_{1,ij} - \mathbf{X}_{2,ij}) \\
&= \frac{1}{2} \sum_{(i,j)\in\Omega} \frac{1}{p} (\mathbf{X}_{1,ij} - \mathbf{X}_{2,ij})^2.
\end{aligned}$$

This matches the form required by Condition 1 with $\mathbf{K}_{ij} = 1$ for all $(i,j) \in \Omega$. Following Zhang et al. (2018b), the uniform RSC and RSS parameters for this loss are bounded by constants, specifically $\mu = \frac{42}{43}$ and $L = \frac{44}{43}$.

Since both conditions hold, we invoke Theorem 3. By Lemma 3, the preprocessing algorithm achieves ϵ-spectral similarity with $\epsilon = O\left(\sqrt{\frac{\log d}{pd}}\right)$. Substituting this ϵ into the error bound yields

$$\left\|\mathbf{U}\mathbf{V}^\top - \mathbf{M}^\star\right\|_{\mathrm{F}}^2 \le \max\left\{ O\left(\beta^2\sigma_1^2 \sqrt{\frac{r^4 \log d}{pd}}\right), O\left(\nu^2 \sqrt{\frac{r^2 \log^3 d}{pd}}\right) \right\},$$

completing the proof. □

C.2 Proof of Corollary 2

Corollary 2 (Semi-Random One-Bit Matrix Completion)**.** *Suppose the ground truth matrix $\mathbf{M}^\star \in \mathbb{R}^{d_1 \times d_2}$ has rank r and is β-incoherent. Consider the one-bit matrix completion problem under the semi-random observation model with sampling probability at least $p = \Omega\left(\frac{r^4 \log^3 d}{d}\right)$. Assume the observation CDF satisfies the steepness condition (6) with parameter s_ρ. Let $\mathcal{F}_{\mathbf{W}}$ be the weighted objective constructed via the preprocessing algorithm. Then, with high probability, any local minimum $\mathbf{Z} = \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix}$ of the constrained optimization problem satisfies*

$$\left\|\mathbf{U}\mathbf{V}^\top - \mathbf{M}^\star\right\|_{\mathrm{F}}^2 \le \max\left\{ O\left(\beta^2\sigma_1^2 \sqrt{\frac{r^4 \log d}{pd}}\right), O\left(s_\rho^2 \sqrt{\frac{r^2 \log^3 d}{pd}}\right) \right\}.$$

Proof of Corollary 2. By Corollary 4.3 in Zhang et al. (2018b), the uniform 1-bit loss function satisfies the global regularity conditions required in Theorem 2. Therefore, it suffices to verify Condition 1 and Condition 2 to apply the general semi-random error bound.

Recall the uniform negative log-likelihood loss function (2) in its standardized form:

$$\begin{aligned}
\mathcal{F}(\mathbf{X}) &= -\frac{1}{pd_1d_2} \sum_{(i,j)\in\Omega} \Big[\mathbb{1}(\mathbf{M}_{ij} = 1) \log f(\mathbf{X}_{ij}) + \mathbb{1}(\mathbf{M}_{ij} = -1) \log\left(1 - f(\mathbf{X}_{ij})\right) \Big] \\
&= -\frac{1}{d_1d_2} \sum_{(i,j)\in\Omega} \frac{1}{p} \Big[\mathbb{1}(\mathbf{M}_{ij} = 1) \log g(\mathbf{X}_{ij}/\tau) + \mathbb{1}(\mathbf{M}_{ij} = -1) \log\left(1 - g(\mathbf{X}_{ij}/\tau)\right) \Big].
\end{aligned}$$

And its gradient:

$$\nabla\mathcal{F}(\mathbf{X}) = \frac{1}{d_1d_2\tau} \sum_{(i,j)\in\Omega} \frac{1}{p} b_{ij}(\mathbf{X}) e_i e_j^\top,$$

where

$$b_{ij}(\mathbf{X}) = -\mathbb{1}(\mathbf{M}_{ij} = 1) \frac{g'(\mathbf{X}_{ij}/\tau)}{g(\mathbf{X}_{ij}/\tau)} + \mathbb{1}(\mathbf{M}_{ij} = -1) \frac{g'(\mathbf{X}_{ij}/\tau)}{1 - g(\mathbf{X}_{ij}/\tau)}.$$

Evaluating at the ground truth $\mathbf{M}^\star$, the scalar $b_{ij}(\mathbf{M}^\star)$ takes the value $-\frac{g'(\mathbf{M}^\star_{ij}/\tau)}{g(\mathbf{M}^\star_{ij}/\tau)}$ with probability $g(\mathbf{M}^\star_{ij}/\tau)$, and $\frac{g'(\mathbf{M}^\star_{ij}/\tau)}{1-g(\mathbf{M}^\star_{ij}/\tau)}$ with probability $1 - g(\mathbf{M}^\star_{ij}/\tau)$. Thus, $\mathbb{E}[b_{ij}(\mathbf{M}^\star)] = 0$. The variance is bounded by the steepness

parameter s_ρ:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[b_{ij}(\mathbf{M}^\star)\right] &= \mathbb{E}\left[b_{ij}^2(\mathbf{M}^\star)\right] \\
&= \left[\frac{g'(\mathbf{M}^\star_{ij}/\tau)}{g(\mathbf{M}^\star_{ij}/\tau)}\right]^2 g(\mathbf{M}^\star_{ij}/\tau) + \left[\frac{g'(\mathbf{M}^\star_{ij}/\tau)}{1-g(\mathbf{M}^\star_{ij}/\tau)}\right]^2 \big(1-g(\mathbf{M}^\star_{ij}/\tau)\big) \\
&= \frac{g'^2(\mathbf{M}^\star_{ij}/\tau)}{g(\mathbf{M}^\star_{ij}/\tau)\big(1-g(\mathbf{M}^\star_{ij}/\tau)\big)} \\
&= \left(\frac{|g'(\mathbf{M}^\star_{ij}/\tau)|}{g(\mathbf{M}^\star_{ij}/\tau)\big(1-g(\mathbf{M}^\star_{ij}/\tau)\big)}\right)^2 g(\mathbf{M}^\star_{ij}/\tau)\big(1-g(\mathbf{M}^\star_{ij}/\tau)\big) \\
&\le s_\rho^2.
\end{aligned}$$

The final inequality uses the definition of s_ρ and the fact that $g \in [0,1]$. Consequently, each scaled gradient component $\frac{b_{ij}(\mathbf{M}^\star)}{d_1 d_2 \tau}$ is a zero-mean sub-exponential random variable with variance bounded by $\frac{s_\rho^2}{d^4\tau^2} = \frac{s_\rho^2}{d^2\nu^2}$ (assuming $d_1 d_2 \approx d^2$), which satisfies Condition 2.

To verify Condition 1, we apply the Mean Value Theorem. For any $\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{C}$, there exists a matrix $M = t\mathbf{X}_1 + (1-t)\mathbf{X}_2$ for some $t \in [0,1]$ such that the second-order Taylor residual is

$$\mathcal{F}(\mathbf{X}_1) - \mathcal{F}(\mathbf{X}_2) - \langle \nabla\mathcal{F}(\mathbf{X}_2), \mathbf{X}_1 - \mathbf{X}_2\rangle = \frac{1}{2}\operatorname{vec}(\mathbf{X}_1 - \mathbf{X}_2)^\top \nabla^2\mathcal{F}(M)\operatorname{vec}(\mathbf{X}_1 - \mathbf{X}_2).$$

Using $\frac{1}{d_1 d_2 \tau^2} = \frac{1}{\nu^2}$, the Hessian of the loss function is:

$$\nabla^2\mathcal{F}(\mathbf{X}) = \frac{1}{\nu^2}\sum_{(i,j)\in\Omega}\frac{1}{p}B_{ij}(\mathbf{X})\operatorname{vec}(e_i e_j^\top)\operatorname{vec}(e_i e_j^\top)^\top,$$

where

$$B_{ij}(\mathbf{X}) = \mathbb{1}(\mathbf{M}_{ij}=1)\left(\frac{g'^2(\mathbf{X}_{ij}/\tau)}{g^2(\mathbf{X}_{ij}/\tau)} - \frac{g''(\mathbf{X}_{ij}/\tau)}{g(\mathbf{X}_{ij}/\tau)}\right) + \mathbb{1}(\mathbf{M}_{ij}=-1)\left(\frac{g''(\mathbf{X}_{ij}/\tau)}{1-g(\mathbf{X}_{ij}/\tau)} + \frac{g'^2(\mathbf{X}_{ij}/\tau)}{\big(1-g(\mathbf{X}_{ij}/\tau)\big)^2}\right).$$

Following Ni and Gu (2016), we define constants bounding the curvature over the domain $|x| \le \rho$:

$$\begin{aligned}
\mu_\rho &= \min\left\{\inf_{|x|\le\rho}\left(\frac{g'^2(x)}{g^2(x)} - \frac{g''(x)}{g(x)}\right), \inf_{|x|\le\rho}\left(\frac{g'^2(x)}{(1-g(x))^2} + \frac{g''(x)}{1-g(x)}\right)\right\}, \\
L_\rho &= \max\left\{\sup_{|x|\le\rho}\left(\frac{g'^2(x)}{g^2(x)} - \frac{g''(x)}{g(x)}\right), \sup_{|x|\le\rho}\left(\frac{g'^2(x)}{(1-g(x))^2} + \frac{g''(x)}{1-g(x)}\right)\right\}.
\end{aligned}$$

Because $\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{C}$, their entries are bounded by α. Thus, $|M_{ij}/\tau| \le \alpha/\tau = \rho$. This implies $\mu_\rho \le B_{ij}(M) \le L_\rho$ for all $(i,j) \in \Omega$. The residual simplifies to:

$$\frac{1}{2\nu^2}\sum_{(i,j)\in\Omega}\frac{1}{p}B_{ij}(M)\langle e_i e_j^\top, \mathbf{X}_1 - \mathbf{X}_2\rangle^2 = \frac{1}{2}\sum_{(i,j)\in\Omega}\frac{1}{p}\mathbf{K}_{ij}(\mathbf{X}_{1,ij} - \mathbf{X}_{2,ij})^2,$$

where $\mathbf{K}_{ij} = B_{ij}(M)/\nu^2$, satisfying $\frac{\mu_\rho}{\nu^2} \le \mathbf{K}_{ij} \le \frac{L_\rho}{\nu^2}$. By Zhang et al. (2018b), the uniform parameters are $\mu = \frac{42}{43}\frac{\mu_\rho}{\nu^2}$ and $L = \frac{44}{43}\frac{L_\rho}{\nu^2}$. Therefore, Condition 1 holds with $\mu \le \mathbf{K}_{ij} \le L$.

Finally, we apply Theorem 3. The condition number parameter $\Gamma = O(1/\mu)$ scales as $O(\nu^2/\mu_\rho) = O(\nu^2)$. The deviation bound squared δ^2 scales with the variance $O(s_\rho^2/\nu^2)$, effectively cancelling the ν^2 factor. Substituting the preprocessing bound $\epsilon = O\left(\sqrt{\frac{\log d}{pd}}\right)$ yields

$$\left\|\mathbf{U}\mathbf{V}^\top - \mathbf{M}^\star\right\|_{\mathrm{F}}^2 \le \max\left\{O(\beta^2 r^2 \sigma_1^2 \epsilon), O(r s_\rho^2 \epsilon \log d)\right\},$$

completing the proof. □

D Matrix Concentration Bounds

For completeness we include a lemma in contrast to Lemma 2, on the operator norm of the gradient (deviation bound) in the uniform case.

Lemma 8 (Lemma B.2 (Zhang et al., 2018b; Negahban and Wainwright, 2012)). *Consider noisy matrix completion with the uniform observation model. Suppose the noise entry $\mathbf{N}_{ij}$ follows i.i.d zero mean distribution with variance $O(\nu^2/d^2)$. Then with probability at least $1 - c_1/d$, we have*

$$\left\| \frac{1}{p} \sum_{(i,j)\in\Omega} \mathbf{N}_{ij} e_i e_j^\top \right\|_2^2 \leq c_2 O(\nu^2/d^2) \frac{d \log d}{p},$$

where c_1, c_2 are universal constants and $p = \frac{|\Omega|}{d_1 d_2}$.

The following theorem provides a tail bound on the operator norm of a sum of random matrices.

Fact 1 (Matrix Berstein, Theorem 1.6 (Tropp, 2012)). *Consider a finite sequence $\{\mathbf{X}_k\} \subset \mathbb{R}^{d_1 \times d_2}$ of independent random matrices. Assume that each random matrix satisfies $\mathbb{E}[\mathbf{X}_k] = \mathbf{0}$ and $\|\mathbf{X}_k\|_2 \leq R$ almost surely, define the variance parameter*

$$\sigma^2 =: \max\left\{ \left\| \sum_k \mathbb{E}\left[\mathbf{X}_k \mathbf{X}_k^\top\right] \right\|_2, \left\| \sum_k \mathbb{E}\left[\mathbf{X}_k^\top \mathbf{X}_k\right] \right\|_2 \right\}.$$

Then for all $t \geq 0$,

$$\mathbb{P}\left[\left\| \sum_k \mathbf{X}_k \right\|_2 \geq t \right] \leq (d_1 + d_2) \cdot \exp\left(\frac{-t^2}{2\sigma^2 + 2Rt/3} \right).$$

The boundedness assumption can be relaxed to a sub-exponential tail condition via a standard reduction.

Fact 2 (A sub-exponential extension of Matrix Bernstein (cf. Tropp (2012))). *Consider a finite sequence $\{\mathbf{X}_k\} \subset \mathbb{R}^{d_1 \times d_2}$ of independent, zero-mean random matrices. Suppose there exists a constant R such that the sub-exponential norm satisfies $\|\mathbf{X}_k\|_{\psi_1} \leq R$ for all k. Define the variance parameter*

$$\sigma^2 =: \max\left\{ \left\| \sum_k \mathbb{E}\left[\mathbf{X}_k \mathbf{X}_k^\top\right] \right\|_2, \left\| \sum_k \mathbb{E}\left[\mathbf{X}_k^\top \mathbf{X}_k\right] \right\|_2 \right\}.$$

Then there exist absolute constants $c_1, c_2 > 0$ such that for all $t \geq 0$,

$$\mathbb{P}\left[\left\| \sum_k \mathbf{X}_k \right\|_2 \geq t \right] \leq (d_1 + d_2) \exp\left(\frac{-t^2}{c_1\sigma^2 + c_2 Rt} \right).$$