OPT2021: 13th Annual Workshop on Optimization for Machine Learning

Sign-RIP: A Robust Restricted Isometry Property for Low-rank Matrix Recovery

Jianhao Ma JIANHAO@UMICH.EDU
Salar Fattahi FATTAHI@UMICH.EDU
University of Michigan, USA

Abstract

Restricted isometry property (RIP), essentially stating that the linear measurements are approximately norm-preserving, plays a crucial role in studying low-rank matrix recovery problem. However, RIP fails in the robust setting, when a subset of the measurements are grossly corrupted with noise. In this work, we propose a robust restricted isometry property, called *Sign-RIP*, and show its broad applications in robust low-rank matrix recovery. In particular, we show that Sign-RIP can guarantee the uniform convergence of the subdifferentials of the robust matrix recovery with nonsmooth loss function, even at the presence of arbitrarily dense and arbitrarily large outliers. Based on Sign-RIP, we characterize the location of the critical points in the robust rank-1 matrix recovery, and prove that they are either close to the true solution, or have small norm. Moreover, in the over-parameterized regime, where the rank of the true solution is over-estimated, we show that subgradient method converges to the true solution at a (nearly) dimension-free rate. We show that the new notion of sign-RIP enjoys almost the same sample complexity as its classical counterparts, but provides significantly better robustness against noise.

1. Introduction

Inspired by the surprising success of simple local-search algorithms in nonconvex optimization arising in modern machine learning tasks, a recent body of work focuses on studying the local and global optimization landscape of these problems. A prototypical class of such problems is low-rank matrix recovery, where the goal is to recover a low-rank matrix from a limited number of linear and noisy measurements. Low-rank matrix recovery is the cornerstone for many modern machine learning problems, including motion detection [2], face recognition [12], recommender systems [13], and system identification [4, 11].

Despite the inherent difficulty of low-rank matrix recovery in its worst case—a fact noted as early as 1995 [14]—it is known that convex relaxation methods can correctly recover the low-rank matrix under the so-called *restricted isometry property* (RIP) [3, 15, 19], but suffer from high computational cost. Roughly speaking, RIP entails that the measurement operator is approximately norm-preserving over the set of low-rank matrices. One of the breakthrough results in this line of research was presented in 2016 NeurIPS papers [1, 7], showing that, for smooth low-rank matrix recovery, simple saddle-escaping algorithms, such as perturbed gradient descent (GD) [6, 8], provably converge to the true low-rank solution. The main intuition behind this result is a simple, yet striking one: *under the same RIP condition, the nonconvex formulation of smooth low-rank matrix recovery problem is devoid of spurious local minima.* This result lead to a flurry of follow-up papers characterizing the landscape of other variants of low-rank matrix recovery [10, 20–22].

Despite the significance of different notions of RIP within the realm of low-rank matrix recovery, they face major breakdowns in robust settings, where a subset of the measurements are grossly corrupted with large noise values. The main reason behind the failure of the existing RIP techniques is that they only apply to nearly clean measurements, and hence, are oblivious to the nature of the noise.

The main goal of this paper is to precisely pinpoint and remedy this challenge. In particular, we study a well-known class of matrix recovery problems with nonsmooth and nonconvex ℓ_1 formulation, called *robust matrix recovery*. We introduce an alternative notion of RIP, called *Sign-RIP*, that can capture and take into account the nature of the noise in robust matrix recovery. Based on Sign-RIP, we take the first step towards demystifying the robustness of the ℓ_1 formulation of the problem against large-and-sparse noise values. Our main contributions are summarized as follows:

- *(Uniform convergence of subdifferentials)* We use Sign-RIP to study the landscape of the robust matrix recovery against large noise values. In particular, we show that, under Sign-RIP, the subdifferentials of the nonsmooth matrix recovery are well-behaved, and they converge uniformly to the gradients of an "ideal", noiseless formulation of the problem, even if the measurements are subject to large noise values. Moreover, we show that Sign-RIP holds, even if an arbitrarily large fraction of the measurements are corrupted with arbitrarily large noise values, provided that the number of measurements scales polynomially with the corruption probability, but only linearly with the true dimension of the problem.

- *(Characterization of the critical points)* We show that Sign-RIP can be used to characterize the locations of the critical points for the robust rank-1 matrix recovery. In particular, we show that, under Sign-RIP, all critical points lie close to the true rank-1 solution, or have small norm.

- *(Implicit regularization with over-parameterization)* Based on Sign-RIP, we show that a subgradient method with decaying step sizes provably converges to the true rank-1 solution in the over-parameterized regime, where the true rank is unknown and over-estimated.

To streamline the presentation, all the proofs are deferred to the appendix.

2. Background and Prior Work

In robust matrix recovery problem, the goal is to recover a rank-r^* positive semidefnite matrix $X^* \in \mathbb{R}^{d\times d}$, from a limited number of linear measurements of the form $\mathbf{y} = \mathcal{A}(X^*) + \mathbf{s}$, where $\mathbf{y} = [y_1, y_2, \dots, y_m]^\top$ is the vector of measurements, and $\mathbf{s}$ is a noise vector. The linear operator $\mathcal{A}$ is defined as $\mathcal{A}(X^*) = [\langle A_1, X^*\rangle, \langle A_2, X^*\rangle, \dots, \langle A_m, X^*\rangle]^\top$, where $\{A_i\}_{i=1}^m$ are measurement matrices. One popular approach for recovering the true low-rank matrix is to consider the following empirical risk minimization (ERM) problem

$$\min_{U\in\mathbb{R}^{d\times r'}} f_{\ell_q}(U) = \frac{1}{2m}\left\|\mathbf{y} - \mathcal{A}(UU^\top)\right\|_{\ell_q}^q, \tag{1}$$

where r' is an upper bound for the rank of the true solution, and $UU^\top$ is used in lieu of X^* to ensure the positive semidefiniteness of the solution.

ℓ_2-RIP: Evidently, the above optimization problem is over-parameterized if $r' > r^*$, since the unknown variable is not restricted to the set of low-rank matrices, and consequently, its globally optimal solution need *not* be low-rank. Nonetheless, it is recently shown that, for the choice of $q = 2$, simple gradient descent (GD) algorithm provably converges to the true rank-r^* solution, even if $r' \gg r^*$ (e.g. $r' = d$) [10, 22]. The key idea behind the convergence proof of GD is the closeness of its gradient to that of the "ideal", noiseless population loss function $\bar{f}_{\ell_2}(U) = \|UU^\top - X^*\|_F^2$. More concretely, the gradient of $f_{\ell_2}(U)$ can be written as $\nabla f_{\ell_2}(U) = Q(UU^\top - X^*)U$, where $Q(M) = \frac{1}{m}\sum_{i=1}^m (\langle A_i, M\rangle + s_i)\frac{A_i + A_i^\top}{2}$. One sufficient condition for $\nabla f_{\ell_2}(U) \approx \nabla \bar{f}_{\ell_2}(U)$ is to ensure that $Q(M)$ remains uniformly close to M for every rank-$(r^* + r')$ matrix X[1]. In the noiseless setting, this condition can be guaranteed by the so-called ℓ_2-RIP.

1. The paper [10] requires the similarity of $Q(X)$ and X for lower rank matrices (rank-r^* as opposed to rank-$(r^* + r')$), but their result only holds for $r' = d$.

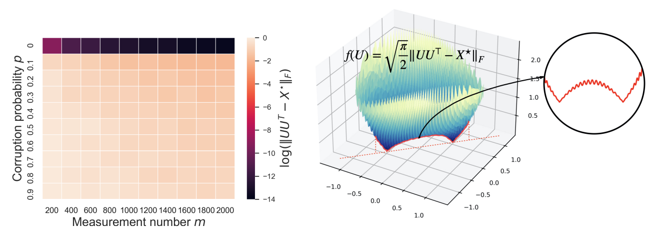
Definition 1 **(ℓ_2-RIP [10, 22])** *The linear operator $\mathcal{A}(\cdot)$ satisfies ℓ_2-RIP with parameters (r, δ) if, for every rank-r matrix M, we have $(1-\delta)\|M\|_F^2 \leq \frac{1}{m}\|\mathcal{A}(M)\|^2 \leq (1+\delta)\|M\|_F^2$.*

Roughly speaking, ℓ_2-RIP entails that the linear operator $\mathcal{A}(\cdot)$ is nearly norm-preserving for every rank-r matrix. It is well-known that with Gaussian measurements, ℓ_2-RIP is satisfied with parameters (r, δ), provided that $m \gtrsim dr$ [15]. However, our next proposition shows that ℓ_2-RIP is not enough to guarantee $Q(X) \approx X$ when the measurements are subject to noise.

Proposition 2 *Suppose that $r' = d$ and the measurement matrices $\{A_i\}_{i=1}^m$ defining the linear operator $\mathcal{A}(\cdot)$ have i.i.d. standard Gaussian entries. Moreover, suppose that the noise vector $\mathbf{s}$ satisfies $s_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ with probability p, and $s_i = 0$ with probability $1-p$, for every $i = 1, \dots, m$. Then we have*

$$\mathbb{P}\left(\sup_{X \in \mathbb{S}} \|Q(X) - X\|_F \gtrsim \sqrt{\frac{(1+p\sigma^2)d^2}{m}}\right) \geq \frac{1}{2}.$$

The above proposition shows that, in order to guarantee $Q(X) \approx X$, the number of measurements should be at least $m \gtrsim (1+p\sigma^2)d^2$, and hence, grow with the variance of the noise. However, for any fixed δ, ℓ_2-RIP is guaranteed to be satisfied with $m \lesssim d^2$, which is independent of the noise variance. This highlights a fundamental pitfall of ℓ_2-RIP in the existence of outliers: the matrices $Q(X)$ and X may be far apart, *even if* the linear mapping $\mathcal{A}(\cdot)$ satisfies ℓ_2-RIP. Figure 1a illustrates that the discrepancy between $Q(X)$ and X in the noisy setting leads to the ultimate failure of the gradient descent algorithm.



(a) (b)

Figure 1: (a) The accuracy of the obtained solutions by solving (1) via gradient descent. Each measurement is corrupted with a large noise with probability p. (b) The linear operator $\mathcal{A}(\cdot)$ satisfies the ℓ_1/ℓ_2-RIP, but the subdifferentials of the ℓ_1-loss show sporadic behavior. Here, $f(U)$ is the population loss.

ℓ_1/ℓ_2-RIP: With the goal of robustifying the solution against outlier noise, recent work has studied the landscape of the nonsmooth optimization (1) with $q = 1$ under ℓ_1/ℓ_2-RIP. Roughly speaking, ℓ_1/ℓ_2-RIP imposes a similar condition to ℓ_2-RIP, but on the ℓ_1-loss function. In particular, it entails that $\frac{1}{m}\|\mathcal{A}(X)\|_1$ remains close to $\sqrt{2/\pi}\|X\|_F$, for every rank-r matrix X. Under ℓ_1/ℓ_2-RIP, it is recently shown that subgradient method converges to the ground truth with $q = 1$, provided that the true rank of the solution is known and the initial point is sufficiently close to the ground truth [9, 17]. However, ℓ_1/ℓ_2-RIP is also oblivious to the nature of the noise, and as a result, cannot guarantee the global convergence of the corresponding subdifferentials of $f_{\ell_1}(U)$. Figure 1b shows an instance of (1) with $q = 1$, where the linear operator $\mathcal{A}(\cdot)$ satisfies ℓ_1/ℓ_2-RIP, and yet the subdifferentials of the loss function suffer from sporadic behavior due to noise, giving rise to numerous undesirable local minima.

The aforementioned challenges highlight the fundamental limitations of the existing notions of RIP in the context of robust matrix recovery with large noise. This calls for a new approach for analyzing the landscape of robust matrix recovery; a goal that is at crux of this paper.

3. Our Approach: Sign-RIP for ℓ_1-loss Function

Our goal is to study the following ERM problem with ℓ_1-loss function:

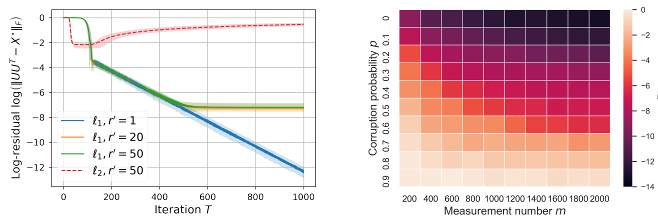
$$\min_{U \in \mathbb{R}^{d \times r'}} f_{\ell_1}(U) = \frac{1}{2m}\left\|\mathbf{y} - \mathcal{A}(UU^\top)\right\|_1, \tag{2}$$

The simplest algorithm for solving (2) is subgradient method (SubGM). At every iteration, SubGM selects an arbitrary direction D_t from the subdifferential of the ℓ_1-loss at the current solution, and then updates the solution by moving towards $-D_t$ with a step size η_t (see Algorithm (1) in Appendix). Figure 2a illustrates that SubGM with diminishing step sizes can successfully recover the true rank-1 solution in both exact and over-parameterized regimes when applied to the ℓ_1-loss (2), even if 10% of the measurements are grossly contaminated with noise. On the other hand, GD on the smooth loss function quickly overfits to the noise within a few iterations. Figure 2b shows the robustness of SubGM against increasing fraction of noisy measurements. It can be seen that SubGM recovers the true solution, even if more than half of the measurements are corrupted with noise, pinpointing its superiority over GD.

To study the convergence of SubGM, it is essential to analyze the behavior of the subdifferential

$$\partial f_{\ell_1}(U_t) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Sign}(\langle A_i, U_tU_t^\top - X^\star\rangle - s_i)\frac{A_i + A_i^\top}{2}U_t.$$

The key idea behind our proposed technique is to study the theoretical superiority of the robust matrix recovery with ℓ_1-loss, by characterizing its distance to an ideal, noiseless problem $\bar{f}_{\ell_2}(U) = \|UU^\top - X^*\|_F^2$. In particular, we show that robust matrix recovery with ℓ_1-loss benefits from strong robustness properties, provided that its subdifferentials resemble the gradients of $\bar{f}_{\ell_2}(U)$. In particular, upon defining $\mathcal{Q}(X) = \frac{1}{m}\sum_{i=1}^{m}\mathrm{Sign}(\langle A_i, X\rangle - s_i)\frac{A_i+A_i^\top}{2}$, our goal is to provide conditions under which, for *any* $Q \in \mathcal{Q}(UU^\top - X^*)$, the subgradient QU of the ℓ_1-loss function points approximately towards the direction $(UU^\top - X^*)U$, which is the gradient of $\bar{f}_{\ell_2}(U)$. This is formalized through the notion of *Sign-RIP*. For a matrix X, its r-restricted Frobenius norm is defined as $\|X\|_{F,r} := \sup_{\|Y\|_F=1,\mathrm{rank}(Y)\le r}\langle X, Y\rangle$.



(a) (b)

Figure 2: (a) SubGD with geometrically decaying step sizes recovers the true rank-1 matrix in both exact ($r' = r = 1$) and over-parameterized ($r' > r$) regimes. GD with small step sizes overfits to the noise. (b) SubGD with geometrically decaying step sizes recovers the true rank-1 matrix in the over-parameterized regime ($r' = d = 20$), for different number of measurements m and corruption probabilities p.

Definition 3 (Sign-RIP) *The measurements are said to satisfy Sign-RIP with parameters (r, δ) and a scaling function φ if, for every rank-r matrix X and every $Q \in \mathcal{Q}(X)$, we have $\left\|Q - \varphi(X)\frac{X}{\|X\|_F}\right\|_{F,r} \le \varphi(X)\delta$.*

At the first glance, one may speculate that Sign-RIP is extremely restrictive: roughly speaking, it requires the uniform concentration of random set-valued function $\mathcal{Q}(X)$, for *every* rank-r matrix X. However, we show that, Sign-RIP enjoys the same linear sample complexity as ℓ_2- [15] and ℓ_1/ℓ_2-RIP [9] for standard Gaussian measurement matrices, and hence, is not statistically more restrictive than its classical counterparts. To this goal, first we introduce our considered noise model:

Assumption 1 (Noise model) *Given a corruption probability p, each entry of the noise vector* s *is independently drawn from a mean-zero distribution $\mathbb{P}$ with probability p and zero otherwise.*

Notice that our proposed noise model does not impose any assumption on the magnitude of the nonzero elements of s, or the specific form of their distribution, which makes it particularly suitable for modeling outliers with arbitrary magnitudes. Next, we characterize the sample complexity of Sign-RIP for Gaussian measurement matrices.

Theorem 4 *Assume that the measurement matrices $\{A_i\}_{i=1}^m$ defining the linear operator $\mathcal{A}(\cdot)$ are i.i.d. standard Gaussian entries, and that the noise vector* $\mathbf{s}$ *satisfies Assumption 1. Then, Sign-RIP holds with parameters (r, δ) and a scaling function $\varphi(X) = \sqrt{\frac{2}{\pi}}\left(1 - p + p\mathbb{E}\left[e^{-s_i^2/(2\|X\|_F)}\right]\right)$ with probability of at least $1 - Ce^{-cm\delta^4}$, provided that $m \gtrsim \frac{dr\left(\log\left(\frac{1}{(1-p)\delta}\right)\vee 1\right)}{\delta^4(1-p)^4}$.*

Based on Sign-RIP, we next study the landscape of the robust rank-1 matrix recovery with outlier noise.

4. Characterization of Critical Points in Robust Rank-1 Matrix Recovery

In this section, we characterize the critical points of the robust rank-1 matrix recovery. In particular, we show that, under Sign-RIP, all critical points lie within a small neighborhood of the ground truth or the origin.

Suppose that $r' = r^\star = 1$, and $X^* = u^* u^{*\top}$ for $u^* \in \mathbb{R}^{d\times 1}$. For simplicity and without loss of generality, we assume that $\|u^*\| = 1$. Recall that $\bar{U}$ is a *critical point* of $f_{\ell_1}(U)$ if it satisfies $0 \in \partial f_{\ell_1}(\bar{U})$. Moreover, $\bar{U}$ is a *local minimum* of $f_{\ell_1}(U)$ if $f_{\ell_1}(\bar{U}) \le f_{\ell_1}(U)$ for every $U \in \mathbb{B}(\bar{U}, \epsilon)$, for some $\epsilon > 0$. All local minima of $f_{\ell_1}(U)$ are also critical points [5]. Our next theorem characterizes the critical points of $f_{\ell_1}(U)$, the proof of which is in Appendix.

Theorem 5 *Assume that the measurements satisfy the Sign-RIP condition with parameters $(2, \delta)$ and a strictly positive and uniformly bounded scaling function $\varphi(X)$. Moreover, suppose that U with $\|U\| \le R$ for some $R \ge 1$ is a critical point of (2) with $r' = r^* = 1$. Then, we have $\|UU^\top - u^*u^{*\top}\|_F \lesssim \delta$ or $\|U\|^2 \lesssim \delta$, provided that $\delta \lesssim 1/R^3$.*

The above theorem shows that, under Sign-RIP with small δ, the critical points of $f_{\ell_1}(U)$ are either close to the ground truth, or have very small norm. Combined with Theorem 4, this leads to the following corollary.

Corollary 6 *Assume that $\{A_i\}_{i=1}^m$ are standard Gaussian matrices, and the noise vector* $\mathbf{s}$ *satisfies Assumption 1. Moreover, suppose that U with $\|U\| \le R$ for some $R \ge 1$ is a critical point of (2) with $r' = r^* = 1$. Then, we have $\|UU^\top - u^*u^{*\top}\|_F \lesssim \delta$ or $\|U\|^2 \lesssim \delta$ with an overwhelming probability, provided that $m \gtrsim \frac{dr' \log\left(\frac{1}{(1-p)\delta}\right)}{\delta^4(1-p)^4}$ and $\delta \lesssim 1/R^3$. Additionally, we have $UU^\top = u^*u^{*\top}$ or $\|U\| \lesssim \delta$, if $p \le \frac{1}{2} - \frac{\delta}{\sqrt{2/\pi} - \delta}$.*

Earlier works on robust matrix recovery with ℓ_1-loss can only characterize the critical points of $f_{\ell_1}(U)$ locally within a very small neighborhood of the global minima [9, 17]. Corollary 6 extends this result in two ways for $r^* = 1$: first, it provides a *global* characterization of the critical points. In particular, it shows that, with sufficiently large number of measurements, all critical points with a bounded norm concentrate around the ground truth or the origin, provided that $p < 1$. Moreover, it shows that, if additionally we have $p < 1/2$, all critical points that are not close to the origin must coincide the ground truth.

5. Over-parameterized Robust Rank-1 Matrix Recovery

In this section, we study the over-parameterized robust matrix recovery problem. In particular, we show that, *under Sign-RIP condition, SubGM converges to the true rank-1 solution in the over-parameterized regime, without any explicit regularization or rank constraint.*

Intuition behind our analysis. Suppose that Sign-RIP holds with sufficiently small δ. Then, we have $D_t \approx \varphi_t \frac{(U_tU_t - X^*)U_t}{\|U_tU_t - X^*\|_F}$ for every $D_t \in \partial f_{\ell_1}(U_t)$, where for simplicity, we define $\varphi_t = \varphi(U_tU_t - X^*)$. Based on this approximation, the iterations of SubGM can be approximated as $U_{t+1} \approx U_t - \eta_t \varphi_t \cdot \frac{(U_tU_t - X^*)U_t}{\|U_tU_t - X^*\|_F}$. Consequently, with the choice of $\eta_t = \eta_0 \varphi_t^{-1} \|U_tU_t - X^*\|_F$, the iterations of SubGM reduce to

$$U_{t+1} \approx U_t - \eta_0 \cdot (U_tU_t - X^*)U_t \tag{3}$$

which are precisely the iterations of GD with a constant step size η_0, applied to the ideal, noiseless ℓ_2-loss function $\bar{f}_{\ell_2}(U) = \|UU^\top - X^*\|_F^2$. This implies that, under Sign-RIP, SubGM behaves similar to GD with a constant step size, when applied to $\bar{f}_{\ell_2}(U)$. A caveat of this analysis is that the proposed step sizes are in terms of $\varphi_t^{-1}\|U_tU_t - X^*\|_F$, which is not known *a priori*. In the noiseless scenario, Sign-RIP can be invoked to show that $\varphi_t^{-1}\|U_tU_t - X^*\|_F$ can be accurately estimated as $\frac{\pi}{2m}\|\mathbf{y} - \mathcal{A}(U_tU_t)\|_1$. However, with noisy measurements, the value of $\|U_tU_t - X^*\|_F$ *cannot* be estimated merely based on $\|\mathbf{y} - \mathcal{A}(U_tU_t^\top)\|_1$, since the ℓ_1-loss function $\|\mathbf{y} - \mathcal{A}(U_tU_t^\top)\|_1$ is no longer an unbiased estimator of $\|U_tU_t^\top - X^*\|_F$ and is highly sensitive to the magnitude of the noise. To alleviate this issue, we instead consider geometrically decaying step sizes $\eta_t = \frac{\eta_0}{\|Q\|_F}\rho^t$, where $Q \in \mathcal{Q}(U_tU_t - X^*)$ and $0 < \rho < 1$ is a predefined decay rate. Due to Sign-RIP, we have $\|Q\|_F \approx \varphi(U_tU_t^\top - X^*)$, which implies that $U_{t+1} \approx U_t - \eta_0\rho^t \left\|U_tU_t^\top - X^\star\right\|_F^{-1} \left(U_tU_t^\top - X^\star\right) U_t$. A closer look at this equation reveals that, at every iteration, SubGD points to the negative gradient of $\|U_tU_t^\top - X^*\|_F^2$, while the geometrically decaying step size $\eta_t = \eta_0\rho^t$ guarantees the convergence of the algorithm.

Theorem 7 *Assume that $r' \geq r^* = 1$, and the measurements satisfy the Sign-RIP condition with parameters $(\min\{r'+1, d\}, \delta)$, where $\delta \lesssim 1$ and $\varphi(X)$ is strictly positive and uniformly bounded. Suppose that the initialization matrix B_0 is chosen via Algorithm 2 in Appendix, and U_T is obtained via SubGM. Moreover, suppose that $\alpha \asymp \sqrt{\delta}/\sqrt[4]{r'}$, and the step size $\eta_t = \eta_0\rho^t$ satisfying $\eta_0 \lesssim \delta$ and $\rho = 1 - \Theta\left(\eta_0/\log\frac{1}{\alpha}\right)$. Then, after $T \asymp \log\left(\frac{r'}{\delta}\right)/\eta_0$ iterations, we have*

$$\left\|U_TU_T^\top - X^\star\right\|_F^2 \lesssim \delta^2 \log^2\left(\frac{r'}{\delta}\right). \tag{4}$$

The above theorem implies that, for any $r' \geq r^* = 1$ (including $r' = d$), SubGM converges to the true low-rank solution at a (nearly) dimension-free rate without any explicit regularization or rank constraint under the Sign-RIP condition. Moreover, our result holds for arbitrarily large values of p, provided that the number of measurements scales accordingly. Combining Theorem 7 and Theorem 4 leads to an end-to-end sample complexity guarantee for SubGM with Gaussian measurements.

Corollary 8 *Assume that $\{A_i\}_{i=1}^m$ are standard Gaussian matrices, and that the noise vector $\mathbf{s}$ satisfies Assumption 1. Moreover, suppose that α, η_t, and T are chosen according to Theorem 7. Then, SubGM satisfies the error bound (4) with an overwhelming probability, provided that $m \gtrsim \frac{dr'\log\left(\frac{1}{(1-p)\delta}\right)}{\delta^4(1-p)^4}$.*

6. Conclusion

Existing techniques for analyzing low-rank matrix recovery presume and rely on different variants of restricted isometry property (RIP). However, these notions fail in the robust settings, where a number of measurements are grossly corrupted with noise. In this work, we propose a robust restricted isometry property, called Sign-RIP, that addresses this fundamental issue. Based on Sign-RIP, we paint a full picture for the landscape of robust rank-1 matrix recovery problem, both in the exact and over-parameterized regimes. In the exact setting, we show that all the critical points of the robust matrix recovery are close to the true solution, or have small norm. In the over-parameterized regime, we show that a simple subgradient method converges to the ground truth.

Although our results on robust matrix recovery is restricted to rank-1 case, the proposed framework is general, and it paves the way towards a better understanding of the problem in more general settings. In particular, our developed guarantees for sign-RIP hold for the general rank-r matrices, and hence, can be potentially used to study the global landscape of more general robust matrix recovery problems in both exact and over-parameterized regimes.

Acknowledgement

The authors sincerely thank Richard Y. Zhang and Cedric Josz for their valuable input on several drafts of this paper. S.F. is supported by MICDE Catalyst Grant and MIDAS PODS Grant.

References

[1] Srinadh Bhojanapalli, Behnam Neyshabur, and Nathan Srebro. Global optimality of local search for low rank matrix recovery. *arXiv preprint arXiv:1605.07221*, 2016.

[2] Thierry Bouwmans and El Hadi Zahzah. Robust pca via principal component pursuit: A review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 122:22–34, 2014.

[3] Emmanuel J Candes. The restricted isometry property and its implications for compressed sensing. *Comptes rendus mathematique*, 346(9-10):589–592, 2008.

[4] Venkat Chandrasekaran, Sujay Sanghavi, Pablo A Parrilo, and Alan S Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

[5] Frank H Clarke. *Optimization and nonsmooth analysis*. SIAM, 1990.

[6] Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping from saddle points—online stochastic gradient for tensor decomposition. In *Conference on learning theory*, pages 797–842. PMLR, 2015.

[7] Rong Ge, Jason D Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. *arXiv preprint arXiv:1605.07272*, 2016.

[8] Chi Jin, Rong Ge, Praneeth Netrapalli, Sham M Kakade, and Michael I Jordan. How to escape saddle points efficiently. In *International Conference on Machine Learning*, pages 1724–1732. PMLR, 2017.

[9] Xiao Li, Zhihui Zhu, Anthony Man-Cho So, and Rene Vidal. Nonconvex robust low-rank matrix recovery. *SIAM Journal on Optimization*, 30(1):660–686, 2020.

[10] Yuanzhi Li, Tengyu Ma, and Hongyang Zhang. Algorithmic regularization in over-parameterized matrix sensing and neural networks with quadratic activations. pages 2–47, 2018.

[11] Zhang Liu and Lieven Vandenberghe. Interior-point method for nuclear norm approximation with application to system identification. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1235–1256, 2010.

[12] Xiao Luan, Bin Fang, Linghui Liu, Weibin Yang, and Jiye Qian. Extracting sparse error of robust pca for face recognition in the presence of varying illumination and occlusion. *Pattern Recognition*, 47(2): 495–508, 2014.

[13] Xin Luo, Mengchu Zhou, Yunni Xia, and Qingsheng Zhu. An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. *IEEE Transactions on Industrial Informatics*, 10(2):1273–1284, 2014.

[14] Balas Kausik Natarajan. Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234, 1995.

[15] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

[16] Bodhisattva Sen. A gentle introduction to empirical process theory and applications, 2018.

[17] Tian Tong, Cong Ma, and Yuejie Chi. Low-rank matrix recovery with scaled subgradient methods: Fast and robust convergence without the condition number. *IEEE Transactions on Signal Processing*, 2021.

[18] Roman Vershynin. High-dimensional probability, 2019.

[19] Min Zhang, Zheng-Hai Huang, and Ying Zhang. Restricted p-isometry properties of nonconvex matrix recovery. *IEEE Transactions on Information Theory*, 59(7):4316–4323, 2013.

[20] Richard Y Zhang. Sharp global guarantees for nonconvex low-rank matrix recovery in the overparameterized regime. *arXiv preprint arXiv:2104.10790*, 2021.

[21] Richard Y Zhang, Somayeh Sojoudi, and Javad Lavaei. Sharp restricted isometry bounds for the inexistence of spurious local minima in nonconvex matrix recovery. *Journal of Machine Learning Research*, 20(114):1–34, 2019.

[22] Jiacheng Zhuo, Jeongyeol Kwon, Nhat Ho, and Constantine Caramanis. On the computational and statistical complexity of over-parameterized matrix sensing. *arXiv preprint arXiv:2102.02756*, 2021.

Appendix

In the appendix, we first provide our pseduo code for subgradient method and our proposed initialization scheme (Algorithms 1 and 2). Then, we provide extensive numerical experiments to verify our theoretical results. Finally, we present the detailed proofs of our theoretical results.

Algorithm 1 Subgradient Method

Input: measurement matrices $\{A_i\}_{i=1}^m$, measurement vector $\mathbf{y} = [y_1, \cdots, y_m]^\top$, number of iterations T, an upper bound on the rank r', and an initialization matrix $B_0 \in \mathbb{R}^{d \times r'}$;
Output: Solution $\hat{X}_T = U_T U_T^\top$ to (2);
Initialize $U_0 = B_0$.
for $t \leq T$ **do**
  Compute a subgradient $D_t \in \partial f_{\ell_1}(U_t)$;
  Select the step size η_t;
  Set $U_{t+1} \leftarrow U_t - \eta_t D_t$;
end for

Algorithm 2 Spectral Initialization

Input: measurement matrices $\{A_i\}_{i=1}^m$, measurement vector $\mathbf{y} = [y_1, \cdots, y_m]^\top$, an upper bound on the rank r', and an initial scaling factor α;
Output: An initialization matrix $B_0 \in \mathbb{R}^{d \times r'}$;
Calculate $C = \frac{1}{m}\sum_{i=1}^m \mathrm{Sign}(y_i) A_i$, and its normalized variant $\hat{X} = C / \|C\|_F$;
Compute the eigenvalue decomposition $\hat{X} = V \Sigma V^\top$;
Define $\Sigma_+^{r'}$ as the top $r' \times r'$ sub-matrix of Σ corresponding to r' largest eigenvalues of $\hat{X}$, whose negative values are replaced by 0;
Set $B_0 = \alpha V \left(\Sigma_+^{r'}\right)^{1/2}$.

Appendix A. Numerical Experiments

In this section, we provide extensive numerical experiments to verify our theoretical guarantees, and to shed light on possible future directions. All simulations are run on a desktop computer with an Intel Core i9 3.50 GHz CPU and 128GB RAM. The reported results are for an implementation in Python.

A.1. Relationship between dimension and measurement number

In this experiment, we analyze the relationship between the number of measurements m and dimension d. Our theoretical result suggests that $m \gtrsim dr'$ is enough to ensure the convergence of SubGM. To empirically verify this, we change d from 10 to 100 and set $r' = d$. Moreover, we set the corruption probability to $p = 0.1$. Moreover, each element of the noise is generated according to a standard Gaussian distribution. The step sizes are selected as $\eta_t = \eta_0 \rho^t$, where $\eta_0 = 0.4$ and $\rho = 0.98$. For each group of parameters, we run 5 independent trials and plot the average log-residual for the last iteration in Figure 3. It can be seen that, in order to ensure the same value for the error, the number of measurements should grow almost linearly with the dimension, which is in line with our theoretical result.

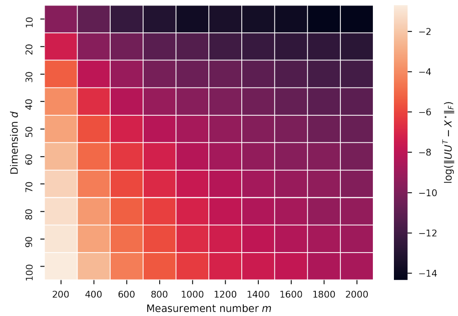


Figure 3: The error with respect to the number of measurements and dimension

A.2. Effect of Different Noise Magnitudes

In this experiment, we certify the robustness of SubGM against large noise values. Our theoretical result suggests that the convergence of SubGM is independent of the noise magnitude. To verify this, we set the dimension and the number of measurements to $d = 50$ and $m = 500$, respectively. Moreover, we set the corruption probability to $p = 0.1$, and select each element of the noise according to a Gaussian distribution $s_i \sim \mathcal{N}(0, \sigma^2)$ with varying variance σ^2. Finally, we set the step size to $\eta_t = \eta_0 \rho^t$, where $\eta_0 = 0.25$, and $\rho = 0.99$. Based on Figure 4, it can be seen that increasing variance slightly deteriorates the error. However, beyond a certain threshold, increasing variance does not have any effect on the error.

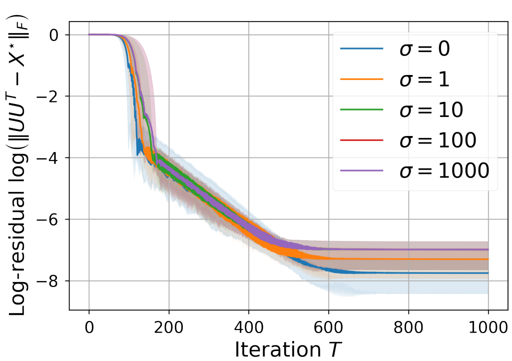


Figure 4: Effect of noise variance.

A.3. Effect of Different Types of Noise

In this experiment, we study the effect of different types of noise on the performance of SubGM. In particular, we choose five different types of distribution for the noise: Gaussian, uniform, Laplace, Cauchy, and Rademacher. The experiments are designed under the same settings as Subsection A.2. Moreover, for all types of noise (except for the Cauchy distribution), we set the variance to 100. As can be seen in Figure 5, SubGM is insensitive to the particular choice of noise.

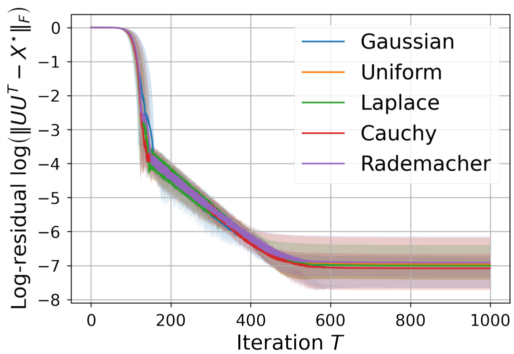


Figure 5: Effect of different types of noise.

A.4. Effect of different step size regimes

Finally, we explore the effect of different step sizes in both noiseless and noisy case under the same settings as Section A.2. In the noiseless case, we compare four different types of step sizes: 1) $\eta_t = \eta_0\rho^t$ with $\eta_0 = 0.25, \rho = 0.99$; 2) $\eta_t = \frac{\eta_0}{t}$ with $\eta_0 = 2.0$; 3) $\eta_t = \frac{\eta_0}{\sqrt{t}}$ with $\eta_0 = 0.3$; and 4) our proposed choice $\eta_t = \eta_0\frac{1}{m}\left\|\mathbf{y} - \mathcal{A}(U_tU_t^\top)\right\|_1$ where $\eta_0 = 0.25$ (see our discussion in Section 5 and Appendix D). As can be seen in Figure 6a, SubGM converges to the true solution with all of the aforementioned step sizes. However, our proposed step size leads to the fastest convergence rate. In the noisy case, we compare the performance of five different step sizes: 1) $\eta_t = \eta_0\rho^t$ with $\eta_0 = 0.45, \rho = 0.98$; 2) $\eta_t = \frac{\eta_0}{t}$ with $\eta_0 = 2.0$; 3) $\eta_t = \frac{\eta_0}{\sqrt{t}}$ with $\eta_0 = 0.3$; 4) $\eta_t = \eta_0\frac{1}{m}\left\|\mathbf{y} - \mathcal{A}(U_tU_t^\top)\right\|_1$ with $\eta_0 = 0.25$; 5) our proposed choice $\eta_t = \frac{\eta_0}{\|D_t\|_F}\rho^t$, where $D_t \in \mathcal{M}(U_tU_t - X^*)$, $\eta_0 = 0.4$, and $\rho = 0.99$. From Figure 6b, it is evident that the step size $\eta_t = \eta_0\frac{1}{m}\left\|\mathbf{y} - \mathcal{A}(U_tU_t^\top)\right\|_1$, which was the best choice in the noiseless case, does not result in the convergence of SubGM to the true solution in the noisy case. As mentioned before, this is due to the sensitivity of $\frac{1}{m}\left\|\mathbf{y} - \mathcal{A}(U_tU_t^\top)\right\|_1$ to outliers. Moreover, our proposed step size outperforms its vanilla counterpart. Finally, the polynomially decaying step size $\eta_t \propto \frac{1}{t}$ performs slightly better than our proposed step size. Motivated by this interesting observation, we will study the performance of SubGM with polynomially decaying step sizes in the future.

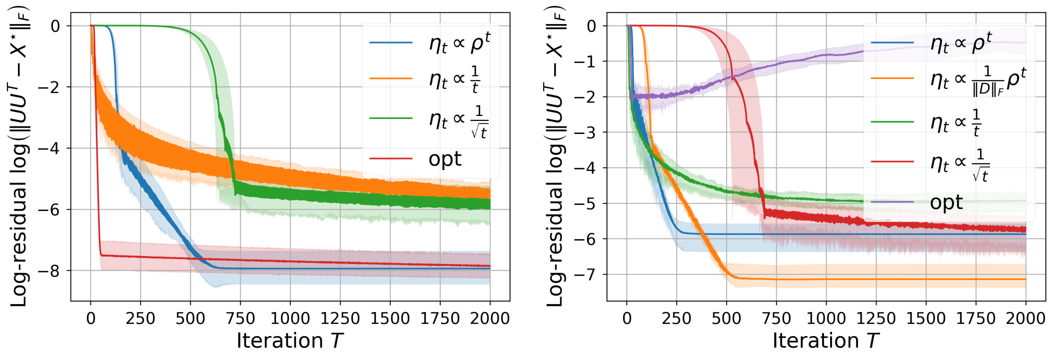
Appendix B. Proofs for Sign-RIP

In this section, we provide the proofs for Theorem 4. As a first step, we start with the noiseless case, and show that a weaker version of Theorem 4 can be obtained directly from the so-called ℓ_1/ℓ_2-RIP condition.

Lemma 9 (ℓ_1/ℓ_2-RIP, Proposition 1 in [9]) *Let $r \geq 1$ be given, suppose measurements $\{A_i\}_{i=1}^m$ have i.i.d. standard Gaussian entries with $m \gtrsim dr$. Then for any $0 < \delta < \sqrt{\frac{2}{\pi}}$, there exists a universal constant $c > 0$, such that with probability of at least $1 - e^{-cm\delta^2}$, we have*

$$\sup_{X \in \mathbb{S}_r}\left|\frac{1}{m}\sum_{i=1}^m |\langle A_i, X\rangle| - \sqrt{\frac{2}{\pi}}\|X\|_F\right| \leq \sqrt{\frac{2}{\pi}}\delta. \tag{5}$$

Based on the above ℓ_1/ℓ_2-RIP condition, we proceed to prove a weaker version of Theorem 4.



(a) noiseless (b) noisy

Figure 6: Effect of different step size regimes.

Proposition 10 *Assume that the measurement matrices $\{A_i\}_{i=1}^m$ have i.i.d. standard Gaussian entries, and that* $\mathbf{s} = 0$*. Then, the sign-RIP condition holds with parameters* $(r, \delta), \delta \leq \sqrt{\frac{2}{\pi}}$ *and a constant scaling function* $\varphi(X) = \sqrt{\frac{2}{\pi}}$ *with probability of at least* $1 - Ce^{-cm\delta^4}$*, provided that* $m \gtrsim d^2$.

Proof Without loss of generality, we assume that $\|X\|_F = 1$. For any given $0 < \delta \leq \sqrt{\frac{2}{\pi}}$ and any $D(X) \in \frac{1}{m}\sum_{i=1}^m \mathrm{Sign}(\langle A_i, X\rangle)A_i$, we have

$$\begin{aligned} &\sup_{X\in\mathbb{S}_r}\left\|D(X) - \sqrt{\frac{2}{\pi}}X\right\|_F^2 \\ &\stackrel{(a)}{=} \sup_{X\in\mathbb{S}_r}\|D(X)\|_F^2 - \sqrt{\frac{8}{\pi}}\frac{1}{m}\sum_{i=1}^m |\langle A_i, X\rangle| + \frac{2}{\pi} \\ &\leq \sup_{X\in\mathbb{S}_r}\|D(X)\|_F^2 - \sqrt{\frac{8}{\pi}}\inf_{X\in\mathbb{S}_r}\frac{1}{m}\sum_{i=1}^m |\langle A_i, X\rangle| + \frac{2}{\pi} \\ &\stackrel{(b)}{\leq} \sup_{X\in\mathbb{S}_r}\|D(X)\|_F^2 + \sqrt{\frac{8}{\pi}}\sqrt{\frac{2}{\pi}}\delta - \sqrt{\frac{8}{\pi}}\sqrt{\frac{2}{\pi}} + \frac{2}{\pi} \\ &= \sup_{X\in\mathbb{S}_r}\|D(X)\|_F^2 + \frac{4}{\pi}\delta - \frac{2}{\pi} \end{aligned} \tag{6}$$

with probability of at least $1 - Ce^{-cm\delta^2}$. Here, (a) follows from $\langle D(X), X\rangle = \frac{1}{m}\sum_{i=1}^m |\langle A_i, X\rangle|$, and (b) uses ℓ_1/ℓ_2-RIP condition from Lemma 9. Then, recall that for an arbitrary $M \in \mathbb{R}^{d\times d}$, we have

$$\|M\|_F = \sup_{Y\in\mathbb{S}}\langle M, Y\rangle. \tag{7}$$

This implies

$$\begin{aligned}
\sup_{X\in\mathbb{S}_r}\|D(X)\|_F^2 &\le \sup_{X,Y\in\mathbb{S}}\left(\langle D(X),Y\rangle\right)^2\\
&\overset{(c)}{=} \sup_{Y\in\mathbb{S}}\left(\frac{1}{m}\sum_{i=1}^m |\langle A_i,Y\rangle|\right)^2\\
&\overset{(d)}{\le} \frac{2}{\pi}(1+\delta)^2\\
&\overset{(e)}{\le} \frac{2}{\pi}+\frac{6}{\pi}\delta
\end{aligned}\tag{8}$$

with high probability $1-Ce^{-cm\delta^2}$. Here, (c) uses the fact that for a fixed Y, the supremum over X is taken exactly at $X=Y$, (d) uses the ℓ_1/ℓ_2-RIP condition, and (e) uses the assumption $\delta\le 1$.

Combining these two parts, we obtain

$$\sup_{X\in\mathbb{S}}\left\|D(X)-\sqrt{\frac{2}{\pi}}X\right\|_F^2 \le \frac{10}{\pi}\delta \tag{9}$$

with probability of at least $1-Ce^{-c'm\delta^2}$. Therefore, upon choosing $\delta'=\sqrt{5\delta}$, we obtain

$$\sup_{X\in\mathbb{S}}\left\|D(X)-\sqrt{\frac{2}{\pi}}X\right\|_F = \sup_{X,Y\in\mathbb{S}}\left\langle D(X)-\sqrt{\frac{2}{\pi}}X, Y\right\rangle \le \sqrt{\frac{2}{\pi}}\delta' \tag{10}$$

with probability of at least $1-Ce^{-cm\delta'^4}$. ■

Despite its simplicity, the above analysis has two major drawbacks: (1) its sample complexity scales with d^2, as opposed to dr in Theorem 4, (2) it is not clear how to extend this analysis to the noisy case. To address these issues and prove Theorem 4, we need a more in-depth analysis of the sign-RIP condition. First, we provide an intermediate lemma.

Lemma 11 *Assume that the measurement matrices $\{A_i\}_{i=1}^m$ defining the linear operator $\mathcal{A}(\cdot)$ have Gaussian entries, and that the noise vector $\mathbf{s}$ satisfies Assumption 1. Then, for every $D\in\mathcal{M}(X)$, we have*

$$\mathbb{E}[D]=\sqrt{\frac{2}{\pi}}\left(1-p+p\mathbb{E}\left[e^{-s_i^2/(2\|X\|_F)}\right]\right)\frac{X}{\|X\|_F} \tag{11}$$

where the expectation is taken with respect to both $\mathbf{s}$ and $\{A_i\}_{i=1}^m$.

Proof To prove this lemma, it is enough to show that for any $X,Y\in\mathbb{R}^{d\times d}$, we have

$$\mathbb{E}\left[\mathrm{Sign}\left(s+\langle A,X\rangle\right)\langle A,Y\rangle\right]=\sqrt{\frac{2}{\pi}}\mathbb{E}\left[e^{-s^2/2\|X\|_F^2}\right]\left\langle\frac{X}{\|X\|_F},Y\right\rangle. \tag{12}$$

Without loss of generality, we assume that $\|X\|_F=\|Y\|_F=1$ and both X and Y are symmetric. Moreover, for the symmetric Gaussian matrix A_i, its off-diagonal elements are from $\mathcal{N}(0,1/2)$, and its diagonal elements are from $\mathcal{N}(0,1)$. Now let us denote $u := \langle A, X\rangle, v := \langle A, Y\rangle, \rho := \mathrm{Cov}(u,v) = \langle X, Y\rangle$. Then

$$\begin{aligned}
\mathbb{E}\left[\mathrm{Sign}\left(s + \langle A, X\rangle\right)\langle A, Y\rangle\right] &= \mathbb{E}\left[\mathrm{Sign}\left(s+u\right)v\right] \\
&\overset{(a)}{=} \rho\mathbb{E}\left[\mathrm{Sign}(u+s)u\right] \\
&= \rho\mathbb{E}_s\left[\int_{-s}^{\infty} u\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du - \int_{-\infty}^{-s} u\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du\right] \\
&= \rho\mathbb{E}_s\left[\int_{-s}^{\infty} u\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du + \int_{s}^{\infty} u\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du\right] \\
&= 2\rho\mathbb{E}_s\left[\int_{|s_i|}^{\infty} u\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du\right] \\
&= \sqrt{\frac{2}{\pi}}\langle X, Y\rangle\mathbb{E}_s\left[\int_{|s|}^{\infty} d\left(-e^{-u^2/2}\right)\right] \\
&= \sqrt{\frac{2}{\pi}}\langle X, Y\rangle\mathbb{E}_s\left[e^{-s^2/2}\right].
\end{aligned} \tag{13}$$

Here (a) uses the fact that $v|u,s \sim \mathcal{N}(\rho u, 1-\rho^2)$ since $s \perp\!\!\!\perp u, v$. This together with the variational form of the Frobenius norm implies

$$\mathbb{E}\left[\mathrm{Sign}(s + \langle A, X\rangle)A\right] = \sqrt{\frac{2}{\pi}}\mathbb{E}\left[e^{-s^2/2\|X\|_F^2}\right]\frac{X}{\|X\|_F}, \tag{14}$$

for any $X \in \mathbb{R}^{d\times d}$. On the other hand, it is easy to verify that $\mathbb{E}\left[\mathrm{Sign}\left(\langle A, X\rangle\right)A\right] = \sqrt{\frac{2}{\pi}}\frac{X}{\|X\|_F}$. The proof is completed by noting that the size of the noisy measurements is equal to pm. ■

B.1. Proof of Theorem 4

For the sake of simplicity, we assume that pm is an integer. Moreover, we abuse the notation and use $\mathrm{Sign}(\cdot)$ as a regular function taking an arbitrary value $\mathrm{Sign}(0) \in [-1, 1]$. To prove Theorem 4, we first present an intermediate lemma, which holds for any fixed $Y, Y \in \mathbb{S}$.

Lemma 12 *There exists a universal constant c, for any $\delta > 0$, we have*

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m}\mathrm{Sign}(\langle A_i, X\rangle + s_i)\langle A_i, Y\rangle - \sqrt{\frac{2}{\pi}}\left(1 - p + p\mathbb{E}\left[e^{-s_i^2/2}\right]\right)\langle X, Y\rangle\right| \geq \delta\right) \leq 2e^{-cm\delta^2}. \tag{15}$$

Proof [Proof of Lemma 12]

We first show that $\mathrm{Sign}(\langle A_i, X\rangle + s_i)\langle A_i, Y\rangle$ is a sub-Gaussian random variable. First notice that $\langle A_i, Y\rangle \sim \mathcal{N}(0,1)$ since $\|Y\|_F = 1$. Moreover, notice that $\|\mathrm{Sign}(\langle A_i, X\rangle + s_i)\langle A_i, Y\rangle\|_{\ell^{2k}} \leq \|\langle A_i, Y\rangle\|_{\ell^{2k}}$ for $\forall k \in \mathbb{N}_+$, where $\|M\|_{\ell^{2k}}$ is defined as $\left(\mathbb{E}\left[|M|^p\right]\right)^{1/p}$. Therefore, based on equivalent definition of sub-Gaussian random variables (see Definition 33), we obtain that $\mathrm{Sign}(\langle A_i, X\rangle + s_i)\langle A_i, Y\rangle$ is also $O(1)$-sub-Gaussian. Moreover, according to the proof of Lemma 11, we have

$\mathbb{E}\left[\frac{1}{m}\sum_{i\in S}\mathrm{Sign}(\langle A_i, X\rangle)\langle A_i, Y\rangle\right] = \sqrt{\frac{2}{\pi}}(1-p)\langle X, Y\rangle$ and $\mathbb{E}\left[\frac{1}{m}\sum_{i\in S}\mathrm{Sign}(\langle A_i, X\rangle + s_i)\langle A_i, Y\rangle\right] = \sqrt{\frac{2}{\pi}}p\mathbb{E}\left[e^{-s_i^2/2}\right]\langle X, Y\rangle$. This together with the standard concentration bound on sub-Gaussian random variables leads to

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{i\notin S}\mathrm{Sign}(\langle A_i, X\rangle)\langle A_i, Y\rangle - \sqrt{\frac{2}{\pi}}(1-p)\langle X, Y\rangle\right| \geq \frac{1}{2}\delta\right) \leq 2e^{-cm\delta^2/(1-p)}, \tag{16}$$

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{i\in S}\mathrm{Sign}(\langle A_i, X\rangle + s_i)\langle A_i, Y\rangle - \sqrt{\frac{2}{\pi}}p\mathbb{E}\left[e^{-s_i^2/2}\right]\langle X, Y\rangle\right| \geq \frac{1}{2}\delta\right) \leq 2e^{-cm\delta^2/p}. \tag{17}$$

which implies

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m}\mathrm{Sign}(\langle A_i, X\rangle + s_i)\langle A_i, Y\rangle - \sqrt{\frac{2}{\pi}}\left(1-p+p\mathbb{E}\left[e^{-s_i^2/2}\right]\right)\langle X, Y\rangle\right| \geq \delta\right)\\ \leq& 4e^{-cm\delta^2\min\left\{\frac{1}{p},\frac{1}{1-p}\right\}} \leq 2e^{-c'm\delta^2}.\end{aligned} \tag{18}$$

■

Consider an ϵ-covering $\mathbb{S}_{\epsilon,r} \subseteq \mathbb{S}_r$ with a property that for every $X \in \mathbb{S}_r$, there exists $\bar{X} \in \mathbb{S}_{\epsilon,r}$ that satisfies $\|X - \bar{X}\|_F \leq \epsilon$. According to Lemma 30, there exists an ϵ-covering that satisfies $|\mathbb{S}_{\epsilon,r}| \leq \left(\frac{9}{\epsilon}\right)^{(2d+1)r}$. For any $\bar{X} \in \mathbb{S}_{\epsilon,r}$, define $B_r(\bar{X}, \epsilon) = \{X \in \mathbb{S}_r : \left\|X - \bar{X}\right\|_F \leq \epsilon\}$. Then, for any $\bar{X}, \bar{Y}$ and $X, Y \in B_r(\bar{X}, \epsilon) \times B_r(\bar{Y}, \epsilon)$, we have

$$\left|\langle X, Y\rangle - \left\langle \bar{X}, \bar{Y}\right\rangle\right| \leq \left|\left\langle X - \bar{X}, \bar{Y}\right\rangle\right| + \left|\left\langle X, Y - \bar{Y}\right\rangle\right| \leq 2\epsilon. \tag{19}$$

Based on the defined ϵ-covering, one can write

$$
\begin{aligned}
&\sup_{X,Y\in\mathbb{S}_r}\left|\frac{1}{m}\sum_{i=1}^m \mathrm{Sign}(\langle A_i,X\rangle+s_i)\langle A_i,Y\rangle-\sqrt{\frac{2}{\pi}}\left(1-p+p\mathbb{E}\left[e^{-s_i^2/2}\right]\right)\langle X,Y\rangle\right|\\
&=\sup_{\bar X,\bar Y\in\mathbb{S}_{\epsilon,r}}\sup_{\substack{X\in B_r(\bar X,\epsilon)\\ Y\in B_r(\bar Y,\epsilon)}}\left|\frac{1}{m}\sum_{i=1}^m \mathrm{Sign}(\langle A_i,X\rangle+s_i)\langle A_i,Y\rangle-\sqrt{\frac{2}{\pi}}\left(1-p+p\mathbb{E}\left[e^{-s_i^2/2}\right]\right)\langle X,Y\rangle\right|\\
&\le\underbrace{\sup_{\bar X,\bar Y\in\mathbb{S}_{\epsilon,r}}\left|\frac{1}{m}\sum_{i=1}^m \mathrm{Sign}(\left\langle A_i,\bar X\right\rangle+s_i)\left\langle A_i,\bar Y\right\rangle-\sqrt{\frac{2}{\pi}}\left(1-p+p\mathbb{E}\left[e^{-s_i^2/2}\right]\right)\left\langle \bar X,\bar Y\right\rangle\right|}_{\text{(A)}}\\
&+\underbrace{\sup_{\bar X,\bar Y\in\mathbb{S}_{\epsilon,r}}\sup_{Y\in B_r(\bar Y,\epsilon)}\left|\frac{1}{m}\sum_{i=1}^m \mathrm{Sign}(\left\langle A_i,\bar X\right\rangle+s_i)\langle A_i,Y\rangle-\mathrm{Sign}(\left\langle A_i,\bar X\right\rangle+s_i)\left\langle A_i,\bar Y\right\rangle\right|}_{\text{(B)}}\\
&+\underbrace{\sup_{\bar X\in\mathbb{S}_{\epsilon,r},Y\in\mathbb{S}_r}\sup_{X\in B(\bar X,\epsilon)}\left|\frac{1}{m}\sum_{i=1}^m \mathrm{Sign}(\left\langle A_i,\bar X\right\rangle+s_i)\langle A_i,Y\rangle-\mathrm{Sign}(\langle A_i,X\rangle+s_i)\langle A_i,Y\rangle\right|}_{\text{(C)}}\\
&+\underbrace{\sup_{\bar X,\bar Y\in\mathbb{S}_{\epsilon,r}}\sup_{\substack{X\in B_r(\bar X,\epsilon)\\ Y\in B_r(\bar Y,\epsilon)}}\sqrt{\frac{2}{\pi}}\left(1-p+p\mathbb{E}\left[e^{-s_i^2/2}\right]\right)\left|\langle X,Y\rangle-\left\langle \bar X,\bar Y\right\rangle\right|.}_{\le\sqrt{\frac{8}{\pi}}\epsilon\text{ by (19)}}
\end{aligned}
\tag{20}
$$

We control the first three terms separately. Based on a union bound and Lemma 12, we have

$$
\text{(A)}\le\delta_1\quad\text{with probability of at least }1-2\left|\mathbb{S}_{\epsilon,r}\right|^2e^{-cm\delta_1^2}. \tag{21}
$$

Moreover, one can write

$$
\begin{aligned}
\text{(B)}&\le\sup_{\bar Y\in\mathbb{S}_{\epsilon,r}}\sup_{Y\in B_r(\bar Y,\epsilon)}\frac{1}{m}\sum_{i=1}^m\left|\left\langle A_i,Y-\bar Y\right\rangle\right|\\
&\overset{(a)}{\le}\epsilon\sup_{Z\in\mathbb{S}_{2r}}\frac{1}{m}\sum_{i=1}^m\left|\langle A_i,Z\rangle\right|\\
&\le\sqrt{\frac{2}{\pi}}\epsilon(1+\delta_2)
\end{aligned}
\tag{22}
$$

with probability of at least $1-Ce^{c_1dr\log\frac{1}{\delta_2}-c_2m\delta_2^2}$. Here we used ℓ_1/ℓ_2 RIP condition from Lemma 9, and the fact for X,Y with ranks at most r, we have $\mathrm{rank}(X-Y)\le\mathrm{rank}(X)+\mathrm{rank}(Y)\le 2r$. Next, we provide an upper bound for (C). First by Cauchy-Schwartz inequality, we have

$$
\text{(C)} \leq \sup_{\bar{X}\in\mathbb{S}_{\epsilon,r}} \sup_{X\in B_r(\bar{X},\epsilon)} \left(\underbrace{\frac{1}{m}\sum_{i=1}^{m}\left(\mathrm{Sign}(\langle A_i,\bar{X}\rangle + s_i) - \mathrm{Sign}(\langle A_i, X\rangle + s_i)\right)^2}_{\text{(C1)}} \right)^{\frac{1}{2}} \sup_{Y\in\mathbb{S}_r}\left(\frac{1}{m}\sum_{i=1}^{m}\langle A_i, Y\rangle^2\right)^{\frac{1}{2}}. \tag{23}
$$

The second term in the above inequality can be readily controlled via ℓ_2-RIP (see Definition 1):

$$
\sup_{Y\in\mathbb{S}_r}\frac{1}{m}\sum_{i=1}^{m}\langle A_i, Y\rangle^2 \leq 1+\delta_3 \tag{24}
$$

which holds with probability of at least $1 - Ce^{c_1 dr\log\frac{1}{\delta_3} - c_2 m\delta_3^2}$ for any $0<\delta_3<1$. For the remaining part (C1), first note that if $|\langle A_i, X-\bar{X}\rangle| \leq |\langle A_i, \bar{X}+s_i\rangle|$, then $\mathrm{Sign}(\langle A_i,\bar{X}\rangle + s_i) = \mathrm{Sign}(\langle A_i, X\rangle + s_i)$. This leads to

$$
\begin{aligned}
\sup_{\bar{X}\in\mathbb{S}_{\epsilon,r}} \sup_{X\in B_r(\bar{X},\epsilon)} \text{(C1)} &\leq \sup_{\bar{X},X}\frac{4}{m}\sum_{i=1}^{m}\mathbb{1}\left(|\langle A_i, X-\bar{X}\rangle| \geq |\langle A_i,\bar{X}\rangle + s_i|\right)\\
&\overset{(a)}{\leq} \sup_{X,\bar{X}}\frac{4}{m}\sum_{i=1}^{m}\mathbb{1}\left(|\langle A_i, X-\bar{X}\rangle| \geq t\right) + \mathbb{1}\left(|\langle A_i,\bar{X}\rangle + s_i| \leq t\right)\\
&\leq \sup_{Z\in\epsilon\mathbb{S}_{2r}}\frac{4}{m}\sum_{i=1}^{m}\mathbb{1}\left(|\langle A_i, Z\rangle| \geq t\right) + \sup_{\bar{X}}\frac{4}{m}\sum_{i=1}^{m}\mathbb{1}\left(|\langle A_i,\bar{X}\rangle + s_i| \leq t\right) \qquad (25)\\
&\overset{(b)}{\leq} \sup_{Z\in\epsilon\mathbb{S}_{2r}}\frac{4}{m}\sum_{i=1}^{m}\mathbb{1}\left(|\langle A_i, Z\rangle| \geq t\right) + 4\mathbb{E}\left[\mathbb{1}\left(|\langle A_i,\bar{X}\rangle + s_i| \leq t\right)\right] + \delta_4\\
&\overset{(c)}{\leq} \sup_{Z\in\epsilon\mathbb{S}_{2r}}\frac{4}{m}\sum_{i=1}^{m}\mathbb{1}\left(|\langle A_i, Z\rangle| \geq t\right) + 4t + \delta_4
\end{aligned}
$$

which holds with probability of at least $1 - C|\mathbb{S}_{\epsilon,r}|e^{-cm\delta_4^2}$, where $t>0$ is an arbitrary scalar. Here, in (a) we use a simple fact that for two arbitrary random variables A, B and a scalar $t\in\mathbb{R}$, the event $\{A\geq B\}$ is included in $\{A\geq t\}\cup\{B\leq t\}$. Moreover, in (b) we use a union bound and Hoeffding's inequality. Finally, in (c) we use the anti-concentration inequality conditioned on s_i. For the first term in the above inequality, we have the following lemma.

Lemma 13 *We have*

$$
\mathbb{E}\left[\sup_{Z\in\epsilon\mathbb{S}_{2r}}\frac{4}{m}\sum_{i=1}^{m}\mathbb{1}\left(|\langle A_i, Z\rangle| \geq t\right) - \mathbb{P}\left(|\langle A_i, Z\rangle| \geq t\right)\right] \lesssim e^{-t^2/4\epsilon^2}\sqrt{\frac{dr}{m}} \vee \frac{dr}{m}, \tag{26}
$$

moreover, for fixed $0<\delta<1$, we have the following tail bound

$$
\mathbb{P}\left(\left|\sup_{Z\in\epsilon\mathbb{S}_{2r}}\frac{4}{m}\sum_{i=1}^{m}\mathbb{1}\left(|\langle A_i, Z\rangle| \geq t\right) - \mathbb{E}\left[\sup_{Z\in\epsilon\mathbb{S}_{2r}}\frac{4}{m}\sum_{i=1}^{m}\mathbb{1}\left(|\langle A_i, Z\rangle| \geq t\right)\right]\right| > \delta\right) \leq 2e^{-cm\delta^2}. \tag{27}
$$

Proof The tail bound directly follows from Theorem 8.5 in [16]. Here we only give the proof sketch for the expectation bound. To apply Theorem 8.7 in [16], we only need to upper bound

$$\sigma^2 := \sup_{Z \in \epsilon \mathbb{S}_{2r}} \mathrm{Var}(\mathbb{1}(|\langle A, Z\rangle| > t)). \tag{28}$$

Note that

$$\begin{aligned} \sigma^2 &\leq \sup_{Z \in \epsilon \mathbb{S}_{2r}} \mathrm{Var}(\mathbb{1}(|\langle A, Z\rangle| > t)) \\ &\leq \sup_{Z \in \epsilon \mathbb{S}_{2r}} \mathbb{E}[\mathbb{1}(|\langle A, Z\rangle| > t)] \\ &\leq \sup_{W \in \mathbb{S}} \mathbb{P}(|\langle A, W\rangle| > t/\epsilon) \\ &\leq 2e^{-t^2/2\epsilon^2}, \end{aligned} \tag{29}$$

where in the last inequality, we used the tail bound for Gaussian random variables. Therefore, by Theorem 8.7 in [16], we have

$$\begin{aligned} \mathbb{E}\left[\sup_{Z \in \epsilon \mathbb{S}_{2r}} \frac{4}{m} \sum_{i=1}^m \mathbb{1}\left(|\langle A_i, Z\rangle| \geq t\right) - \mathbb{P}\left(|\langle A_i, Z\rangle| \geq t\right)\right] &\lesssim \sigma \sqrt{\frac{dr}{m} \log \frac{1}{\sigma}} \vee \frac{dr}{m} \log \frac{1}{\sigma} \\ &\lesssim e^{-t^2/4\epsilon^2} \sqrt{\frac{dr}{m}} \vee \frac{dr}{m}. \end{aligned} \tag{30}$$

This completes the proof. ■

Based on Lemma 13, we have

$$\sup_{\bar{X}, X} \text{(C1)} \lesssim e^{-t^2/4\epsilon^2} + 4t + \delta_4 + \delta_5 \tag{31}$$

with probability of at least $1 - C|\mathbb{S}_{\epsilon,r}|e^{-cm\delta_4^2} - Ce^{-cm\delta_5^2}$ given $m \gtrsim dr$. Combining all derived bounds, we have

$$\begin{aligned} &\sup_{X, Y \in \mathbb{S}_r} \left| \frac{1}{m} \sum_{i=1}^m \mathrm{Sign}(\langle A_i, X\rangle + s_i) \langle A_i, Y\rangle - \left(1 - p + p\mathbb{E}\left[e^{-s_i^2/2}\right]\right) \langle X, Y\rangle \right| \\ &\leq \delta_1 + C\epsilon(1 + \delta_2) + \sqrt{1 + \delta_3} \sqrt{e^{-t^2/4\epsilon^2} + 4t + \delta_4 + \delta_5} \end{aligned} \tag{32}$$

with probability of at least $1 - 2\left|\mathbb{S}_{\epsilon,r}\right|^2 e^{-cm\delta_1^2} - Ce^{-cm\delta_2^2} - Ce^{c_1 dr \log \frac{1}{\delta_3} - c_2 m\delta_3^2} - C\left|\mathbb{S}_{\epsilon,r}\right| e^{-cm\delta_4^2} - Ce^{-cm\delta_5^2}$. Upon choosing $\delta_2 = \delta_3 = \frac{1}{2}$, $\delta_4 = \delta_5 = \delta^2$, $\delta_1 = \delta$, $\epsilon = \delta^3$, $t = 8\delta^3\sqrt{\log \frac{1}{\delta}}$, and $m \gtrsim dr(\log \frac{1}{\delta} \vee 1)/\delta^4$, we have

$$\sup_{X, Y \in \mathbb{S}} \left| \frac{1}{m} \sum_{i=1}^m \mathrm{Sign}(\langle A_i, X\rangle + s_i) \langle A_i, Y\rangle - \sqrt{\frac{2}{\pi}} \left(1 - p + p\mathbb{E}\left[e^{-s_i^2/2}\right]\right) \langle X, Y\rangle \right| \lesssim \delta, \tag{33}$$

with probability of at least $1 - Ce^{-cm\delta^4}$. This leads to

$$\sup_{X \in \mathbb{S}, D \in \mathcal{M}(X)} \left\| D - \sqrt{\frac{2}{\pi}} \left(1 - p + p\mathbb{E}\left[e^{-s_i^2/2}\right]\right) X \right\|_F \lesssim \delta \tag{34}$$

Finally, note that

$$\sqrt{\frac{2}{\pi}}\left(1-p+p\mathbb{E}\left[e^{-s_i^2/2}\right]\right) \geq \sqrt{\frac{2}{\pi}}(1-p). \tag{35}$$

Therefore, we have

$$\sup_{X\in\mathbb{S}, D\in\mathcal{M}(X)} \left\|D-\sqrt{\frac{2}{\pi}}\left(1-p+p\mathbb{E}\left[e^{-s_i^2/2}\right]\right)X\right\|_F \lesssim \sqrt{\frac{2}{\pi}}\left(1-p+p\mathbb{E}\left[e^{-s_i^2/2}\right]\right)\delta, \tag{36}$$

with probability of at least $1-Ce^{-cm\delta^4}$, given $m \gtrsim \frac{dr\left(\log(\frac{1}{(1-p)\delta})\vee 1\right)}{(1-p)^4\delta^4}$. $\square$

Appendix C. Proof of Theorem 5

Define $\Delta = UU^\top - u^*u^{*\top}$. The conditions $\|u^*\| \leq 1$ and $\|U\| \leq R$ imply that $\|\Delta\|_F \leq (1+R^2)$. Since U is a critical point of (2), we have $0 \in \partial f_{\ell_1}(U)$, or equivalently, $QU = 0$ for some $Q \in \mathcal{Q}(\Delta)$, which in turn implies that

$$\left(Q - \varphi(\Delta)\frac{\Delta}{\|\Delta\|_F}\right)U = -\varphi(\Delta)\frac{\Delta U}{\|\Delta\|_F}.$$

Invoking Sign-RIP on the left side, we have

$$\|\Delta U\| \leq \delta\|\Delta\|\|U\| \leq R(1+R^2)\delta. \tag{37}$$

Consider the decomposition $U = \alpha u^* + \beta u^*_\perp$, where $\alpha = \langle u^*, U\rangle$, $u^*_\perp = \frac{U-\langle u^*,U\rangle u^*}{\|U-\langle u^*,U\rangle u^*\|}$, and $\beta = \|U - \langle u^*, U\rangle u^*\|$. Note that $\|u^*\| = \|u^*_\perp\| = 1$ and $\langle u^*, u^*_\perp\rangle = 0$. Substituting this decomposition in inequality (37) leads to

$$\begin{aligned}
\|\Delta U\|^2 &= \|(UU^\top - u^*u^{*\top})U\|^2 \\
&= \|\alpha(\alpha^2+\beta^2-1)u^* + \beta(\alpha^2+\beta^2)u^*_\perp\|^2 \\
&= \alpha^2(\alpha^2+\beta^2-1)^2 + \beta^2(\alpha^2+\beta^2)^2 \\
&\leq R^2(1+R^2)^2\delta^2
\end{aligned} \tag{38}$$

The above inequality implies that

$$\alpha(\alpha^2+\beta^2-1) \leq R(1+R^2)\delta \qquad \text{and} \qquad \beta(\alpha^2+\beta^2) \leq R(1+R^2)\delta$$

which in turn leads to

$$\begin{gathered}
\left\{\underbrace{|\alpha| \leq (R(1+R^2)\delta)^{1/3}}_{(A)} \quad \text{or} \quad \underbrace{|\alpha^2+\beta^2-1| \leq (R(1+R^2)\delta)^{2/3}}_{(B)}\right\} \\
\text{and} \\
\left\{\underbrace{|\beta| \leq (R(1+R^2)\delta)^{1/3}}_{(C)} \quad \text{or} \quad \underbrace{|\alpha^2+\beta^2| \leq (R(1+R^2)\delta)^{2/3}}_{(D)}\right\}
\end{gathered} \tag{39}$$

Now, we consider different cases

- *Case 1:* Suppose that (A) and (C) holds. This implies that

$$\alpha^2+\beta^2 \leq 2(R(1+R^2)\delta)^{2/3} \implies \|U\|^2 \leq 2(R(1+R^2)\delta)^{2/3}.$$

- *Case 2:* Suppose that (A) and (D) holds. Similar to the previous case, we have

$$\alpha^2+\beta^2 \leq (R(1+R^2)\delta)^{2/3} \implies \|U\|^2 \leq (R(1+R^2)\delta)^{2/3}.$$

- *Case 3:* Suppose that (B) and (D) holds. This implies that $1-(R(1+R^2)\delta)^{2/3} \leq \alpha^2+\beta^2$ and $\alpha^2+\beta^2 \leq (R(1+R^2)\delta)^{2/3}$. However, these inequalities are mutually exclusive if $\delta < \frac{1}{4\sqrt{2}R^3}$.

- *Case 4:* Suppose that (B) and (C) holds. This implies that

$$1-2(R(1+R^2)\delta)^{2/3} \leq \alpha^2 \leq 1+(R(1+R^2)\delta)^{2/3} \implies |\alpha^2-1| \leq 2(R(1+R^2)\delta)^{2/3}$$

  which leads to

$$\|UU^\top - u^*{u^*}^\top\|_F \leq \sqrt{5}(R(1+R^2)\delta)^{2/3}.$$

In summary, we have shown that for sufficiently small δ, the critical point $\|UU^\top - u^*{u^*}^\top\|_F \leq \sqrt{5}(R(1+R^2)\delta)^{2/3}$ or $\|U\|^2 \leq 2(R(1+R^2)\delta)^{2/3}$. Our next goal is to show that these bounds can be improved to $\|U\|^2 \leq 2\sqrt{2}\delta$ and $\|UU^\top - u^*{u^*}^\top\|_F \leq 2\sqrt{2}\delta$.

Suppose that U is close to 0, i.e., $\|U\|^2 \leq 2(R(1+R^2)\delta)^{2/3}$ (Cases 1 and 2). We use the following intermediate lemma:

Lemma 14 *Suppose that $\|U\| \leq C\delta^\epsilon$ for some $C \geq 1$ and $1 \geq \epsilon \geq 1/3$. Moreover, suppose that Sign-RIP holds with $\delta \leq \frac{1}{4C^{1/\epsilon}}$. Then, we have $\|U\| \leq \sqrt{2}(C)^{1/3}\delta^{(\epsilon+1)/3}$.*

Proof Since $\|U\| \leq C\delta^\epsilon$, we have $\|\Delta U\|^2 \leq 4C^2\delta^{2+2\epsilon}$, provided that $\delta \leq \frac{1}{(C)^{1/\epsilon}}$. A similar argument to (39) can be used to show that

$$\left\{\underbrace{|\alpha| \leq (2C)^{1/3}\delta^{(1+\epsilon)/3}}_{(A)} \quad \text{or} \quad \underbrace{|\alpha^2+\beta^2-1| \leq (2C)^{2/3}\delta^{2(1+\epsilon)/3}}_{(B)}\right\}$$
$$\text{and}$$
$$\left\{\underbrace{|\beta| \leq (2C)^{1/3}\delta^{(1+\epsilon)/3}}_{(C)} \quad \text{or} \quad \underbrace{|\alpha^2+\beta^2| \leq (2C)^{2/3}\delta^{2(1+\epsilon)/3}}_{(D)}\right\}$$

It is easy to verify that (B) is infeasible due to the upper bound on δ. Therefore, we have $\|U\|^2 = \alpha^2+\beta^2 \leq 2(2C)^{2/3}\delta^{2(1+\epsilon)/3}$, which completes the proof. ∎

Upon choosing $C_0 = \sqrt{2}(R(1+R^2))^{1/3}$ and $\epsilon_0 = 1/3$, the repeated applications of Lemma 14 implies that

$$\|U\| \leq C_k\delta^{\epsilon_k}, \quad \text{where} \quad C_k = \sqrt{2}C_{k-1}^{1/3},\ \epsilon_k = \frac{1+\epsilon_{k-1}}{3}$$

for every $k = 1, 2, \ldots, \infty$, provided that $\delta \leq \frac{1}{4C_k^{1/\epsilon_k}}$. It is easy to verify that $C_\infty = 2^{3/4}$ and $\epsilon_\infty = 1/2$. Moreover, the choice of $\delta \leq \frac{1}{32R(1+R^2)}$ is enough to guarantee $\delta \leq \frac{1}{4C_k^{1/\epsilon_k}}$. This in turn implies that $\|U\|^2 \leq 2\sqrt{2}\delta$. A similar approach can be used to show that $\|UU^\top - u^* {u^*}^\top\|_F \leq 2\sqrt{2}\delta$. $\square$

Based on the proof of Theorem 5, we prove Corollary 6.

Proof of Corollary 6. The first part of the corollary is directly implied by Theorem 4 combined with Theorem 5. To prove the second part, note that, due to Lemma 21, Sign-RIP implies ℓ_1/ℓ_2-RIP. On the other hand, Li et al. [9] show that, with the choice of $p \leq \frac{1}{2} - \frac{\delta}{\sqrt{2/\pi - \delta}}$, the loss function is α-weakly-sharp [9, Proposition 2] and τ-weakly-convex [9, Proposition 3], with appropriate choices of α and τ (for simplicity of presentation, we omit the exact definition of these parameters, and refer the reader to [9]). This implies that the $f_{\ell_1}(U)$ does not have a critical point other than $U = u^*$ within a neighborhood of u^* with radius $\frac{2\alpha}{\tau}$. On the other hand, Theorem 5 implies that, with the provided bounds on δ and m, all the critical points of that are not close to the origin must be within a neighborhood of u^* with radius $\frac{2\alpha}{\tau}$. This implies that these critical points must coincide with true solution, i.e., $UU^\top = u^* {u^*}^\top$. $\square$

Appendix D. Proofs for Convergence Analysis

Before diving into details, we first provide a sketch of the proof for Theorem 7.

Sketch of the proof for Theorem 7. Suppose that $X^* = u^* {u^*}^\top$ for $u^* \in \mathbb{R}^{d\times 1}$. Moreover, without loss of generality, we assume that $\|u^*\| = 1$. Inspired by [10], we decompose the solution U_t as

$$U_t = u^* {u^*}^\top U_t + \left(1 - u^* {u^*}^\top\right) U_t := u^* r_t^\top + E_t, \tag{40}$$

where $r_t = U_t^\top u^*$ is called *signal term*, and $E_t = (1 - u^* {u^*}^\top) U_t$ is referred to as *error term*, which is the projection of U_t onto the orthogonal complement of the subspace spanned by u^*. Evidently, we have $U_t U_t^\top = X^\star$ if and only if $\|r_t\| = 1$ and $\|E_t\|_F = 0$. More generally, our next lemma shows that the error $\|U_t U_t - X^*\|_F$ can be controlled in terms of $\|E_t\|_F$ and $\|r_t\|$.

Lemma 15 *The following inequality holds:*

$$\left\|U_t U_t^\top - X^\star\right\|_F^2 \leq \left(1 - \|r_t\|^2\right)^2 + 2\,\|E_t\|^2 \|r_t\|^2 + \|E_t\|_F^4. \tag{41}$$

Based on Lemma 15, we provide a high-level idea of our proof technique:

1. (Spectral Initialization) It is shown in Lemma 16 that the proposed initialization scheme (see Algorithm 2) results in $\|r_0\| = \alpha(1 \pm O(\sqrt{\delta}))$ and $\|E_0\| = O(\alpha\sqrt{\delta})$. Therefore, the signal term dominates the error term at the beginning.

2. It is shown in Lemma 17 that the signal term $\|r_t\|^2$ approaches 1 at a geometric rate. Therefore, $1 - \|r_t\|^2$ converges to zero at a geometric rate.

3. It is proven in Lemma 18 that the error term $\|E_t\|_F$ grows at most sublinearly, and its growth rate is significantly slower than that of the signal term.

4. This discrepancy in the growth rates of the signal and error terms ensures that after a certain number of iterations T, the signal term $\|r_t\|$ is sufficiently close to 1, while the error term $\|E_t\|_F$ remains small. Combined with Lemma 15, this establishes the convergence of SubGD with early stopping of the algorithm.

In particular, our main proof is based on the following three key lemmas, the proofs of which can be found in Appendix D of the supplementary material.

Lemma 16 (Spectral Initialization) *Suppose that $U_0 = B_0$ is chosen by Algorithm 2. Then under the conditions of Theorem 7 , we have*

$$\|r_0\| = \alpha\sqrt{\varphi_0}(1 \pm O(\sqrt{\delta})), \quad \|E_0\| = O(\alpha\sqrt{\varphi_0}\sqrt{\delta}), \quad \|E_0\|_F = O(\alpha\sqrt{\varphi_0}\sqrt[4]{r'}\sqrt{\delta}), \tag{42}$$

where $\varphi_0 = \varphi(U_0U_0^\top/\alpha^2 - X^\star) \in [\sqrt{2/\pi}(1-p), \sqrt{2/\pi}]$ is the initial scaling factor.

Given with this lemma, we next show that the signal term grows much faster than the error term.

Lemma 17 (Signal Dynamics) *Assume that the measurements satisfy the sign-RIP with parameters $(\min\{r'+1, d\}, \delta)$ and a strictly positive and uniformly bounded scaling function $\varphi(X)$. Moreover, suppose that $\|E_t\|_F \leq 1, \|r_t\| \leq 2, \delta \leq \frac{1}{2}$, and the step size η_t is chosen as (**??**). Then, we have*

$$\begin{aligned}\left\|r_{t+1} - \left(1 + \frac{\eta_0\rho^t(1-\|r_t\|^2)}{\left\|U_tU_t^\top - X^\star\right\|_F}\right) r_t\right\| &\leq 2\delta\eta_0\rho^t(\|E_t\| + \|r_t\|) + \frac{2\eta_0\rho^t}{\left\|U_tU_t^\top - X^\star\right\|_F}\|E_t\|^2\|r_t\| \\ &\quad + \frac{2\delta\eta_0\rho^t}{\left\|U_tU_t^\top - X^\star\right\|_F}(1-\|r_t\|^2)\|r_t\|.\end{aligned} \tag{43}$$

The above lemma shows that, when $\|r_t\|$ and t are small, the signal term grows geometrically fast. On the other hand, the growth rate of the error is sublinear, as shown in the following lemma.

Lemma 18 (Error Dynamics) *Suppose that the conditions of Proposition 17 are satisfied and $\eta_0 \lesssim \delta \lesssim \left\|U_tU_t^\top - X^*\right\|_F$. Then, we have*

$$\|E_{t+1}\|_F \leq \|E_t\|_F + 10\delta\eta_0\rho^t. \tag{44}$$

Combining the aforementioned lemmas, we establish the global convergence of SubGD for robust rank-1 matrix recovery. □

First, we consider the noiseless scenario with clean measurements. Our proof for the noisy case is built upon the developed results for the the noiseless scenario. In particular, different from the noisy setting, we choose the step-size $\eta_t = \frac{\pi}{2m}\left\|\mathrm{y} - \mathcal{A}(U_tU_t^\top)\right\|_1$. As will be shown later, this choice of η_t will remain close to $\|UU^\top - u^*u^{*\top}\|_F$ due to sign-RIP.

Proofs for the Noiseless Case

D.1. Proof of Lemma 15

Due to (40), one can write

$$\begin{aligned}&\left\|U_tU_t^\top - X^\star\right\|_F^2 \\ =&\left\|(u^\star r_t^\top + E_t)(r_tu^{\star\top} + E_t^\top) - u^\star u^{\star\top}\right\|_F^2 \\ =&\left\|\underbrace{(\|r_t\|^2-1)u^\star u^{\star\top}}_{\text{(A)}} + \underbrace{E_tE_t^\top}_{\text{(B)}} + \underbrace{E_tr_tu^{\star\top}}_{\text{(C)}} + \underbrace{u^\star r_t^\top E_t^\top}_{\text{(D)}}\right\|_F^2.\end{aligned} \tag{45}$$

Now, note that $\|\mathrm{A}\|_F^2 = \left(1-\|r_t\|^2\right)^2$, and

$$\|\mathrm{C}\|_F = \|\mathrm{D}\|_F = \|E_t r_t\| \, \|u^\star\| = \|E_t r_t\| \, . \tag{46}$$

On the other hand, $u^{\star\top} E_t = u^{\star\top}(I - u^\star u^{\star\top})U_t = 0$, and therefore, $\langle \mathrm{A}, \mathrm{B}\rangle = 0$. Similarly, $\langle \mathrm{A}, \mathrm{C}\rangle = (\|r_t\|^2 - 1)\mathrm{Tr}\left(u^\star u^{\star\top} E_t r_t u^{\star\top}\right) = 0$, and $\langle \mathrm{A}, \mathrm{D}\rangle = 0$ since $\mathrm{Tr}\left(u^\star u^{\star\top} u^\star r_t^\top E_t^\top\right) = \mathrm{Tr}\left(u^\star u^{\star\top} U_t U_t^\top (I - u^\star u^{\star\top})\right) = \mathrm{Tr}\left(U_t U_t^\top (I - u^\star u^{\star\top}) u^\star u^{\star\top}\right) = 0$. Similarly, we have $\langle \mathrm{B}, \mathrm{C}\rangle = 0, \langle \mathrm{B}, \mathrm{D}\rangle = 0, \langle \mathrm{C}, \mathrm{D}\rangle = 0$. This leads to

$$\begin{aligned}\left\|U_t U_t^\top - X^\star\right\|_F^2 &= \left(1-\|r_t\|^2\right)^2 + 2\left\|E_t r_t\right\|^2 + \left\|E_t E_t^\top\right\|_F^2 \\ &\le \left(1-\|r_t\|^2\right)^2 + 2\left\|E_t\right\|^2 \left\|r_t\right\|^2 + \left\|E_t\right\|_F^4 .\end{aligned} \tag{47}$$

□

D.2. Error Dynamics

Proposition 19 (Error Dynamics) *Assume that the measurements are noiseless and satisfy the sign-RIP with parameters $(\min\{r'+1, d\}, \delta)$, and constant scaling function $\varphi(X) = \sqrt{\frac{2}{\pi}}$. Moreover, suppose that $\|E_t\|_F \le 1, \|r_t\| \le 2, \delta \le \frac{1}{2}$, and the step size η_t is chosen as (3) with $\eta_0 \le \frac{2}{45}$. Then, we have*

$$\|E_{t+1}\|_F \le \|E_t\|_F + 22\delta\eta_0. \tag{48}$$
$$\|E_{t+1}\| \le \|E_t\| + 15\delta\eta_0. \tag{49}$$

Proof

For simplicity of notation, we define $\Delta_t = U_t U_t^\top - X^*$ throughout the proof. First, we provide a useful fact, which will be widely used in our subsequent arguments.

Fact 1 *Suppose $\|E_t\|_F \le 1, \|r_t\| \le 2$, then by Lemma 15, we have $\|\Delta_t\|_F^2 \le 1 + 2\times 4 + 1 = 10$.*

We first prove the error dynamics under a general learning rate η_t.

Lemma 20 *Suppose that $\|E_t\|_F \le 1, \|r_t\| \le 2$, $\delta \le \frac{1}{2}$, then, the following inequalities hold*

$$\|E_{t+1}\|_F^2 \le \|E_t\|_F^2 + \sqrt{\frac{8}{\pi}}\eta_t\left(-\frac{\|E_t U_t^\top\|_F^2}{\|\Delta_t\|_F} + \delta\left\|E_t U_t^\top\right\|_F\right) + \frac{20}{\pi}\eta_t^2\left(\delta^2 + \left\|E_t U_t^\top\right\|_F^2 / \|\Delta_t\|_F^2\right), \tag{50}$$

$$\|E_{t+1}\| \le \left\|I - \frac{\eta_t U_t^\top U_t}{\sqrt{\frac{\pi}{2}}\|\Delta_t\|_F}\right\| \cdot \|E_t\| + \sqrt{\frac{2}{\pi}}\delta\eta_t(\|r_t\| + \|E_t\|). \tag{51}$$

Proof For simplicity, we denote $M_t \in \frac{1}{m}\sum_{i=1}^m \mathrm{Sign}(\langle A_i, U_t U_t^\top - X^\star\rangle)A_i$, and $\bar{M}_t = \sqrt{\frac{2}{\pi}}\frac{\Delta_t}{\|\Delta_t\|_F}$. It is easy to verify that

$$E_{t+1} = E_t - \eta_t(I - u^\star u^{\star\top})M_t U_t, \tag{52}$$

Based on the above equation, one can write

$$\|E_{t+1}\|_F^2 = \|E_t\|_F^2 - 2\eta_t\left\langle E_t, (I - u^\star u^{\star\top})M_t U_t\right\rangle + \eta_t^2\left\|(I - u^\star u^{\star\top})M_t U_t\right\|_F^2 . \tag{53}$$

Next, we will provide separate upper bounds for the second and third terms in the above equation. First, note that

$$\begin{aligned}\langle E_t, M_t U_t\rangle &= \left\langle M_t, E_t U_t^\top \right\rangle \\ &\overset{(a)}{\geq} \frac{1}{\sqrt{\frac{\pi}{2}}\left\|\Delta_t\right\|_F}\left\langle \Delta_t, E_t U_t^\top \right\rangle - \sqrt{\frac{2}{\pi}}\delta\left\|E_t U_t^\top\right\|_F \\ &\overset{(b)}{=} \sqrt{\frac{2}{\pi}}\frac{\left\|E_t U_t^\top\right\|_F^2}{\left\|\Delta_t\right\|_F} - \sqrt{\frac{2}{\pi}}\delta\left\|E_t U_t^\top\right\|_F.\end{aligned} \tag{54}$$

where we used sign-RIP condition in (a), and (b) follows from $\left\langle U_t U_t^\top - X^\star, E_t U_t^\top\right\rangle = \left\langle E_t U_t^\top, E_t U_t^\top\right\rangle = \left\|E_t U_t^\top\right\|_F^2$. On the other hand, we have

$$\left\langle E_t, u^\star u^{\star\top} M_t U_t\right\rangle = \mathrm{Tr}\left(E_t^\top u^\star u^{\star\top} M_t U_t\right) = \mathrm{Tr}\left(U_t^\top (I - u^\star u^{\star\top}) u^\star u^{\star\top} M_t U_t\right) = 0. \tag{55}$$

Combining (54) and (55) leads to

$$-2\eta_t\left\langle E_t, (I - u^\star u^{\star\top}) M_t U_t\right\rangle \leq -2\eta_t\left(\sqrt{\frac{2}{\pi}}\frac{\left\|E_t U_t^\top\right\|_F^2}{\left\|\Delta_t\right\|_F} - \sqrt{\frac{2}{\pi}}\delta\left\|E_t U_t^\top\right\|_F\right) \tag{56}$$

Now, we provide an upper bound for the third term in (53). One can write

$$\begin{aligned}\left\|(I - u^\star u^{\star\top}) M_t U_t\right\|_F^2 &\leq 2\left\|(I - u^\star u^{\star\top})(M_t - \bar{M}_t) U_t\right\|_F^2 + 2\left\|(I - u^\star u^{\star\top})\bar{M}_t U_t\right\|_F^2 \\ &\overset{(a)}{\leq} 2\left\|(M_t - \bar{M}_t) U_t\right\|_F^2 + \frac{4}{\pi}\frac{\left\|E_t U_t^\top U_t\right\|_F^2}{\left\|\Delta_t\right\|_F^2} \\ &\overset{(b)}{\leq} \frac{4}{\pi}\delta^2\left\|U_t\right\|_F^2 + \frac{4}{\pi}\left\|E_t U_t^\top\right\|_F^2\left\|U_t\right\|_F^2 / \left\|\Delta_t\right\|_F^2 \\ &\overset{(c)}{\leq} \frac{20}{\pi}\delta^2 + \frac{20}{\pi}\left\|E_t U_t^\top\right\|_F^2 / \left\|\Delta_t\right\|_F^2.\end{aligned} \tag{57}$$

where we used the contraction of projection and $(I - u^\star u^{\star\top})(U_t U_t^\top - X^\star) U_t = E_t U_t^\top U_t$ in (a). Moreover, (b) directly follows from the sign-RIP condition. Finally, we used the following fact in (c).

Fact 2 *Assuming* $\|E_t\|_F \leq 1, \|r_t\| \leq 2$, *we have* $\|U_t\|_F^2 = \|r_t\|^2 + \|E_t\|_F^2 \leq 1^2 + 2^2 = 5$.

Finally, combining all the three terms, we finally have:

$$\|E_{t+1}\|_F^2 \leq \|E_t\|_F^2 + \sqrt{\frac{8}{\pi}}\eta_t\left(-\frac{\left\|E_t U_t^\top\right\|_F^2}{\left\|\Delta_t\right\|_F} + \delta\left\|E_t U_t^\top\right\|_F\right) + \frac{20}{\pi}\eta_t^2\left(\delta^2 + \frac{\left\|E_t U_t^\top\right\|_F^2}{\left\|\Delta_t\right\|_F^2}\right). \tag{58}$$

Now, we turn to control the spectral norm. First, notice that

$$
\begin{aligned}
&\left\|(I - u^\star u^{\star\top})M_t U_t - (I - u^\star u^{\star\top})\frac{\Delta_t U_t}{\sqrt{\frac{\pi}{2}}\left\|\Delta_t\right\|_F}\right\| \\
\overset{(a)}{\leq} &\left\|M_t U_t - \frac{\Delta_t U_t}{\sqrt{\frac{\pi}{2}}\left\|\Delta_t\right\|_F}\right\| \\
\leq &\left\|M_t - \bar{M}_t\right\| \cdot \left\|U_t\right\| \\
\leq &\sqrt{\frac{2}{\pi}}\delta\left\|U_t\right\| \\
\leq &\sqrt{\frac{2}{\pi}}\delta(\left\|E_t\right\| + \left\|r_t\right\|).
\end{aligned} \tag{59}
$$

Here in (a) we used the contraction of projection. On the other hand, observing that $(I - u^\star u^{\star\top})(U_t U_t^\top - X^\star)U_t = E_t U_t^\top U_t$, we have

$$
\begin{aligned}
\left\|E_{t+1}\right\| &= \left\|E_t - \eta_t(I - u^\star u^{\star\top})M_t U_t\right\| \\
&\leq \left\|E_t - \eta_t(I - u^\star u^{\star\top})\bar{M}_t U_t\right\| + \eta_t\left\|(I - u^\star u^{\star\top})(M_t - \bar{M}_t)U_t\right\| \\
&\leq \left\|E_t\left(I - \frac{\eta_t U_t^\top U_t}{\sqrt{\frac{\pi}{2}}\left\|\Delta_t\right\|_F}\right)\right\| + \sqrt{\frac{2}{\pi}}\delta\eta_t(\left\|r_t\right\| + \left\|E_t\right\|) \\
&\leq \left\|I - \frac{\eta_t U_t^\top U_t}{\sqrt{\frac{\pi}{2}}\left\|\Delta_t\right\|_F}\right\| \cdot \left\|E_t\right\| + \sqrt{\frac{2}{\pi}}\delta\eta_t(\left\|r_t\right\| + \left\|E_t\right\|),
\end{aligned} \tag{60}
$$

which completes the proof. ∎

Before presenting the proof of Proposition 19, we need the following intermediate result

Lemma 21 *Suppose that the measurements satisfy sign-RIP with parameters (δ, r) and a constant scaling function $\varphi(X) = \sqrt{\frac{2}{\pi}}$. Then, for every $X \in \mathbb{S}_r$, we have*

$$
\left|\frac{1}{m}\left\|\mathcal{A}(X)\right\|_1 - \sqrt{\frac{2}{\pi}}\left\|X\right\|_F\right| \leq \sqrt{\frac{2}{\pi}}\delta \tag{61}
$$

Proof Due to the sign-RIP condition, we have $\left\|D - \sqrt{\frac{2}{\pi}}X\right\|_F \leq \sqrt{\frac{2}{\pi}}\delta$ for every $X \in \mathbb{S}_r$ and $D \in \frac{1}{m}\sum_{i=1}^m \mathrm{Sign}(\langle A_i, X\rangle)A_i$. This implies that

$$
\sqrt{\frac{2}{\pi}}\delta = \sup_{Y \in \mathbb{S}}\left\langle D - \sqrt{\frac{2}{\pi}}X, Y\right\rangle \geq \left\langle D - \sqrt{\frac{2}{\pi}}X, X\right\rangle \geq \frac{1}{m}\left\|\mathcal{A}(X)\right\|_1 - \sqrt{\frac{2}{\pi}}\left\|X\right\|_F. \tag{62}
$$

Similarly, it can be shown that $\sqrt{\frac{2}{\pi}}\delta \geq -\frac{1}{m}\left\|\mathcal{A}(X)\right\|_1 + \sqrt{\frac{2}{\pi}}\left\|X\right\|_F$. This completes the proof. ∎

Based on the above lemma, the sign-RIP condition implies the ℓ_1/ℓ_2-RIP condition. Given Lemmas 20 and 21, we are ready to present the proof of Proposition 19.

Proof [Proof of Proposition 19] It is enough to show that the choice of $\eta_t = \frac{\pi}{2}\eta_0 \cdot \frac{1}{m}\sum \left|\left\langle A_i, U_tU_t^\top - X^\star\right\rangle\right|$ in Lemma 20 leads to the desired bounds. Based on Lemma 21 and $\delta \le \frac{1}{2}$, we have

$$\eta_t \le \sqrt{\frac{\pi}{2}}\eta_0 \left\|\Delta_t\right\|_F (1+\delta) \le \sqrt{\frac{9\pi}{8}}\eta_0 \left\|\Delta_t\right\|_F. \tag{63}$$

Similarly, we have the lower bound $\eta_t \ge \sqrt{\frac{\pi}{8}}\eta_0 \left\|\Delta_t\right\|_F$. Substituting these inequalities in Lemma 20 leads to

$$\begin{aligned}
\left\|E_{t+1}\right\|_F^2 &\le \left\|E_t\right\|_F^2 - \eta_0 \left\|E_tU_t^\top\right\|_F^2 + 3\delta\eta_0 \left\|\Delta_t\right\|_F \left\|E_tU_t^\top\right\|_F + \frac{45}{2}\eta_0^2\left(\delta^2 \left\|\Delta_t\right\|_F^2 + \left\|E_tU_t^\top\right\|_F^2\right)\\
&\overset{(a)}{\le} \left\|E_t\right\|_F^2 + 3\delta\eta_0 \left\|\Delta_t\right\|_F \left\|E_tU_t^\top\right\|_F + \frac{45}{2}\delta^2\eta_0^2 \left\|\Delta_t\right\|_F^2\\
&\overset{(b)}{\le} \left\|E_t\right\|_F^2 + 3\sqrt{10}\delta\eta_0 \left\|E_t\right\|_F \left\|U_t\right\|_F + 225\delta^2\eta_0^2\\
&\overset{(c)}{\le} \left\|E_t\right\|_F^2 + 22\delta\eta_0 \left\|E_t\right\|_F + 225\delta^2\eta_0^2\\
&= \left(\left\|E_t\right\|_F + 11\delta\eta_0\right)^2 + 104\delta^2\eta_0^2.
\end{aligned} \tag{64}$$

where (a) follows from the assumption $\eta_0 \le \frac{2}{45}$, (b) follows form Fact 1, and (c) follows from Fact 2. Therefore, we have

$$\left\|E_{t+1}\right\|_F \le \left\|E_t\right\|_F + 11\delta\eta_0 + 11\delta\eta_0 = \left\|E_t\right\|_F + 22\delta\eta_0. \tag{65}$$

Similarly,

$$\left\|E_{t+1}\right\| \le \left\|I - \frac{\eta_tU_t^\top U_t}{\sqrt{\frac{\pi}{2}}\left\|\Delta_t\right\|_F}\right\| \cdot \left\|E_t\right\| + \sqrt{\frac{2}{\pi}}\delta\eta_t(\left\|r_t\right\| + \left\|E_t\right\|). \tag{66}$$

Note that

$$\left\|U_t^\top U_t\right\| \le \left\|U_t\right\|^2 \le (\left\|E_t\right\| + \left\|u^\star r_t^\top\right\|)^2 \le (\left\|E_t\right\| + \left\|r_t\right\|)^2 \le 9. \tag{67}$$

which, together with $\eta_t \le \sqrt{\frac{9\pi}{8}}\eta_0$, implies

$$\left\|\frac{\eta_tU_t^\top U_t}{\sqrt{\frac{\pi}{2}}\left\|\Delta_t\right\|_F}\right\| < 1. \tag{68}$$

Therefore, the first term in the right hand side of the above inequality is upper bounded by $\left\|E_t\right\|$. On the other hand, the second term in (66) can be bounded as

$$\sqrt{\frac{2}{\pi}}\delta\eta_t(\left\|r_t\right\| + \left\|E_t\right\|) \le \frac{3}{2}\delta\eta_0\|\Delta_t\|_F(\left\|r_t\right\| + \left\|E_t\right\|) \le 15\delta\eta_0 \tag{69}$$

This completes the proof. ∎

∎

D.3. Signal Dynamics

Proposition 22 (Signal Dynamics) *Under the conditions of Proposition 19, we have*

$$\left\|r_{t+1} - \left(1 + \eta_0(1 - \|r_t\|^2)\right) r_t\right\| \leq 10\eta_0\delta(\|E_t\| + \|r_t\|) + 2\eta_0 \|E_t\|^2 \|r_t\| . \tag{70}$$

Proof Similarly, we first prove the following lemma which holds for a general choice of η_t.

Lemma 23 *Assuming $\|E_t\|_F \leq 1, \|r_t\| \leq 2$, we have*

$$\begin{aligned}\left\|r_{t+1} - \left(1 + \frac{\eta_t(1 - \|r_t\|^2)}{\sqrt{\frac{\pi}{2}} \left\|U_tU_t^\top - X^\star\right\|_F}\right) r_t\right\| &\leq \sqrt{\frac{2}{\pi}}\eta_t\delta(\|E_t\| + \|r_t\|) \\ &\quad + \frac{\eta_t}{\sqrt{\frac{\pi}{2}} \left\|U_tU_t^\top - X^\star\right\|_F} \|E_t\|^2 \|r_t\| .\end{aligned} \tag{71}$$

Proof Recalling the notations $M_t \in \frac{1}{m}\sum_{i=1}^m \mathrm{Sign}(\langle A_i, \Delta_t\rangle)A_i$ and $\bar{M}_t = \sqrt{\frac{2}{\pi}}\frac{\Delta_t}{\|\Delta_t\|_F}$, we have

$$r_{t+1} = r_t - \eta_t U_t^\top M_t^\top u^\star. \tag{72}$$

Therefore,

$$\begin{aligned}\left\|r_{t+1} - r_t + \eta_t\frac{U_t^\top \Delta_t u^\star}{\sqrt{\frac{\pi}{2}} \|\Delta_t\|_F}\right\| &\leq \eta_t \left\|U_t^\top (M_t - \bar{M}_t)^\top u^\star\right\| \\ &\leq \eta_t \|U_t\| \cdot \left\|M_t - \bar{M}_t\right\| \cdot \|u^\star\| \\ &\leq \sqrt{\frac{2}{\pi}}\eta_t\delta(\|E_t\| + \|r_t\|),\end{aligned} \tag{73}$$

where the last inequality follows from the sign-RIP condition. On the other hand, since $U_t^\top \left(U_tU_t^\top - X^\star\right) u^\star = U_t^\top U_t r_t - r_t = \left(r_tr_t^\top + E_t^\top E_t\right) r_t - r_t = \left(\|r_t\|^2 - 1\right) r_t - E_t^\top E_t r_t$, one can write

$$\begin{aligned}&\left\|r_{t+1} - \left(1 + \frac{\eta_t(1 - \|r_t\|^2)}{\sqrt{\frac{\pi}{2}} \|\Delta_t\|_F}\right) r_t\right\| \\ \leq&\sqrt{\frac{2}{\pi}}\eta_t\delta(\|E_t\| + \|r_t\|) + \frac{\eta_t}{\sqrt{\frac{\pi}{2}} \|\Delta_t\|_F} \left\|E_t^\top E_t r_t\right\| \\ \leq&\sqrt{\frac{2}{\pi}}\eta_t\delta(\|E_t\| + \|r_t\|) + \frac{\eta_t}{\sqrt{\frac{\pi}{2}} \|\Delta_t\|_F} \|E_t\|^2 \|r_t\| .\end{aligned} \tag{74}$$

■

Equipped with this lemma and (63), we write

$$\begin{aligned}\left\|r_{t+1} - \left(1 + \eta_0(1 - \|r_t\|^2)\right) r_t\right\| &\leq \eta_0\delta(1 - \|r_t\|^2) \|r_t\| \\ &\quad + 2\eta_0 \|\Delta_t\|_F \delta(\|E_t\| + \|r_t\|) \\ &\quad + 2\eta_0 \|E_t\|^2 \|r_t\| .\end{aligned} \tag{75}$$

Note that the first term is dominated by the second term since $\|\Delta_t\|_F \geq 1 - \|r_t\|^2$ due to (47) and the fact that $\|\Delta_t\|_F \leq \sqrt{10}$. We finally have

$$\left\| r_{t+1} - \left(1 + \eta_0(1 - \|r_t\|^2)\right) r_t \right\| \leq 10\eta_0\delta(\|E_t\| + \|r_t\|) + 2\eta_0 \|E_t\|^2 \|r_t\| . \tag{76}$$

■

D.4. Convergence Result

Now we can formally state the following convergence theorem for the noiseless setting.

Theorem 24 *Assume that the measurements are noiseless and satisfy the sign-RIP condition with parameters* $(\min\{r'+1, d\}, \delta)$, $\delta \lesssim 1$, *and constant scaling function* $\varphi(X) = \sqrt{\frac{2}{\pi}}$. *Suppose that* $\alpha \asymp \sqrt{\delta}/\sqrt[4]{r'}$ *and the step size* η_t *is chosen as* (3) *with* $\eta_0 \lesssim 1$. *Then, after* $T \asymp \frac{\log(r'/\delta)}{\eta_0}$ *iterations, we have*

$$\left\| U_T U_T^\top - X^\star \right\|_F^2 \lesssim \delta^2 \log^2\left(\frac{r'}{\delta}\right). \tag{77}$$

Proof To start the proof, we first provide the following intermediate lemma on the initial signal and error terms.

Lemma 25 *Suppose that* $U_0 = \alpha B$ *is chosen by Algorithm* 2. *Then under the conditions of Theorem* 24 *, we have*

$$\|r_0\| = \alpha(1 \pm O(\sqrt{\delta})), \quad \|E_0\| = O(\alpha\sqrt{\delta}), \quad \|E_0\|_F = O(\alpha\sqrt[4]{r'}\sqrt{\delta}). \tag{78}$$

Proof For convenience, we list the diagonal elements of Σ in a descending order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d$. Moreover, suppose that $u_1, \cdots, u_d$ are the corresponding eigenvectors. For simplicity, we define $\sigma_i' = \max\{\sigma_i, 0\}$. Due to sign-RIP, we have

$$\left\| C - \sqrt{\frac{2}{\pi}} X^\star \right\|_F \lesssim \delta. \tag{79}$$

Therefore, we have

$$\begin{aligned}\left\| \hat{X} - X^\star \right\|_F &= \left\| \frac{C}{\|C\|_F} - X^\star \right\|_F \\ &\leq \left\| \sqrt{\frac{\pi}{2}} C - X^\star \right\|_F + \|X^\star\|_F - \sqrt{\frac{\pi}{2}} \|C\|_F \\ &= O(\delta),\end{aligned} \tag{80}$$

here we use triangle inequality twice. Since $\mathrm{span}\{u_1, \cdots, u_d\} = \mathbb{R}^d$ and $\|u^\star\| = 1$, we can write $u^\star = \sum_{i=1}^d \beta_i u_i$, where $\sum_{i=1}^d \beta_i^2 = 1$. Therefore, we have

$$\begin{aligned}
\left\|\hat{X} - X^\star\right\|_F^2 &= \left\langle \sum_{i=1}^d \sigma_i u_i u_i^\top - u^\star u^{\star\top}, \sum_{i=1}^d \sigma_i u_i u_i^\top - u^\star u^{\star\top} \right\rangle \\
&= \sum_{i=1}^d \sigma^2 - 2\sum_{i=1}^d \sigma_i \beta_i^2 + 1 \\
&= \sum_{i=1}^d \left(\sigma_i - \beta_i^2\right)^2 + 1 - \sum_{i=1}^d \beta_i^4 \lesssim \delta^2,
\end{aligned} \tag{81}$$

which implies

$$\sum_{i=1}^d \left(\sigma_i - \beta_i^2\right)^2 = O\left(\delta^2\right), \quad \sum_{i=1}^d \beta_i^4 = 1 - O\left(\delta^2\right). \tag{82}$$

Without loss of generality, we define $\beta_{\max} = |\beta_1| = \max\{|\beta_i| : 1 \le i \le d\}$. Therefore, we have

$$1 - O\left(\delta^2\right) = \sum_{i=1}^d \beta_i^4 \le \beta_{\max}^2 \sum_{i=1}^d \beta_i^2 = \beta_{\max}^2. \tag{83}$$

Here we use the fact that $\sum_{i=1}^d \beta_i^2 = 1$. Moreover, it is easy to see that $\sum_{i=2}^d \beta_i^2 = 1 - \beta_{\max}^2 = O\left(\delta^2\right)$. Therefore, we have

$$\begin{aligned}
\left\|BB^\top - X^\star\right\|_F^2 &= \sum_{i=1}^{r'} \sigma_i'^2 - \sum_{i=1}^{r'} \sigma_i' \beta_i^2 + 1 \\
&= \sum_{i=1}^{r'} \left(\sigma_i' - \beta_i^2\right)^2 - \sum_{i=1}^{r'} \beta_i^4 + 1 \\
&= 1 - \beta_{\max}^4 + O\left(\delta^2\right) \\
&= \left(1 + \beta_{\max}^2\right)\left(1 - \beta_{\max}^2\right) + O\left(\delta^2\right) = O\left(\delta^2\right).
\end{aligned} \tag{84}$$

By Lemma 15, we immediately have

$$\left(1 - \left\|B^\top u^\star\right\|^2\right)^2 \le \left\|BB^\top - X^\star\right\|_F^2 \lesssim \delta^2. \tag{85}$$

On the other hand, note that $\|r_0\| = \alpha \left\|B^\top u^\star\right\|$, which together with the above inequality, implies $\|r_0\| = \alpha(1 \pm O(\sqrt{\delta}))$. Similarly, we have

$$\begin{aligned}
\left\|(I - u^\star u^{\star\top})B\right\|^2 &= \sup_{\|x\| \le 1} x^\top (I - u^\star u^{\star\top}) BB^\top (I - u^\star u^{\star\top}) x \\
&= \sup_{\|x\| \le 1} x^\top (I - u^\star u^{\star\top}) (BB^\top - X^\star) (I - u^\star u^{\star\top}) x \\
&\le \left\|BB^\top - X^\star\right\| \le \left\|BB^\top - X^\star\right\|_F = O(\delta),
\end{aligned} \tag{86}$$

which leads to $\|E_0\| = O(\alpha\sqrt{\delta})$. As for the Frobenius norm, we have

$$\begin{aligned}
\left\|(I - u^\star u^{\star\top})B\right\|_F^2 &= \left\langle (I - u^\star u^{\star\top})B, (I - u^\star u^{\star\top})B \right\rangle \\
&= \left\langle I - u^\star u^{\star\top}, BB^\top - X^\star \right\rangle \\
&= \sum_{i=1}^{r'} \sigma_i' - \sum_{i=1}^{r'} \sigma_i' \beta_i^2 \\
&\leq \sigma_1' \left(1 - \beta_{\max}^2\right) + \sum_{i=2}^{r'} \sigma_i' \\
&= \sum_{i=2}^{r'} \left(\sigma_i' - \beta_i^2\right) + \sum_{i=2}^{r'} \beta_i^2 + O\left(\delta^2\right) \\
&\leq \sqrt{r'-1}\sqrt{\sum_{i=2}^{r'} \left(\sigma_i' - \beta_i^2\right)^2} + 1 - \beta_{\max}^2 + O\left(\delta^2\right) \\
&= O\left(\sqrt{r'}\delta\right).
\end{aligned} \tag{87}$$

Therefore, we have $\|E_0\|_F = O\left(\alpha\sqrt[4]{r'}\sqrt{\delta}\right)$, which completes the full proof. ■

Based on the above lemma, the assumptions of Propositions 19 and 22 are valid for the base case. Next, we control the Frobenius norm of the error term. From Proposition 19, we have

$$\|E_{t+1}\|_F \leq \|E_t\|_F + 22\delta\eta_0, \tag{88}$$

which implies

$$\|E_T\|_F = \|E_0\|_F + \sum_{t=1}^{T} \left(\|E_t\|_F - \|E_{t-1}\|_F\right) \leq \|E_0\|_F + 22\delta\eta_0 T. \tag{89}$$

Therefore, since $T \asymp \log\left(\frac{1}{\delta}\right)/\eta_0$ and $\alpha \asymp \sqrt{\delta}/\sqrt[4]{r'}$, we have $\|E_T\|_F \lesssim \delta \log \frac{r'}{\delta}$. Therefore, we have $\|E_T\| \leq \|E_T\|_F \lesssim \delta \log \frac{r'}{\delta}$. This shows that the error term remains small throughout the iterations of SubGM. Without loss of generality and to simplify our subsequent analysis, we assume that $\|E_T\| \leq \delta \log \frac{r'}{\delta}$, which can be ensured with sufficiently small η_0. Next, we control the signal term. Due to Proposition 22, we have

$$\|r_{t+1}\| \geq \left(1 + \eta_0(1 - \|r_t\|^2)\right)\|r_t\| - 10\eta_0\delta(\|E_t\| + \|r_t\|) - 2\eta_0 \|E_t\|^2 \|r_t\|. \tag{90}$$

Now, we separate our analysis into two stages. In the first stage, we show that the signal grows at a linear rate, provided that $\|r_t\| \leq 1/2$. To show this, we first prove that during the whole training process, the signal term is always larger than the error term.

Lemma 26 *Suppose that $\delta \leq 1/50$. Then, for any $0 \leq t \leq T = \Theta\left(\log\frac{1}{\alpha}/\eta_0\right)$, we have*

$$\|E_t\| \leq \|r_t\|. \tag{91}$$

Proof We prove this lemma by induction. For the base case, (91) holds since we have $\|r_0\| = \alpha(1 \pm O(\sqrt{\delta}))$, $\|E_0\| = \alpha O(\sqrt{\delta})$. Now, suppose that (91) holds at time t. Based on (66) and (69), we have

$$\|E_{t+1}\| \leq (1 + 5\eta_0\delta)\|E_t\| + 5\eta_0\delta\|r_t\| \leq (1 + 10\eta_0\delta)\|r_t\|. \tag{92}$$

On the other hand, due to (90), we have

$$\begin{aligned}\|r_{t+1}\| &\geq (1 + \eta_0(1 - \|r_t\|^2))\|r_t\| - 10\eta_0\delta(\|E_t\| + \|r_t\|) - 2\eta_0\|E_t\|^2\|r_t\| \\ &\geq (1 + \eta_0(1 - 3\|r_t\|^2))\|r_t\| - 20\delta\eta_0\|r_t\| \\ &\geq \left(1 + \frac{1}{5}\eta_0\right)\|r_t\|.\end{aligned} \tag{93}$$

Here we used the induction hypothesis $\|E_t\| \leq \|r_t\|$. The above two inequalities, together with $\delta \leq 1/50$, imply that $\|E_{t+1}\| \leq \|r_{t+1}\|$. ■

During the proof of the above lemma, we showed that

$$\|r_{t+1}\| \geq (1 + \eta_0/5)\|r_t\|. \tag{94}$$

provided that $\delta \leq 1/50$. Now, assuming that $T_1 \gtrsim \log\frac{1}{\alpha}/\eta_0$, we have

$$\|r_{T_1}\| \geq (1 - O(\sqrt{\delta}))\alpha(1 + \eta_0/5)^{T_1} \geq \frac{1}{2}, \tag{95}$$

This implies that, after T_1 iterations, the signal term will have a norm of at least $1/2$. In the second stage, we assume that $1 \geq \|r_t\| \geq 1/2$. One can write

$$\begin{aligned}\|r_{t+1}\| &\geq (1 + \eta_0(1 - \|r_t\|^2))\|r_t\| - 10\eta_0\delta(\|E_t\| + \|r_t\|) - 2\eta_0\|E_t\|^2\|r_t\| \\ &\geq (1 + \eta_0(1 - \|r_t\|))\|r_t\| - 20\eta_0\delta\|r_t\| - 4\eta_0\delta^2\log^2\frac{r'}{\delta},\end{aligned} \tag{96}$$

where we used $1 - \|r_t\|^2 \geq 1 - \|r_t\|$ given $\|r_t\| \leq 1$, and Lemma 26.

For the sake of simplicity, we define $x_t = 1 - \|r_{t+T_1}\|$. Hence, (96) can be simplified as

$$\begin{aligned}x_{t+1} &\leq 1 - (1 - 20\eta_0\delta + \eta_0 x_t)(1 - x_t) + 4\eta_0\delta^2\log^2\frac{r'}{\delta} \\ &\leq (1 - \eta_0 + 20\eta_0\delta)x_t + \eta_0 x_t^2 + 20\eta_0\delta + 4\eta_0\delta^2\log^2\frac{r'}{\delta} \\ &\leq (1 - \frac{3}{4}\eta_0)x_t + \frac{1}{2}\eta_0 x_t + 20\eta_0\delta\left(1 + \delta\log^2\frac{r'}{\delta}\right) \\ &\leq (1 - \eta_0/4)x_t + 20\eta_0\delta\left(1 + \delta\log^2\frac{r'}{\delta}\right),\end{aligned} \tag{97}$$

Here, we used $x_t \leq 1/2$ and $\delta \leq 1/80$. Then, we have

$$x_{t+1} - 80\delta\left(1 + \delta\log^2\frac{1}{\delta}\right) \leq \left(1 - \frac{\eta_0}{4}\right)\left(x_t - 80\delta\left(1 + \delta\log^2\frac{r'}{\delta}\right)\right), \tag{98}$$

which implies

$$x_{T_2} \leq 80\delta\left(1+\delta\log^2\frac{r'}{\delta}\right)+\frac{1}{2}\left(1-\eta_0/4\right)^{T_2}. \tag{99}$$

Upon choosing $T_2 \gtrsim \log\frac{1}{\delta}/\eta_0$, we have $x_{T_2} \lesssim \delta \vee \delta^2\log^2\frac{r'}{\delta}$, which is equivalent to $\|r_{T_1+T_2}\| \geq 1 - O\left(\delta+\delta^2\log^2\frac{r'}{\delta}\right)$.

This completes the proof under the assumption $\|r_{T_1+T_2}\| \leq 1$. Now, it remains to show that the error bound holds even if $\|r_{T_1+T_2}\| > 1$. To this goal, first we show that $T_3 = \Omega\left(\log\frac{r'}{\delta}/\eta_0\right)$ is necessary to guarantee the convergence of SubGM. In particular, we prove that we need at least $T_3 = \Omega\left(\log\frac{r'}{\delta}/\eta_0\right)$ to ensure $\|r_{T_3}\| \geq \frac{1}{2}$. To this goal, suppose that $\|r_t\| \leq 1/2$ for every $t \leq T$. Due to Proposition 22, we have

$$\begin{aligned}
\|r_{t+1}\| &\leq (1+\eta_0(1-\|r_t\|^2))\|r_t\| + 10\eta_0\delta(\|E_t\|+\|r_t\|) + 2\eta_0\|E_t\|^2\|r_t\|. \\
&\overset{(a)}{\leq} (1+20\eta_0\delta+\eta_0(1-\|r_t\|^2))\|r_t\| + 2\eta_0\|r_t\|^2\cdot\|r_t\| \\
&\overset{(b)}{\leq} \left(1+20\eta_0\delta+\eta_0+\frac{1}{2}\eta_0\right)\|r_t\| \\
&\leq (1+2\eta_0)\|r_t\|.
\end{aligned} \tag{100}$$

Here we used Lemma 26 and $\|r_t\| \leq \frac{1}{2}$ in (a) and (b), respectively. Therefore,

$$\|r_T\| \leq \alpha(1+2\eta_0)^T. \tag{101}$$

This shows that we need at least $T_3 = \Omega\left(\log\frac{1}{\alpha}/\eta_0\right) = \Omega\left(\log\frac{r'}{\delta}/\eta_0\right)$ iterations to guarantee $\|r_t\| \geq 1/2$. Now, suppose $\|r_{T_1+T_2}\| > 1$. Without loss of generality, we assume that $\|r_{T_1+T_2-1}\| \leq 1 < \|r_{T_1+T_2}\|$ (since T_3 and T_1+T_2 have the same order). Under this assumption, we show that $\|r_{T_1+T_2}\| \leq 1+O\left(\delta+\delta^2\log^2\frac{r'}{\delta}\right)$. By Proposition 19, we have

$$\begin{aligned}
\|r_{t+1}\| - \|r_t\| &\leq \eta_0(1-\|r_t\|^2)\|r_t\| + 10\eta_0\delta(\|E_t\|+\|r_t\|) + 2\eta_0\|E_t\|^2\|r_t\| \\
&\leq 6\eta_0(1-\|r_t\|) + 40\eta_0\delta + 4\eta_0\delta^2\log^2\frac{r'}{\delta}.
\end{aligned} \tag{102}$$

where we used the Lemma 26 and $\|r_t\| \leq 2$. Then, by our choice of $\|r_{T_1+T_2-1}\|$ and $\|r_{T_1+T_2}\|$, we have

$$\begin{aligned}
\|r_{T_1+T_2}\| - \|r_{T_1+T_2-1}\| &\leq 6\eta_0(1-\|r_{T_1+T_2-1}\|) + 40\eta_0\delta + 4\eta_0\delta^2\log^2\frac{r'}{\delta} \\
&\leq 6\eta_0\left(\|r_{T_1+T_2}\| - \|r_{T_1+T_2-1}\|\right) + 40\eta_0\delta + 4\eta_0\delta^2\log^2\frac{r'}{\delta}.
\end{aligned} \tag{103}$$

Then, since $\eta_0 \lesssim 1$, we have

$$\|r_{T_1+T_2}\| - 1 \leq \|r_{T_1+T_2}\| - \|r_{T_1+T_2-1}\| \lesssim \delta \vee \delta^2\log^2\frac{r'}{\delta}. \tag{104}$$

This implies that $|1-\|r_{T_1+T_2}\||\lesssim \delta \vee \delta^2\log^2\frac{r'}{\delta}$.

Finally, these two stages characterize the behavior of r_t and its convergence to the true solution. In particular, with the choice of $T = T_1 + T_2 = O(\log \frac{r'}{\delta}/\eta_0)$, and according to Lemma 15, we have

$$\begin{aligned}
\left\|U_T U_T^\top - X^\star\right\|_F^2 &\leq (1 - \|r_T\|^2)^2 + 2\|E_T\|^2\|r_T\|^2 + \|E_T\|_F^4 \\
&\lesssim \delta^2 + \delta^4 \log^4 \frac{r'}{\delta} + \delta^2 \log^2 \frac{r'}{\delta} + \delta^4 \log^4 \frac{r'}{\delta} \\
&\lesssim \delta^2 \log^2 \frac{r'}{\delta},
\end{aligned} \tag{105}$$

which completes the proof. □

■

Proofs for the Noisy Case

For simplicity of notation, we denote $\varphi_t = \varphi(\Delta_t)$, where $\Delta_t = U_t U_t^\top - X^*$.

D.5. Proof of Proposition 18

Analogous to the proof of Proposition 19, first we provide a general result which holds for arbitrary learning rates.

Lemma 27 *Suppose that $\|E_t\|_F \leq 1$, $\|r_t\| \leq 2$, $\delta \leq \frac{1}{2}$, then, the following inequalities hold*

$$\|E_{t+1}\|_F^2 \leq \|E_t\|_F^2 + 2\eta_t\varphi_t\left(-\frac{\|E_t U_t^\top\|_F^2}{\|\Delta_t\|_F} + \delta\left\|E_t U_t^\top\right\|_F\right) + 10\eta_t^2\varphi_t^2\left(\delta^2 + \frac{\|E_t U_t^\top\|_F^2}{\|\Delta_t\|_F^2}\right), \tag{106}$$

$$\|E_{t+1}\| \leq \left\|I - \frac{\eta_t\varphi_t U_t^\top U_t}{\|\Delta_t\|_F}\right\| \cdot \|E_t\| + \delta\eta_t\varphi_t(\|r_t\| + \|E_t\|). \tag{107}$$

Proof The proof is similar to that of Lemma 20. The details are omitted for brevity. ■

Now, we are ready to present the proof of Proposition 18.

Proof [Proof of Proposition 18]

Based on the sign-RIP condition, the step sizes satisfy

$$\eta_t = \frac{\eta_0\rho^t}{\|D\|_F} \leq \frac{\eta_0\rho^t}{\varphi_t(1-\delta)} \leq \frac{2\eta_0\rho^t}{\varphi_t}, \tag{108}$$

where $D \in \mathcal{M}(U_t U_t - X^*)$. For the Frobenius norm, we have

$$\begin{aligned}
\|E_{t+1}\|_F^2 &\leq \|E_t\|_F^2 + 2\eta_t\varphi_t\left(-\frac{\|E_t U_t^\top\|_F^2}{\|\Delta_t\|_F} + \delta\left\|E_t U_t^\top\right\|_F\right) + 10\eta_t^2\varphi_t^2\left(\delta^2 + \frac{\|E_t U_t^\top\|_F^2}{\|\Delta_t\|_F^2}\right) \\
&\leq \|E_t\|_F^2 + 2\eta_t\varphi_t\delta\left\|E_t U_t^\top\right\|_F + 10\delta^2\eta_t^2\varphi_t^2 \\
&\leq \|E_t\|_F^2 + 4\delta\eta_0\rho^t\left\|E_t U_t^\top\right\|_F + 20\delta^2\eta_0^2\rho^{2t}.
\end{aligned} \tag{109}$$

where in the second inequality, we used the assumption $\eta_0 \lesssim \delta \lesssim \|\Delta_t\|_F$, which implies

$$-2\eta_t\varphi_t \frac{\left\|E_t U_t^\top\right\|_F^2}{\|\Delta_t\|_F} + 10\eta_t^2\varphi_t^2 \frac{\left\|E_t U_t^\top\right\|_F^2}{\|\Delta_t\|_F^2} \leq 0 \tag{110}$$

Furthermore, note that

$$\begin{aligned}\left\|E_t U_t^\top\right\|_F^2 &= \left\|E_t E_t^\top\right\|_F^2 + \left\|E_t r_t u^{\star\top}\right\|_F^2 \\ &\leq \|E_t\|_F^4 + \|E_t\|_F^2 \|r_t\|^2 \\ &\leq (1+4)\|E_t\|_F^2,\end{aligned} \tag{111}$$

which implies

$$\begin{aligned}\|E_{t+1}\|_F^2 &\leq \|E_t\|_F^2 + 20\delta\eta_0\rho^t \|E_t\|_F + 20\delta^2\eta_0^2\rho^{2t} \\ &\leq \left(\|E_t\|_F + 10\delta\eta_0\rho^t\right)^2.\end{aligned} \tag{112}$$

This leads to $\|E_{t+1}\|_F \leq \|E_t\|_F + 10\delta\eta_0\rho^t$.

For the spectral norm, since we suppose $\eta_0 \lesssim \delta \lesssim \|\Delta_t\|_F$, we have $\left\|I - \frac{\eta_t\varphi_t U_t^\top U_t}{\|\Delta_t\|_F}\right\| \leq 1$. Combined with Lemma 27, this implies that

$$\|E_{t+1}\| \leq \|E_t\| + \delta\eta_t\varphi_t(\|r_t\| + \|E_t\|) \leq \|E_t\| + 2\delta\eta_0\rho^t(\|r_t\| + \|E_t\|). \tag{113}$$

thereby completing the proof. ■

D.6. Proof of Proposition 17

Similar to the proof of Proposition 22, we first present a general result which holds for arbitrary learning rates.

Lemma 28 *For any learning rate η_t, if $\|E_t\|_F \leq 1$, $\|r_t\| \leq 2$, then we have*

$$\begin{aligned}&\left\|r_{t+1} - \left(1 + \frac{\varphi_t\eta_t(1-\|r_t\|^2)}{\left\|U_tU_t^\top - X^\star\right\|_F}\right) r_t\right\| \\ \leq& \delta\eta_t\varphi_t(\|E_t\| + \|r_t\|) + \frac{\varphi_t\eta_t}{\left\|U_tU_t^\top - X^\star\right\|_F}\|E_t\|^2\|r_t\|.\end{aligned} \tag{114}$$

Proof The proof is similar to that of Lemma 23. The details are omitted for brevity. ■

Proof [Proof of Proposition 17] Assuming that $\delta \leq \frac{1}{2}$, we have

$$\left|\eta_t - \frac{\eta_0}{\varphi_t}\rho^t\right| = \left|\frac{\eta_0}{\|D\|_F}\rho^t - \frac{\eta_0}{\varphi_t}\rho^t\right| \leq \frac{\delta\varphi_t\eta_0\rho^t}{(1-\delta)\varphi_t^2} \leq \frac{2\delta\eta_0\rho^t}{\varphi_t}. \tag{115}$$

Combined with Lemma 28, this implies that

$$\begin{aligned}\left\|r_{t+1} - \left(1 + \frac{\eta_0\rho^t(1-\|r_t\|^2)}{\left\|U_tU_t^\top - X^\star\right\|_F}\right) r_t\right\| &\leq 2\delta\eta_0\rho^t(\|E_t\| + \|r_t\|) + \frac{2\eta_0\rho^t}{\left\|U_tU_t^\top - X^\star\right\|_F}\|E_t\|^2\|r_t\| \\ &\quad + \frac{2\delta\eta_0\rho^t}{\left\|U_tU_t^\top - X^\star\right\|_F}(1-\|r_t\|^2)\|r_t\|.\end{aligned} \tag{116}$$

which completes the proof. ■

D.7. Proof of Theorem 7

Recall that $T = \Theta(\log \frac{1}{\alpha}/\eta_0)$. First, we show that the error term remains small during the iterations of SubGM. Proposition 18 leads to

$$\|E_t\| \leq \|E_t\|_F = \|E_0\|_F + \sum_{t=1}^{t} (\|E_t\|_F - \|E_{t-1}\|_F) \leq \sqrt[4]{r'}\sqrt{\delta}\alpha + 10\delta\eta_0 T \lesssim \delta \log\frac{r'}{\delta}, \tag{117}$$

provided that $\|\Delta_t\|_F \geq 1 - \|r_t\|^2 \gtrsim \delta$. To verify this assumption, we show that $\|\Delta_t\|_F \geq 1 - \|r_t\|^2 \gtrsim \delta \log \frac{r'}{\delta}$ for every $t \leq \bar{T}$, where $\bar{T} \gtrsim \log \frac{1}{\alpha}/\eta_0$. To this goal, first we present a preliminary claim

Claim 1 *For every $0 \leq t \leq T$, we have $\|E_t\| \leq \|r_t\|$.*

Proof It follows an argument analogous to the proof of Lemma 26. The details are omitted for brevity. ■

Based on this claim, we are ready to show that $\|\Delta_t\|_F \geq 1 - \|r_t\|^2 \gtrsim \delta \log \frac{r'}{\delta}$ for every $t \leq \bar{T}$, where $\bar{T} \gtrsim \log \frac{1}{\alpha}/\eta_0$.

Claim 2 *Suppose that $\delta \leq 1/6$. Then, for every $0 \leq t \lesssim \log \frac{1}{\alpha}/\eta_0$, we have $\|r_t\| \leq \frac{1}{2}$.*

Proof The statement holds for $t = 0$ since $\|r_0\| = \Theta(\alpha)$. Now, suppose that $\|r_t\| \leq \frac{1}{2}$. Then, we have

$$\begin{aligned}
\|r_{t+1}\| &\leq \left(1 + \frac{4}{3}\frac{\eta_0\rho^t}{\|\Delta_t\|_F}(1 - \|r_t\|^2)\right)\|r_t\| + 2\delta\eta_0\rho^t(\|E_t\| + \|r_t\|) + \frac{2\eta_0\rho^t}{\|\Delta_t\|_F}\|E_t\|^2\|r_t\| \\
&\overset{(a)}{\leq} \left(1 + O(1)\,\eta_0\rho^t\right)\|r_t\| + 2\delta\eta_0\rho^t(\|E_t\| + \|r_t\|) \\
&\overset{(b)}{\leq} \left(1 + O(1)\eta_0\rho^t\right)\|r_t\|.
\end{aligned} \tag{118}$$

where in (a) we use the fact that $\|\Delta_t\|_F \geq 1 - \|r_t\|^2 \geq 3/4$ and $\|E_t\| \lesssim 1$; and in (b) we use Claim 1. Without loss of generality, we assume that $\|r_{t+1}\| \leq \left(1 + \eta_0\rho^t\right)\|r_t\|$. Hence, it suffices to show that

$$\|r_0\| \prod_{s=1}^{t}(1 + \eta_0\rho^s) = \Theta(\alpha)\prod_{s=1}^{t}(1 + \eta_0\rho^s) \leq \frac{1}{2}, \tag{119}$$

for every $0 \leq t \lesssim \log \frac{1}{\alpha}/\eta_0$. This is equivalent to

$$\sum_{s=1}^{t}\log\left(1 + \eta_0\rho^s\right) \leq \log\frac{1}{2\alpha}. \tag{120}$$

On the other hand, note that

$$\sum_{s=1}^{t}\log\left(1 + \eta_0\rho^s\right) \leq \sum_{s=1}^{t}\eta_0\rho^s \leq \eta_0\frac{1-\rho^t}{1-\rho} \leq C\log\frac{1}{\alpha}\left(1 - \rho^t\right). \tag{121}$$

Therefore, to finish the proof, we need to show that $C\log\frac{1}{\alpha}\left(1 - \rho^t\right) \leq \log\frac{1}{2\alpha}$, which implies $1 - \frac{1}{C} + \frac{\log 2}{C\log\frac{1}{\alpha}} \leq \rho^T$. This can be easily verified for every $t \lesssim \log\frac{1}{\alpha}/\eta_0$, by noting that $\rho = 1 - \Theta(\eta_0/\log\frac{1}{\alpha})$. ■

Based on the above claim and upon choosing $\bar{T} \asymp \log\frac{1}{\alpha}/\eta_0$, the error term is bounded as (117) for every $t \leq \bar{T}$. Now, note that the proof is completed if $\|\Delta_t\|_F \lesssim \delta\log\frac{r'}{\delta}$ for some $\bar{T} \leq t \leq T$. Therefore, suppose that $\|\Delta_t\|_F \gtrsim \delta\log\frac{r'}{\delta}$ for every $\bar{T} \leq t \leq T$. This implies that the error bound (117) holds for every $\bar{T} \leq t \leq T$. Moreover, we assume that $1 - \|r_t\|^2 \geq 3\|E_t\|_F$, since otherwise, we have $1-\|r_t\|^2 \lesssim \delta\log(r'/\delta)$, and the proof is completed together with $\|E_t\|_t \leq \delta\log(r'/\delta)$ and Lemma 15. This leads to

$$1-\|r_t\|^2 \leq \|\Delta_t\|_F \leq 1-\|r_t\|^2 + \|E_t\|\|r_t\| + \|E_t\|_F^2 \leq \frac{13}{9}(1-\|r_t\|^2). \tag{122}$$

assuming that $\|r_t\| \leq 1$. Then, according to Proposition 17, we have

$$\begin{aligned}\|r_{t+1}\| &\geq \left(1+\frac{2}{3}\frac{\eta_0\rho^t}{\|\Delta_t\|_F}(1-\|r_t\|^2)\right)\|r_t\| - 2\delta\eta_0\rho^t(\|E_t\|+\|r_t\|) - \frac{2\eta_0\rho^t}{\|\Delta_t\|_F}\|E_t\|^2\|r_t\| \\ &\overset{(a)}{\geq} \left(1+\Omega(1)\eta_0\rho^t\right)\|r_t\| - 2\delta\eta_0\rho^t\|E_t\|.\end{aligned} \tag{123}$$

where in (a) we used $\|E_t\|^2 \leq (1-\|r_t\|^2)/9$, inequality (122), and $\delta \lesssim 1$. To proceed, note that $\|E_t\| \leq \|r_t\|$ due to Claim 1. Hence, we have

$$\|r_{t+1}\| \geq \left(1+\Omega(1)\eta_0\rho^t\right)\|r_t\|. \tag{124}$$

for every $0 \leq t \leq T$. Now, it remains to show that after $T = O\left(\log\left(\frac{1}{\alpha}\right)/\eta_0\right)$ iterations, the signal term approaches 1. Without loss of generality, we assume that $\|r_{t+1}\| \geq (1+\eta_0\rho^t)\|r_t\|$, which implies $\|r_T\| \geq \alpha\prod_{t=1}^{T}\left(1+\eta_0\rho^t\right)$. Taking the logarithm of the right hand side leads to

$$\sum_{t=1}^{T}\log\left(1+\eta_0\rho^t\right) \geq \sum_{t=1}^{T}\frac{\eta_0\rho^t}{1+\eta_0\rho^t} \geq \frac{\eta_0}{2}\frac{1-\rho^T}{1-\rho}. \tag{125}$$

where we used the lower bound $\log(1+x) \geq \frac{x}{1+x}$ for $x \geq -1$. Now, upon defining $\gamma = 1-\rho$, we have

$$\begin{aligned}\frac{\eta_0}{2}\frac{1-\rho^T}{1-\rho} &= \frac{\eta_0}{2}\frac{1-(1-\gamma)^T}{\gamma} \\ &\geq \frac{\eta_0}{2\gamma}\left(1-\left(1-\frac{\gamma T}{1+(T-1)\gamma}\right)\right) \\ &\geq \frac{\eta_0}{2\gamma}\frac{\gamma T}{2}.\end{aligned} \tag{126}$$

where we used the basic inequality $(1-x)^r \leq 1-\frac{rx}{1+(r-1)x}$ for $x \in [0,1], r > 1$. Now, recalling $T = \Theta\left(\log\frac{1}{\alpha}/\eta_0\right)$ and $\gamma = \Theta\left(\eta_0/\log\frac{1}{\alpha}\right)$, we have $\frac{\eta_0}{2\gamma}\frac{\gamma T}{2} \geq \log(1/\alpha)$, which implies that after $T = \Theta(\log\frac{1}{\alpha}/\eta_0)$ iterations, the signal term satisfies $\|r_T\| \geq 1$. So, the only remaining part is to show that $\|r_T\| = 1 \pm O(\delta\log\frac{r'}{\delta})$. Recall that, based on the definition of $\bar{T}$, we have $\|r_{\bar{T}}\| < 1$. Now, we assume that $\|r_{T-1}\| < 1$, and $\|r_T\| \geq 1$. Note that this assumption is without loss of generality, since $\bar{T}$ and T have the same order. Then we have the following claim.

Claim 3 *Either* $1-\delta\log\frac{r'}{\delta} \lesssim \|r_{T-1}\|^2$, *or* $\|r_T\| \lesssim 1+\delta^2\log\frac{r'}{\delta}$.

Proof Assume that $\|\Delta_{T-1}\|_F \geq 1 - \|r_{T-1}\|^2 \gtrsim \delta \log\frac{r'}{\delta}$. Then, by Proposition 17, we have

$$\begin{aligned}
\|r_T\| - \|r_{T-1}\| &\leq \frac{4}{3}\frac{\eta_0\rho^{T-1}(1-\|r_{T-1}\|^2)}{\|\Delta_{T-1}\|_F}\|r_{T-1}\| + \frac{2\eta_0\rho^{T-1}\|E_{T-1}\|^2}{\|\Delta_{T-1}\|_F}\|r_{T-1}\| + O(\delta\eta_0\rho^T) \\
&\lesssim \frac{1}{\log\frac{r'}{\delta}}(1-\|r_{T-1}\|) + \delta^2\log\frac{r'}{\delta} \\
&\lesssim \frac{1}{\log\frac{r'}{\delta}}(\|r_T\| - \|r_{T-1}\|) + \delta^2\log\frac{r'}{\delta}
\end{aligned} \tag{127}$$

This implies that, for sufficiently small δ, we have $\|r_T\| - \|r_{T-1}\| = O(\delta^2\log\frac{r'}{\delta})$, thereby completing the proof. ■

In summary, we showed that $1 - \delta\log\frac{r'}{\delta} \lesssim \|r_{T-1}\|^2 \leq 1$, or $1 \leq \|r_T\| \lesssim 1 + \delta^2\log\frac{r'}{\delta}$. On the other hand, we know that $\|E_t\| \lesssim \delta\log\frac{r'}{\delta}$ for every $t \leq T$. This together with Lemma 15 completes the proof. □

Appendix E. Proof of Proposition 2

We divide our analysis into two cases. In the first case, we assume $p\sigma^2 = \Omega(1)$. We have

$$\begin{aligned}
\sup_{X\in\mathbb{S}}\|Q(X)-X\|_F &= \sup_{X,Y\in\mathbb{S}}\left|\frac{1}{m}\sum_{i=1}^m \langle A_i, X\rangle\langle A_i, Y\rangle + \frac{1}{m}\sum_{i\in S} s_i\langle A_i, Y\rangle - \langle X, Y\rangle\right| \\
&\overset{(a)}{\geq} \sup_{Y\in\mathbb{S}}\left|\frac{1}{m}\sum_{i=1}^m \langle A_i, Y\rangle^2 + \frac{1}{m}\sum_{i\in S} s_i\langle A_i, Y\rangle - 1\right| \\
&\geq \sup_{Y\in\mathbb{S}}\left|\frac{1}{m}\sum_{i\in S} s_i\langle A_i, Y\rangle\right| - \sup_{Y\in\mathbb{S}}\left|\frac{1}{m}\sum_{i=1}^m \langle A_i, Y\rangle^2 - 1\right| \\
&\overset{(b)}{=} \left\|\frac{1}{m}\sum_{i\in S} s_i A_i\right\|_F - \sup_{Y\in\mathbb{S}}\left|\frac{1}{m}\sum_{i=1}^m \langle A_i, Y\rangle^2 - 1\right|.
\end{aligned} \tag{128}$$

where in (a) we add a constraint $X = Y$ to the supremum; and in (b) we use the Cauchy-Schwartz inequality and the variational form of the Frobenius norm. By the ℓ_2-RIP for Gaussian measurements (Lemma 31), we have

$$\sup_{X,Y\in\mathbb{S}}\left|\frac{1}{m}\sum_{i=1}^m \langle A_i, X\rangle\langle A_i, Y\rangle + \frac{1}{m}\sum_{i\in S} s_i\langle A_i, Y\rangle - \langle X, Y\rangle\right| \geq \left\|\frac{1}{m}\sum_{i\in S} s_i A_i\right\|_F - \delta_1 \tag{129}$$

with probability of at least $1 - Ce^{-cm\delta_1^2}$, given $m \gtrsim d^2$. The expectation and tail bound of $\left\|\frac{1}{m}\sum_{i\in S} s_i A_i\right\|_F$ is provided in the following lemma.

Lemma 29 *For any $0 < t < 1$, we have*

$$\mathbb{P}\left(\left|\left\|\frac{1}{m}\sum_{i\in S} s_i A_i\right\|_F - \mathbb{E}\left[\left\|\frac{1}{m}\sum_{i\in S} s_i A_i\right\|_F\right]\right| \geq t\right) \leq 2e^{-\frac{Cmt^2}{p\sigma^2 d^2}}, \tag{130}$$

where C is a universal constant. Moreover, the expectation is lower bounded as

$$\mathbb{E}\left[\left\|\frac{1}{m}\sum_{i\in S} s_i A_i\right\|_F\right] \gtrsim \sqrt{\frac{p\sigma^2 d^2}{m}}. \tag{131}$$

Before providing the proof of Lemma 29, we complete the proof of Proposition 2. Based on the above lemma and (129), we have

$$\sup_{X,Y\in\mathbb{S}}\left|\frac{1}{m}\sum_{i=1}^{m}\langle A_i, X\rangle\langle A_i, Y\rangle + \frac{1}{m}\sum_{i\in S} s_i\langle A_i, Y\rangle - 1\right| \geq C\sqrt{\frac{p\sigma^2 d^2}{m}} - \delta_1 - \delta_2, \tag{132}$$

with probability of at least $1 - Ce^{-c_1 m\delta_1^2} - e^{-c_2\frac{m\delta_2^2}{p\sigma^2 d^2}}$. Hence, with the proper choice of δ_1, δ_2, we have

$$\mathbb{P}\left(\sup_{X\in\mathbb{S}}\|Q(X) - X\|_F \geq C\sqrt{\frac{p\sigma^2 d^2}{m}}\right) \geq \frac{1}{2}. \tag{133}$$

Since $p\sigma^2 = \Omega(1)$, we can choose C' such that

$$\mathbb{P}\left(\sup_{X\in\mathbb{S}}\|Q(X) - X\|_F \geq C'\sqrt{\frac{(1+p\sigma^2)d^2}{m}}\right) \geq \frac{1}{2}. \tag{134}$$

In the second case, we assume that $p\sigma^2 = O(1)$. Making a similar argument, we can show that there exists a universal constant C such that

$$\mathbb{P}\left(\sup_{X\in\mathbb{S}}\|Q(X) - X\|_F \geq C\sqrt{\frac{d^2}{m}}\right) \geq \frac{1}{2}. \tag{135}$$

Combining the two cases, the following inequality holds for an arbitrary $\sigma > 0$

$$\mathbb{P}\left(\sup_{X\in\mathbb{S}}\left\|M_2(X) - \bar{M}_2(X)\right\|_F \geq C'\sqrt{\frac{(1+p\sigma^2)d^2}{m}}\right) \geq \frac{1}{2}. \tag{136}$$

Which completes the proof of Proposition 2. $\square$

Now, we present the proof for Lemma 29.

Proof [Proof of Lemma 29] For simplicity, we denote $B = \frac{1}{m}\sum_{i\in S} s_i A_i$. First, we prove the lower bound on the expectation. Note that, conditioned on s_i, we have $B_{j,k} = \frac{1}{m}\sum_{i\in S} s_i A^i_{j,k} \sim N\left(0, \frac{1}{m^2}\sum_{i\in S} s_i^2\right)$. Then, by invoking Theorem 3.1.1. in [18], we have

$$\begin{aligned}\mathbb{E}\left[\|B\|_F\right] &= \mathbb{E}\left[\mathbb{E}\left[\|B\|_F\right]|s_i, i\in S\right]\\ &\gtrsim \mathbb{E}\left[\frac{d}{m}\sqrt{\sum_{i\in S} s_i^2}\right]\\ &\gtrsim \frac{\sigma d}{m}\sqrt{pm} = \sqrt{\frac{p\sigma^2 d^2}{m}}.\end{aligned} \tag{137}$$

Now, we show that $\|B\|_F$ is a sub-exponential random variable. First, for arbitrary indices i, j, k, the random variable $s_i A^i_{j,k}$ is sub-exponential according to Lemma 36 since $\left\|s_i A^i_{j,k}\right\|_{\psi_1} \leq \|s_i\|_{\psi_2} \left\|A^i_{j,k}\right\|_{\psi_2} = \Theta(\sigma)$. This implies that $\|B_{j,k}\|_{\psi_1} = \Theta\left(\sqrt{\frac{p\sigma^2}{m}}\right)$. Finally, we have

$$\begin{aligned}
\|\|B\|_F\|_{\ell^{2k}} &= \left(\left\|\sum_{j,k} B^2_{j,k}\right\|_{\ell^k}\right)^{1/2} \\
&\overset{(a)}{\leq} \left(\sum_{j,k} \|B^2_{j,k}\|_{\ell^k}\right)^{1/2} \\
&= d\,\|B_{j,k}\|_{\ell^{2k}} \lesssim \sqrt{\frac{p\sigma^2 d^2}{m}} k.
\end{aligned} \tag{138}$$

which implies that $\|B\|_F$ is sub-exponential with sub-exponential norm $O\left(\sqrt{\frac{p\sigma^2 d^2}{m}}\right)$ due to the equivalent definition of sub-exponential random variable (see Definition 35). Note that in (a) we used the Minkowski inequality. Given the lower bound on the expected value, the tail bound directly follows from the tail of sub-exponential distribution. ■

Appendix F. Auxiliary Lemmas

F.1. Restricted Isometry Property

Lemma 30 *Let $\mathbb{S}_r = \{X \in \mathbb{R}^{d\times d} : \mathrm{rank}(X) \leq r, \|X\|_F = 1\}$. Then, there exists an ϵ-covering $\mathbb{S}_{\epsilon,r}$ with respect to the Frobenius norm satisfying $|\mathbb{S}_\epsilon| \leq \left(\frac{9}{\epsilon}\right)^{(2d+1)r}$.*

Lemma 31 (ℓ_2-RIP, Theorem 4.2 in [15]) *Fix $0 < \delta < 1$, suppose that the measurement matrices $\{A_i\}_{i=1}^m$ have i.i.d. standard Gaussian entries. Then, we have*

$$\sup_{X\in\mathbb{S}_r} \left|\frac{1}{m}\sum_{i=1}^m \langle A_i, X\rangle^2 - \|X\|_F^2\right| \leq \delta. \tag{139}$$

with probability of at least $1 - Ce^{c_1 dr \log\frac{1}{\delta} - c_2 m\delta^2}$.

F.2. Basic Probability

Lemma 32 (Conditional Gaussian Variable in Bivariate Case) *For two Gaussian random variables $X \sim \mathcal{N}(\mu_1, \sigma_1^2), Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ with correlation coefficient ρ, we have*

$$X|Y = a \sim \mathcal{N}\left(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho\,(a - \mu_2)\,, \left(1-\rho^2\right)\sigma_1^2\right). \tag{140}$$

Definition 33 (Sub-Gaussian random variable) *We say a random variable $X \in \mathbb{R}$ with expectation $\mathbb{E}[X] = \mu$ is σ^2-sub-Gaussian if for all $\lambda \in \mathbb{R}$, we have $\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}}$. This definition is equivalent to the following statements*

- *(Tail bound) For any $t > 0$, we have $\mathbb{P}(|X - \mu| \geq t) \leq 2e^{-\frac{t^2}{2\sigma^2}}$.*

- *(Moment bound) For any positive integer p, we have $\|X\|_{\ell^p} = \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p} \lesssim \sigma\sqrt{p}$.*

Moreover, the sub-Gaussian norm of X is defined as $\|X\|_{\psi_2} := \sup_{p\geq 1}\left\{p^{-1/2}\|X\|_{\ell^p}\right\}$.

For sum of independent sub-Gaussian random variables, their sub-Gaussian norm can be bounded via the following lemma.

Lemma 34 (Proposition 2.6.1 in [18]) *Let $X_1, \cdots, X_m$ be a series independent zero-mean sub-Gaussian variables, then $\sum_{i=1}^m$ is sub-Gaussian and*

$$\left\|\sum_{i=1}^m X_i\right\|_{\psi_2}^2 \lesssim \sum_{i=1}^m \|X_i\|_{\psi_2}^2. \tag{141}$$

Definition 35 (Sub-exponential random variable) *A random variable X with expectation μ is sub-exponential if there exists (μ, α), such that $\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\frac{\lambda^2\nu^2}{2}}$ for all $|\lambda| \leq \alpha$. This definition is equivalent to the following statements:*

- *(tail bound) There exists a universal constant C, for any $t > 0$, we have $\mathbb{P}(|X - \mu| \geq t) \leq 2e^{-Ct}$.*

- *(moment bound) For any positive integer p, we have $\|X\|_{\ell^p} = \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p} \lesssim p$.*

Moreover, the sub-exponential norm of X is defined as $\|X\|_{\psi_1} := \sup_{p\geq 1}\left\{p^{-1}\|X\|_{L^p}\right\}$.

For sub-Gaussian and sub-exponential random variables, we have the following lemma to illustrate their relations.

Lemma 36 *The following statements hold*

- *(Lemma 2.7.6 in [18]) A random variable X is sub-Gaussian if and only if X^2 is sub-exponential. Moreover, $\|X\|_{\psi_2}^2 = \left\|X^2\right\|_{\psi_1}$.*

- *(Lemma 2.7.7 in [18]) Let X and Y be sub-Gaussian random variables. Then XY is sub-exponential. Moreover, $\|XY\|_{\psi_1} \leq \|X\|_{\psi_2}\|Y\|_{\psi_2}$.*

F.3. Basic Inequalities

Lemma 37 (Bernoulli inequality) *The following inequality holds*

$$(1+x)^r \leq 1 + \frac{rx}{1-(r-1)x}, \quad \textit{for } x \in \left[-1, \frac{1}{r-1}\right), r \geq 1. \tag{142}$$