Robust spectral compressive sensing via vanilla gradient descent
Xumneng Wu, Zai Yang and Zongben Xu

Abstract—This paper investigates robust recovery of an undamped or damped spectrally sparse signal from its partially revealed noisy entries within the framework of spectral compressed sensing. Nonconvex optimization approaches such as projected gradient descent (PGD) based on low-rank Hankel matrix completion model have recently been proposed for this problem. However, the analysis of PGD relies heavily on the operation of projection onto feasible set involving two tuning parameters, and the theoretical guarantee in noisy case is still missing. In this paper, we propose a vanilla gradient descent (VGD) algorithm without projection based on low-rank Hankel noisy matrix completion, and prove that VGD can achieve the sample complexity $O\left(K^2 \log^2 N\right)$, where $K$ is the number of the complex exponential functions and $N$ is the signal dimensions, to ensure robust recovery from noisy observations when noise parameter satisfies some mild conditions. Moreover, we show the possible performance loss of PGD, suffering from the inevitable estimation of the above two unknown parameters of feasible set. Numerical simulations are provided to corroborate our analysis and show more stable performance obtained by VGD than PGD when dealing with damped spectrally sparse signal.

Index Terms—Robust spectral compressive sensing, nonconvex optimization, low-rank Hankel noisy matrix completion.

I. INTRODUCTION

Robust spectral compressive sensing refers to reconstruct a spectrally sparse signal from its incomplete and noisy observation. In particular, the spectrally sparse signal sequence $\{x_n\}$ with or without damping that is corrupted by noise is given by

$$x_n = x_n^* + \varepsilon_n, \quad x_n^* = \sum_{k=1}^{K} s_k e^{(i2\pi f_k - r_k)n}, \quad n = 0, 1, \ldots, N-1$$

where $i = \sqrt{-1}$, $x_n^*$ denotes the noiseless signal, $\{\varepsilon_n\}$ is noise, $f_k \in [0, 1)$, $k = 1, \ldots, K$ are the unknown frequencies, $s_k \in \mathbb{C}$ denotes the amplitude, $r_k \geq 0$ is a damping factor and the sparsity origins from the fact that the complex exponential functions number $K$ is small compared with the samples number $N$. This signals are encountered in many practical applications such as analog-to-digital conversion [1], target localization in radar system [2], magnetic resonance imaging [3], etc. In this paper, our objective is to recovery the true signal $\{x_n^*\}$ from partial noisy observed entries in $\{x_n\}$.

Unlike conventional compressive sensing, spectral compressive sensing [4] doesn’t assume the frequencies in the model lie in the fixed grid, and it tries to estimate signal and frequencies in a continuous way, which essentially belongs to the grid-less sparse methods [5] and can eliminate the mismatch error between the true frequencies and the discrete frequencies [6]. Interestingly, all famous spectral compressive sensing methods [7–12] transform the sparsity signal to the low-rankness in special corresponding matrix, and so are closely related with the low-rank matrix recovery problem [13].

Roughly speaking, these methods can be divided into two class, positive-semidefinite- (PSD-) Toeplitz-based and Hankel-based. For the first class, there exist two approaches, i.e. atomic norm minimization (ANM) [7] and covariance fitting [11, 12]. The author in [7] proved that ANM could achieve exact recovery from $O\left(K \log K \log N\right)$ random samples via semidefinite programming (SDP) provided the frequencies are well separated in noiseless case. When dealt with noisy data, [14] [15] showed that the mean-squared error rate of recovery by ANM via SDP was bounded by $O\left(\tilde{\sigma}^2 \left(K \log N \right) / N\right)$, where $\tilde{\sigma}^2$ is Gaussian noise variance.

For Hankel-based methods, [8] developed enhanced matrix completion (EMaC) approach that solved a convex Hankel matrix nuclear norm minimization problem, and showed that $O\left(K \log^4 N\right)$ samples number could guarantee perfect recovery with high probability under some mild incoherence conditions in noiseless data, and it was also stable against bounded noise. Moreover, some nonconvex algorithms like fast iterative hard thresholding (FIHT) [9] and PGD [10] were introduced to cope with the high computational complexity of convex approaches. Particularly, they showed that the required sampling number for exact recovery is $O\left(K^2 \log N\right)$ for PGD in absence of noise.

Although PGD shows absolute superiority in terms of both memory and computation when dealing with large-scale data, there are some aspects that remain unclear. Firstly, the theoretical analysis of gradient descent based on low-rank Hankel matrix completion from noisy data is still lacking [16]. Secondly, as mentioned in numerical experiments [10] that the projection onto feasible set of PGD may be unnecessary in practice, it is ill-defined whether there exists similar theoretical guarantee for VGD without projection.

In this paper, we prove that robust recovery from incomplete noisy data via VGD can be guaranteed with high probability as soon as the number of samples exceeds the order of $O\left(K^2 \log^2 N\right)$ under mild noise condition. The proof is motivated by the application of the leave-one-out techniques in [17] [18], whose effectiveness has been widely verified in high dimensional regression [19], stochastic spectral problems [20] and low-rank recovery problems such as phase retrieval.
Our Contributions

The main contributions of our paper are threefold. Firstly, we note in this paper that, for PGD in solving the low-rank Hankel matrix completion problem, the discrepancy between its theoretical results heavily relying on the projection operation and its numerical experiments in [10] oppositely showing the useless of projection has not been well understood.

So the first contribution of our paper is that we prove the VGD without projection in our problem has linear convergence and the same iteration complexity as PGD. Specifically speaking, different from the route of proving the regularity condition [23] [10] for PGD, the analysis of VGD demonstrates the iterates automatically lie on the region of incoherence and smoothness. As a product, the existence of implicit regularization [17] for gradient descent in low-rank Hankel matrix completion problem is also confirmed.

Secondly, we present the required sample complexity and noise condition of VGD for robust recovery of a spectrally sparse signal from the observed incomplete noisy data. To be specific, we show that the signal recovery error bound of VGD matches the minimax lower bound in [14]. When degrading to the noiseless case, $$O(K^2 \log N)$$ random observed entries are sufficient to exactly recover the original signal by VGD, almost keeping the same as that of PGD.

Thirdly, the potential issue of PGD resulting from the unavoidable choice of two tuning parameters in the projection operation is discussed and verified by extensive numerical simulations. Besides, for VGD, the recovery guarantee and its stable performance and superiority of simplified algorithm step comparing with PGD are also validated through experiments.

Notations and Organization

Notations used in this paper are as follows. The set of real and complex numbers are denoted $$\mathbb{R}$$ and $$\mathbb{C}$$ respectively. Symbols for matrices (upper case) and vectors (lower case) are in boldface. The number of vector and matrix elements starts at zero except that the index $$k$$ starts from one. $$[N]$$ denotes the set $$\{0, 1, ..., N - 1\}$$ for a natural number $$N$$. For vector, $$x^T$$, $$x^H$$, $$||x||_1$$ and $$||x||_2$$ denote its transpose, complex transpose, spectral norm and Frobenius norm of its rows and $$||A||_\infty$$ represents the largest magnitude of all entries of $$A$$. The inner product of matrices $$A$$ and $$B$$ is defined as $$\langle A, B \rangle = \text{trace} (A^H B)$$. We use $$A_{i,j}$$ and $$A_{i,:}$$ to denote its $$i$$th column and $$i$$th row, respectively. The sign matrix of $$A$$ is denoted as $$\text{sign}(A) = UV^H$$ if the singular value decomposition (SVD) of $$A$$ is $$U \Sigma V^H$$. The identity matrix and identity operator are denoted as $$I$$ and $$\mathcal{I}$$. The diagonal matrix with vector $$x$$ on the diagonal is denoted diag$$(x)$$. The notation $$O(K)$$ means the set of all $$K \times K$$ unitary matrices and let $$\mathbb{R}(\cdot)$$ denote the real part of a complex number. Denote $$e_i$$ is the standard basis of $$\mathbb{R}$$ with proper dimensions. We use $$C_0, C_1, C_2, ...$$ to denote positive constants whose values may change from place to place.

The rest of the paper is organized as follows. Section II introduces some preliminaries. Section III presents the VGD algorithm and our main result Theorem I. In Section IV we give the proof of Theorem I by the help of three Propositions. Some novel key Lemmas in our paper is presented in Section V. In Section VI numerical simulations are provided to demonstrate the performance of VGD algorithm and validate our theory results. Section VII concludes this paper. We present some other lemmas and the proof of Proposition 13 in Appendix.
diag \((\sigma_1, \sigma_2, \ldots, \sigma_K)\), \(\sigma_1 \geq \ldots \geq \sigma_K\) and the condition number \(\kappa = \frac{\sigma_1}{\sigma_K}\) are used across this paper. Then the \(\mu\)-incoherence means
\[
\|U\|_{2,\infty}^2 = \max_{0 \leq i \leq N_1-1} \left\| e_i^T A_1 \left( A_1^H A_1 \right)^{-1/2} \right\|_2^2 \\
\leq \max_{0 \leq i \leq N_1-1} \left\| e_i^T A_1 \right\|_2 \left\| \left( A_1^H A_1 \right)^{-1/2} \right\|_2 \leq \frac{\mu \epsilon N}{K} \tag{5}
\]
and \(\|V\|_{2,\infty}^2 \leq \frac{\mu \epsilon N}{K} \tag{8}\). Therefore the \(\mu\)-incoherence of \(\mathcal{G} y^*\) can be understood that \(U\) and \(V\) remain incoherent with respect to the sensing vector \(e_i\). Moreover, we denote
\[
M = \begin{bmatrix} M_U \\ M_V \end{bmatrix} = \begin{bmatrix} U \Sigma^{1/2} \\ V \Sigma^{1/2} \end{bmatrix}
\]
so we have \(M_U M_V^H = \mathcal{G} y^*\).

III. OUR ALGORITHM AND MAIN RESULTS

A. Objective function and Algorithm

Let \(\Omega \subseteq [N]\) be the sampling index set of size \(M \leq N\), and \(\mathcal{P}_\Omega\) be the projection onto the subspace supported on \(\Omega\) that sets all entries out of \(\Omega\) to zero. Denote the sampling ratio \(p = \frac{M}{N}\). In this paper, we consider the case that \(\Omega\) obeys an i.i.d. Bernoulli model with probability \(p\). Then the low-rank Hankel noisy matrix completion with rank constraint can be formally expressed as
\[
\min_{x \in \mathbb{C}^N} \| \mathcal{P}_\Omega (x - \bar{x}) \|_2^2, \quad \text{subject to rank}(\mathcal{H} \bar{x}) = K. \tag{6}
\]

To reduce computational complexity, we adopt the famous Burman-Monte Carlo factorization \[24\] to make \(\mathcal{H} \bar{x} = Z_U Z_V^H\) with two factor matrices \(Z_U \in \mathbb{C}^{N_1 \times K}\) and \(Z_V \in \mathbb{C}^{N_2 \times K}\). Moreover, there exists an another constraint that \(Z_U Z_V^H\) must be a Hankel matrix, i.e., \((I - \mathcal{G} \mathcal{G}^H) (Z_U Z_V^H) = 0\). Therefore, we can rewrite the above problem \[6\] into an unconstrained nonconvex minimization problem following from \[10\]:
\[
\min_{Z_U, Z_V} f(Z_U, Z_V) = \frac{1}{2} \left\| (I - \mathcal{G} \mathcal{G}^H) (Z_U Z_V^H) \right\|_F^2 \\
+ \frac{1}{2p} \left\| \mathcal{P}_\Omega \mathcal{G} \mathcal{G}^H (Z_U Z_V^H - \mathcal{G} y^*) \right\|_2^2 \\
+ \frac{1}{8} \left\| Z_U^H Z_U - Z_V^H Z_V \right\|_F^2, \tag{7}
\]
where \(\left\| Z_U^H Z_U - Z_V^H Z_V \right\|_F^2\) is the balanced regularizer \[2.4\] which has been widely used to control the possible imbalance between factor matrices and enhance the convergence of gradient descent in nonconvex optimization. Namely, it reduces the solutions set from \(S_1 = \{ M : P, P \in \mathbb{C}^{K \times K} \text{ is invertible} \}\) to \(S_2 = \{ M : Q, Q \in \mathbb{C}^{K \times K} \text{ is unitary} \}\). In the noiseless case, we have \(f(M_U, M_V) = 0\).

Denote \(f(Z) := f(Z_U, Z_V), Z := [Z_U^H, Z_V^H]^H \in \mathbb{C}^{(N_1 + N_2) \times K}\) and \(C_0 = \frac{1}{p} \mathcal{P}_\Omega(y)\). We present a vanilla gradient descent (VGD) algorithm without projection in Algorithm \[1\] to solve the problem \[7\]. Specifically speaking, the algorithm follows the framework of two-stage nonconvex approaches. The first stage focuses on spectral initialization by one step hard thresholding, which is implemented by truncating the SVD of the Hankel matrix formulated by the observed noisy entries from \(y\). The operator \(T_K(\cdot)\) is the top-\(K\) partial SVD. This initialization manner can guarantee the initial point for gradient descent in the subsequent stage keeps close to the true solution.

**Algorithm 1** Vanilla gradient descent (VGD) for low-rank Hankel noisy matrix completion

**Initialization:** Let \(\mathcal{T}_K(G^0) = U^0 \Sigma^0 (V^0)^H\), then \(Z^0 = \begin{bmatrix} Z_U^0 \\ Z_V^0 \end{bmatrix}\).

\[Z_{t+1} = Z_t - \eta \nabla f(Z^t) \]

\[\text{end for}\]

**Output:** \(Z^t, y^t = \mathcal{G}^H \left( Z_U^t \left( Z_V^t \right)^H \right)\) and \(x^t = \mathcal{D} \mathcal{D}^{-1} y^t\).

In the second stage, simple gradient descent with a fixed step size \(\eta\) is executed to update the estimate \(Z^t\). The gradient without noise of \(f(Z)\) is given by
\[
\nabla f(Z) = \begin{bmatrix} \nabla f_U(Z) \\ \nabla f_V(Z) \end{bmatrix}, \tag{8}
\]
where
\[
\nabla f_U(Z) = (I - \mathcal{G} \mathcal{G}^H) \left( Z_U Z_V^H \right) Z_V \\
+ \frac{1}{p} \mathcal{P}_\Omega \mathcal{G} \mathcal{G}^H \left( Z_U Z_V^H - \mathcal{G} y^* \right) Z_V \\
+ \frac{1}{2} Z_U \left( Z_V^H Z_U - Z_V^H Z_U \right), \\
\nabla f_V(Z) = \left[ (I - \mathcal{G} \mathcal{G}^H) \left( Z_U Z_V^H \right) \right]^H Z_U \\
+ \frac{1}{p} \left[ \mathcal{P}_\Omega \mathcal{G} \mathcal{G}^H \left( Z_U Z_V^H - \mathcal{G} y^* \right) \right]^H Z_U \\
+ \frac{1}{2} Z_V \left( Z_U^H Z_V - Z_U^H Z_U \right). \\
\]

The true gradient \(\nabla f(Z)\) of \[7\] is given by
\[
\nabla f(Z) = \begin{bmatrix} \nabla f_U(Z) \\ \nabla f_V(Z) \end{bmatrix}. \tag{9}
\]
where
\[
\nabla f_U(Z) = \nabla f_U(Z) + \frac{1}{p} \mathcal{H} \mathcal{P}_\Omega(\epsilon) Z_V, \\
\nabla f_V(Z) = \nabla f_V(Z) + \frac{1}{p} \left[ \mathcal{H} \mathcal{P}_\Omega(\epsilon) \right]^H Z_U,
\]
and we recall that \(\epsilon\) is the noise vector.

B. Comparison between PGD and VGD

VGD in this paper is similar to PGD in \[10\] despite of the difference of dealing with noiseless or noisy signal. However, the projection operation is an significant component in PGD algorithm, while it is removed in our VGD algorithm. To be more specific, the author of PGD defines a convex set as \(\mathcal{C} = \{ Z : \| Z \|_{2,\infty} \leq \beta = \sqrt{\frac{2c_K \kappa}{N}} \}\) with two tuning parameters \(\nu \geq \mu \) and \(\alpha \geq \sigma_1\). Because there holds \(\|M\|_{2,\infty} \leq \beta\) when \(\mathcal{G} y^*\) is \(\mu\)-incoherent, it is obvious that \(S_2 \subset \mathcal{C}\) and then \(\mathcal{C}\) is a feasible set. Then projection operation \(\mathcal{P}_\mathcal{C}(Z^0)\) and \(\mathcal{P}_\mathcal{C}(Z^t)\) are adopted to restrict the solution on
the $C$ during initialization and iteration procedure, respectively, which essentially constrains the iterates keep incoherence with the sensing vector $e_i$ in an artificial manner. The projection operation $\mathcal{P}_C(Z)$ in PGD is computed by row-wise trimming, i.e.

$$(\mathcal{P}_C(Z))_{i,:} = \begin{cases} Z_{i,:} & \text{if } \|Z_{i,:}\|_2 \leq \beta, \\ \frac{Z_{i,:}}{\beta} & \text{otherwise}. \end{cases}$$

(10)

Due to the absence of original complete noiseless signal $x^*$, the true incoherence parameter $\mu$ and the maximum singular value $\sigma_1$ are obviously unknown for us. Hence, a natural question is how to determine the two tuning parameters $\nu$ and $\alpha$ in $C$. The way adopted by PGD in experiments of (10) is to estimate them from the initialization, i.e. select the estimation of $\beta$ be $\tilde{\beta}$ as:

$$\tilde{\beta} = \max \left\{ \|U\|_{2,\infty}, \|V\|_{2,\infty} \right\} \cdot \sqrt{\sigma_0^2},$$

(11)

where $\sigma_0^2$ is the maximum singular value from $\Sigma^0$.

We debate this manner (11) has potential trouble and will possibly bring terrific loss in signal recovery performance. Obviously, there doesn’t exist enough theoretical evidence that guarantees $\tilde{\beta} \geq \frac{\mu c \sqrt{\kappa^2 + 1}}{N}$. In fact, opposite phenomena can be observed in the Section VI which implies the simulation results of PGD with (11) may be inconsistent with its theoretical consideration.

Therefore, it can be divided into two case for analysis. Firstly, when $\beta \geq \frac{\mu c \sqrt{\kappa^2 + 1}}{N}$ or $\frac{\mu c \sqrt{\kappa^2 + 1}}{N} > \beta \geq \|M\|_{2,\infty}$ holds, the iterates $\{Z^t\}$, originating from the initialization $Z^0$ that satisfying $\|Z^0\|_{2,\infty} \leq \tilde{\beta}$ from (11), will converge to the global optimum $M$, as illustrated in Figure 1(a). In other words, it holds that $\|Z^t\|_{2,\infty} \leq \tilde{\beta}$ and the projection $\mathcal{P}_C(Z)$ essentially doesn’t work according to (10). The PGD algorithm running in this case equals to the VGD. This case will be verified by the subsequent theoretical results of VGD, which shows the iterates keep incoherence throughout.

Secondly, if $\beta < \|M\|_{2,\infty}$ holds, the row-wise trimming in the second line of (10) will be executed when $\|Z^t\|_2 > \beta$ in the iteration, and then the projection $\mathcal{P}_C(Z)$ becomes an additional constrain that preserves the iterates converge to $M$, as shown in Figure 1(b). In this case, there is always a gap between the convergence point $Z^*$ of PGD and the global optimum $M$, which results in poor performance for PGD in signal recovery error. However, our VGD without projection operation will succeed to converge to the $M$, proved in Theorem I and verified by simulations in Section VI.

Our VGD has the same computation complexity per iteration $O(NK^2 + KN \log N)$ flops as PGD. Such operation of $Z^H_U Z_U$, $G^H (Z_U Z_U^H)$ and $(G v) Z_V$ for vector $v \in \mathbb{C}^N$ in the computation of $\nabla f(Z)$ dominate the main computational cost per iteration. Specifically, $Z^H_U Z_U$ can be computing using $O(NK^2)$ flops, and $G^H (Z_U Z_U^H)$ and $(G v) Z_V$ can be computing using $K$ fast convolutions and $K$ fast Hankel matrix-vector multiplications with $O(KN \log N)$ flops, respectively. Moreover, $\mathcal{H}$ becomes a symmetric but not Hermitian matrix and hence $\mathcal{H} \tilde{x} = Z_U Z_U^T$ holds when $N_1 = N_2$, which can save nearly half of the computational costs compared with the rectangular case.

C. Main results

Our main results for the theoretical guarantee of VGD are stated as follows:

**Theorem 1.** Assume $G y^*$ is a rank $K$, $\mu$-incoherent matrix, and noise entries $\{\epsilon_n\}$ are independent sub-Gaussian random variables with sub-Gaussian norm $\sigma$ [20]. Recall that $\sigma_1$ and $\sigma_K$ are the largest and smallest singular values of $G y^*$, respectively. Let $\rho = 1 - 0.01 \sigma_K$ and the gradient descent step size $\eta \leq \frac{\sigma_0^2}{\rho \sigma_{\infty}^2}$, then with probability exceeding $1 - N^{-4}$, the iteration in Algorithm I converges linearly for at least the first $N^5$ steps:

$$\min_{Q \in O(K)} \|Z^t Q - M\|_F \leq \rho^t \sqrt{\sigma_K + C_2 \sigma_{\infty} \sqrt{\kappa^5 KN^2}}$$

(12)

for $0 \leq t \leq N^5$ provided $M \geq C_m \mu^2 c^2 \kappa^{11} K^2 \log^2 N$ and $\sigma_{\infty} \sqrt{c s M} \leq C_n \sqrt{\mu c^2 K \log N}$ with proper absolute constants $C_m$ and $C_n$.

**Remark 1.** (1) Our VGD algorithm stays the same sample complexity with PGD in (10) which can achieve exact recovery if the number of observed entries is of order $O(\mu^2 c^2 K^2 \log N)$ when $\kappa = O(1)$ in absence of noise. Besides, the number of iterations of VGD to attain $\epsilon$-accuracy is bounded by $O(\kappa^2 \log \frac{1}{\epsilon})$, which means VGD also nearly has the same iteration complexity with PGD.

(2) When $t$ increases, (12) converges to

$$\|Z^t \tilde{x} - M\|_F \leq C_2 \sigma_{\infty} \sqrt{\kappa^5 KN^2} \sqrt{\sigma_0^2},$$

(13)

Additionally, we have

$$\|x^t - x^*\|_2 \leq \frac{1}{\sqrt{2}} \left( \|Z^t\|_2 + \|M\| \right) \|Z^t \tilde{x} - M\|_F$$

1The $\log^2 N$ item in VGD’s sample complexity stems from the noise control in Lemma I which can be reduced to $\log N$ in the noiseless case.
where the first inequality follows from (10), the second one holds due to $\|Z^t\| \leq \|Z^t Q^t - M\|$, and (22) and (13), and the last one holds provided $\frac{\sigma}{\sigma_K} \sqrt{\frac{N}{c_M}} \leq \frac{1}{\sqrt{2}} \frac{\sqrt{\kappa}}{\sqrt{\kappa}}$.

It means that $\frac{1}{\sqrt{2}} \frac{\|x^t - \hat{x}^t\|}{\|x^t - \hat{x}^t\|}$ is the order of $O(\epsilon^2)$ when omitting other items and the order of $O(\frac{K}{N})$ since $\frac{\sigma}{\sigma_K} \sqrt{\frac{N}{c_M}} \leq \frac{1}{\sqrt{2}} \frac{\sqrt{\kappa}}{\sqrt{\kappa}}$, which gives the signal recovery error bound by rank-constrained Hankel noisy matrix completion and matches the minimax lower bound in (14).

(3) Extra projection operation in PGD, which essentially can be understood as an explicit regularization, is unnecessary in our VGD algorithm. Namely, we unveil the implicit regularization property of gradient descent in low-rank Hankel noisy matrix completion.

IV. PROOF OF THEOREM 1

Before starting the proof of Theorem 1, it should be mentioned that although our proof follows the leave-one-out analysis framework from [17], extending to unstructured rectangular matrix completion in [18], the objective function in this work is rather different and the details in our proof are quite technical.

Specifically speaking, our objective function consists of an extra but essential regularizer to keep the Hankel structure of the matrix, and the loss term contains the transformation between matrix and vector. About the proof procedure, starting from spectral initialization, we provide the near-optimal bound of $\|\frac{1}{p}G_P(y) - \hat{G}y^t\|$ in noisy case in Lemma 3 based on Lemma 1 and [9] Lemma 2 (restated as Lemma 2 in our paper). Moreover, different from the manner of control spectral norm in [13] Lemma 7, we prove Lemma 5 according to our specific model by applying matrix Bernstein inequality [27] Theorem 6.1.1. Finally, when analyzing local properties of the Hessian, we carefully demonstrate Lemma 6 to give the suitable bound. Owing to these new theoretical results, the sample complexity of VGD without projection can still keep almost the same compared with PGD’s.

Here we begin to prove Theorem 1. The proof uses the fact that the empirical loss function (7) has benign geometry along certain directions near the global optimum in the average case. Following a similar trajectory analysis in [17], the first step is to establish the region of incoherence and contraction (RIC) (cf.19) and (20) where our objective loss function (7) enjoys local strong convexity along certain directions (cf.21) and smoothness (cf.22) as shown in Proposition 1. The second step is to demonstrate the spectral initial guess in Algorithm 1 lies inside RIC (revealed in Proposition 2), and to prove that the iterates satisfy the $\mu$-incoherence condition throughout by induction (shown in Proposition 3).

It’s difficult to directly prove the iterates always stays in the RIC because there doesn’t exist extra regularization operation in Algorithm 1 to enforce incoherence constraints. Namely, due to the complicated statistical dependence between the sequence $\{Z^t\}$ and the sampling operator $\mathcal{P}_l$, we need resort to the leave-one-out perturbation argument to break it and make them independent so that matrix Bernstein inequality can be used (such as (85)).

Concretely speaking, we consider the modified nonconvex optimization problem in (15) and define the leave-one-out sequences $\{Z_{l,(i)}\}$ for each $l \in [0, N_l - 1]$ as the corresponding iterates of gradient descent. The new loss function is

$$\min_{Z_U, Z_V} h(Z_U, Z_V) = \frac{1}{2} \left\| (I - G \hat{G}^H) \left( Z_U Z_V^H \right) \right\|_F^2 + \frac{1}{2p} \left\| \mathcal{P}_{l,-1} \mathcal{P}_l \hat{G} \hat{G}^H \left( Z_U Z_V^H - \hat{G}y\right) \right\|_F^2 + \frac{1}{2} \left\| \mathcal{P}_l \mathcal{G} \hat{G}^H \left( Z_U Z_V^H - \hat{G}y^t\right) \right\|_F^2 + \frac{1}{8} \left\| Z_U^H Z_U - Z_V^H Z_V \right\|_F^2,$$

where $\mathcal{P}_l(\cdot)$ represents the orthogonal projection onto the subspace of matrices that vanish outside of the l-th row, and $\mathcal{P}_{l,-1}(\cdot)$ is the projector that transforms a matrix by setting the l-th row into zeros and remaining all other entries. Notice that we assume $0 \leq l \leq N_l - 1$ in this paper by default, if there is no special instructions. The similar loss function and gradient descent iterates can be obtained by transforming the projection from l-th row onto l-th column when $N_l \leq l \leq N_l + N_l - 2$, so we omit it here for simplicity.

The gradient of $h(Z_U, Z_V)$ is given by

$$\nabla h(Z) = \left[ \nabla h_{U}(Z) \quad \nabla h_{V}(Z) \right],$$

where

$$\nabla h_{U}(Z) = \left( I - G \hat{G}^H \right) \left( Z_U Z_V^H \right) Z_V + \frac{1}{p} \mathcal{P}_{l,-1} \mathcal{P}_l \hat{G} \hat{G}^H \left( Z_U Z_V^H - \hat{G}y\right) Z_V + \mathcal{P}_l \mathcal{G} \hat{G}^H \left( Z_U Z_V^H - \hat{G}y^t\right) Z_V + \frac{1}{2} Z_U \left( Z_U^H Z_U - Z_V^H Z_V \right),$$

$$\nabla h_{V}(Z) = \left[ (I - G \hat{G}^H) \left( Z_U Z_V^H \right) \right]^H Z_U + \frac{1}{p} \left[ \mathcal{P}_{l,-1} \mathcal{P}_l \hat{G} \hat{G}^H \left( Z_U Z_V^H - \hat{G}y\right) \right]^H Z_U + \mathcal{P}_l \mathcal{G} \hat{G}^H \left( Z_U Z_V^H - \hat{G}y^t\right)^H Z_U + \frac{1}{2} Z_V \left( Z_U^H Z_V - Z_V^H Z_U \right).$$

Denote $G_{0,(l)} = \frac{1}{p} \mathcal{P}_{l,-1} \mathcal{P}_l \hat{G} \hat{G}^H + \mathcal{P}_l \hat{G} y^t$ and $\mathcal{T}_K \left( G_{0,(l)} \right) = U_{0,(l)} \Sigma_{0,(l)} \left( V_{0,(l)} \right)^H.$ The spectral initialization and gradient descent for $\{Z_{l,(i)}\}$ are shown in Algorithm 2. It should be noticed that the iteration in Algorithm 2 only plays an auxiliary role for proving Theorem 1 and needn’t be executed in solving the primal problem (7) in practice.
Algorithm 2: The leave-one-out sequence for low-rank Hankel noisy matrix completion

Initialization:

Let \( Z_0^{(l)} = \begin{bmatrix} Z_U^{(l)} \\ Z_V^{(l)} \end{bmatrix} = \begin{bmatrix} U^{0,(l)} \left( \Sigma^{0,(l)} \right)^{1/2} \\ V^{0,(l)} \left( \Sigma^{0,(l)} \right)^{1/2} \end{bmatrix} \).

for \( t = 0, 1, \ldots, J - 1 \) do

\( Z^{t+1,(l)} = Z^{t,(l)} - \eta \nabla h \left( Z^{t,(l)} \right) \).

end for

This leave-one-out sequence \( \{ Z^{t,(l)} \} \) introduced above will be the indispensable component in the proof of Proposition 2 and Proposition 3. Obviously, the calculations between \( Z^t \) and \( Z^{t,(l)} \) only differ by one row/column, and so they remain greatly close. Moreover, \( Z^{t,(l)} \) is independent with the sampling operator \( \mathcal{P}_l \mathcal{S} \mathcal{P}_{l+1} \), and hence we can easily bound its relevant norm of the l-th row/column and show its incoherence. Therefore, \( \{ Z^{t,(l)} \} \) is used as the bridge for proving the incoherence of \( \{ Z \} \) in the following discussion.

Proposition 1. Suppose that \( Z = \begin{bmatrix} Z_U^H & Z_V^H \end{bmatrix} \in \mathbb{C}^{(N_1+N_2)\times K} \) and \( W = \begin{bmatrix} W_U^H & W_V^H \end{bmatrix} \in \mathbb{C}^{(N_1+N_2)\times K} \) satisfy

\[
\| Z - M \|_{2,\infty} \leq \frac{1}{800 \sqrt{N_1 + N_2}} \sqrt{\sigma_1}, \tag{19}
\]

and \( W = BQ_B - C \), with \( Q_B = \arg \min_{Q \in \mathcal{O}(K)} \| BQ - C \|_F \) and

\[
\| C - M \| \leq \frac{1}{800 \sqrt{\sigma_1}}, \tag{20}
\]

where \( B = \begin{bmatrix} B_U^H \\ B_V^H \end{bmatrix} \) and \( C = \begin{bmatrix} C_U^H \\ C_V^H \end{bmatrix} \).

Then we have

\[
\Re \left\{ \text{vec}(W)^H \nabla^2 f^* (Z) \text{vec}(W) \right\} \geq \frac{1}{10} \sigma_K \| W \|_F^2, \tag{21}
\]

and

\[
\| \nabla^2 f^* (Z) \| \leq 10 \sigma_1, \tag{22}
\]

provided \( M \geq C_{10} \mu c_s K \log N \) with probability at least \( 1 - N^{-10} \) for an absolute constant \( C_{10} \).

Proposition 1 characterizes the Hessian \( \nabla^2 f^* (Z) \) without noise local properties of the objective function, including strong convexity and smoothness. Its proof mainly invokes Lemma 6 and Lemma 7 in our paper. According to (19), we have

\[
\| Z \|_{2,\infty} \leq \| Z - M \|_{2,\infty} + \| M \|_{2,\infty} \leq \frac{1}{800 \sqrt{N_1 + N_2}} + \sqrt{\frac{\mu c_s K}{N}} \sqrt{\sigma_1}, \tag{23}
\]

where the final inequality holds because \( 800 \sqrt{\mu c_s K^2} \geq 1 \). Hence (19) means \( Z \) satisfies the incoherence in (5).

To control the distances between initialization and the truth, we need denote some unitary matrices:

\[
Q^0 := \arg \min_{Q \in \mathcal{O}(K)} \| Z^0 Q - M \|_F, \tag{24}
\]

\[
Q^{0,(l)} := \arg \min_{Q \in \mathcal{O}(K)} \| Z^{0,(l)} Q - M \|_F, \tag{25}
\]

\[
T^{0,(l)} := \arg \min_{Q \in \mathcal{O}(K)} \| Z^0 Q^{0,(l)} - Z^{0,(l)} Q \|_F. \tag{26}
\]

Proposition 2. Suppose the observed samples number satisfies \( M \geq 6400 \mu c_s^2 c_s^2 K^2 \log^2 N \) and the noise satisfies \( \frac{\sigma}{\sigma_K} \sqrt{\frac{N^2}{c_M}} \leq \frac{1}{800 \sqrt{\sigma_1}} \), then the following inequalities

\[
\| Z^0 Q^0 - M \| \leq 90 \left( C_1 \sqrt{\frac{\mu c_s K \log N}{M}} + \frac{C_2 \sigma}{\sigma_K} \sqrt{\frac{K \log N}{c_M}} \right) \sqrt{\sigma_1}, \tag{27}
\]

\[
\| (Z^{0,(l)} Q^{0,(l)} - M) \|_2 \leq 68 \left( C_1 \sqrt{\frac{\mu c_s K \log N}{M}} + \frac{C_2 \sigma}{\sigma_K} \sqrt{\frac{K \log N}{c_M}} \right) \sqrt{\sigma_1}, \tag{28}
\]

\[
\| Z^0 Q^0 - Z^{0,(l)} T^{0,(l)} \|_F \leq 912 \left( C_6 \sqrt{\frac{\mu c_s^2 K \log N}{MN}} + \frac{C_7 \sigma}{\sigma_K} \sqrt{\frac{K \log N}{c_M}} \right) \sqrt{\sigma_1}, \tag{29}
\]

\[
\| Z^0 Q^0 - M \|_{2,\infty} \leq 4560 \left( C_1 \sqrt{\frac{\mu c_s^2 K \log N}{MN}} + \frac{C_2 \sigma}{\sigma_K} \sqrt{\frac{K \log N}{c_M}} \right) \sqrt{\sigma_1}, \tag{30}
\]

hold for \( l = 0, 1, \ldots, N_1 + N_2 - 1 \) with probability at least \( 1 - N^{-8} \) for some absolute constants \( C_1, C_2, C_6, \) and \( C_7 \).

Proposition 2 contains error bounds for the initialization \( Z^0 \) and the leave-one-out sequences \( \{ Z^{0,(l)} \} \). The proof is similar to [18, Lemma 3] and need mainly invoke [17, Lemma 45-47], and Lemma 14 in this paper. When \( M \geq 9120 c_s^2 \mu c_s K \log N \) and \( \frac{\sigma}{\sigma_K} \leq \frac{1}{800 \sqrt{\sigma_1}} \), we can obtain \( \| Z^0 Q^0 - M \|_{2,\infty} \leq \frac{\mu c_s K}{N} \sqrt{\sigma_1} \) from (30), which means that the initialization lies in the RIC.

To guarantee the convergence of iteration procedure, we denote:

\[
Q^t := \arg \min_{Q \in \mathcal{O}(K)} \| Z^t Q - M \|_F, \tag{31}
\]

\[
Q^{t,(l)} := \arg \min_{Q \in \mathcal{O}(K)} \| Z^{t,(l)} Q - M \|_F, \tag{32}
\]

\[
T^{t,(l)} := \arg \min_{Q \in \mathcal{O}(K)} \| Z^t Q^t - Z^{t,(l)} Q \|_F. \tag{33}
\]

Proposition 3. Suppose the sample size satisfies \( M \geq C_n \mu c_N^2 K^{11} \log^2 N \) and noise satisfies \( \frac{\sigma}{\sigma_K} \sqrt{\frac{N^2}{c_M}} \leq \frac{1}{C_n \mu c_N^2 K \log N} \) for some positive constants \( C_n \) and \( C_n \).
related to $C_1, C_6, C_8$ and $C_0, C_2, C_7, C_8$ respectively, and the following inequalities

$$
\|Z^t Q^t - M\| \\
\leq \left( C_1 \rho \sqrt{\frac{\mu_c \kappa^5 K \log N}{M}} + C_2 \frac{\sigma}{\kappa \sigma K} \sqrt{\frac{k^5 N^2}{c \sigma M}}\right) \sqrt{\sigma_1},
$$  
(34)

$$
\left\| \left(Z^{t,(l)} Q^{t,(l)} - M\right)_{l_i} \right\|_2 \\
\leq \left( C_1 \rho \sqrt{\frac{\mu_c \kappa^5 K \log N}{MN}} + C_2 \frac{\sigma}{\kappa \sigma K} \sqrt{\frac{\mu_c \kappa \log N}{M}}\right) \sqrt{\sigma_1},
$$  
(35)

$$
\left\| Z^t Q^t - Z^{t,(l)} T^{t,(l)} \right\|_F \\
\leq \left( C_0 \rho \sqrt{\frac{\mu_c \kappa^5 K \log N}{MN}} + C_2 \frac{\sigma}{\kappa \sigma K} \sqrt{\frac{\mu_c \kappa \log N}{M}}\right) \sqrt{\sigma_1},
$$  
(36)

$$
\left\| Z^t Q^t - M \right\|_{2,\infty} \\
\leq \left( C_1 \rho \sqrt{\frac{\mu_c \kappa^5 K \log N}{MN}} + C_2 \frac{\sigma}{\kappa \sigma K} \sqrt{\frac{\mu_c \kappa \log N}{M}}\right) \sqrt{\sigma_1},
$$  
(37)

hold for $l = 0, 1, ..., N_1 + N_2 - 1$ and $\rho = 1 - 0.01 \eta \sigma K$ with probability at least $1 - N^{-10}$. Then these inequalities also hold for $t + 1$.

Proposition 8 gives the error bounds for the iterates $\{Z^t\}$ and the leave-one-out sequences $\{Z^{t,(l)}\}$ in an induction manner. Its proof primarily invokes on our novel Lemma 9 and other lemmas like Lemma 3, Lemma 10, Lemma 11 and so on.

The error contraction with respect to spectral norm is presented in (34), and (35) depicts the incoherence of $\{Z^{t,(l)}\}$ with respect of the given $l$-th row/column. What is expressed in (36) is the fact that the iterates $\{Z^t\}$ and the auxiliary iterates $\{Z^{t,(l)}\}$ remain close enough to some proper unitary transformations.

When $M \geq 16 C_1 \mu_c \kappa^5 K \log N$ and $\frac{\sigma}{\kappa \sigma K} \sqrt{\frac{N^2}{c \sigma M}} \leq \frac{1}{4c \sqrt{\kappa}}$, we can obtain

$$
\left\| Z^t Q^t - M \right\|_{2,\infty} \leq \frac{1}{2} \sqrt{\frac{\mu_c \kappa \log N}{c \sigma M}} \sqrt{\sigma_1}
$$

It means that the iterates $Z^t$ in VGD remain weakly correlated with the sensing vectors $e_i, 0 \leq i \leq N_1 - 1$, throughout. In other words, $\{Z^t\}$ always keeps incoherence and lies on the feasible convex set $C$ in the iterative process even if there is no designed projection operation in VGD.

Proof of Theorem 1 According to Proposition 2 and Proposition 3 we can obtain that

$$
\left\| Z^t Q^t - M \right\|_F \leq \sqrt{K} \left\| Z^t Q^t - M\right\| \\
\leq \sqrt{K} \left( C_1 \rho \sqrt{\frac{\mu_c \kappa^5 K \log N}{M}} + C_2 \frac{\sigma}{\kappa \sigma K} \sqrt{\frac{\mu_c \kappa^5 N^2}{c \sigma M}}\right) \sqrt{\sigma_1},
$$  
(38)

$$
\leq \rho \sqrt{\sigma_1} + C_2 \frac{\sigma}{\kappa \sigma K} \sqrt{\frac{\mu_c \kappa^5 N^2}{c \sigma M}} \sqrt{\sigma_1},
$$

where the final inequality follows from $M \geq C_0^2 \mu_c \kappa^5 K^2 \log N$. By applying (38) iteratively for $t = 1, 2, ..., N^5$, it holds with probability exceeding $1 - (1 + N^5)N^{-10} \geq 1 - N^{-4}$ for all $t \in [0, N^5]$.

V. Key Lemmas

In this section, we list some novel key lemmas that are used in the proof of Proposition 11.

Lemma 1. Recall from [1] that $\varepsilon \in \mathbb{C}^{N \times 1}$ is a random noise vector and each entries $\varepsilon_n$ is an independent sub-Gaussian random variable with parameter $\sigma$. Suppose $M \geq c_0 \mu_c \log^2 N_2$ for some sufficiently large constant $c_0 > 0$, then

$$
\frac{1}{p} \|H \mathcal{P}_\Omega (\varepsilon)\| \leq C_0 \sigma \sqrt{\frac{N^2}{c_0 M}}
$$

holds with probability exceeding $1 - N^{-12}$ for a universal constant $C_0 > 0$.

Proof. See Appendix A in [?].

Lemma 2 (Lemma 2). Assume $G\hat{y}$ is $\mu$-incoherent, then $\|G\hat{y}\|_{2,\infty} \leq \sqrt{\frac{\mu_c \kappa}{K}} \|G y\|$ and $\|G \hat{y}\|_{\infty} \leq \sqrt{\frac{\mu_c \kappa}{N}} \|G y\|$.

Moreover, there exists a universal constant $C_1 > 0$ such that

$$
\frac{1}{p} \|G \mathcal{P}_\Omega (y^*) - G \hat{y}\| \leq C_1 \sqrt{\frac{\mu_c \kappa \log N}{M}} \|G \hat{y}\|
$$

with probability at least $1 - N^{-12}$ provided $M \geq \mu_c K \log N$.

Lemma 3. Assume $G \hat{y}$ is $\mu$-incoherent, and each noise entries $\varepsilon_n$ is independent sub-Gaussian random variable with sub-Gaussian norm $\sigma$. Notice $y = Dx = Dx^* + De = y^* + De$, then

$$
\frac{1}{p} \|G \mathcal{P}_\Omega (y) - G \hat{y}\| \\
\leq \left( C_1 \sqrt{\frac{\mu_c \kappa \log N}{M}} + C_2 \frac{\sigma}{\kappa \sigma K} \sqrt{\frac{N^2}{c \sigma M}}\right) \|G \hat{y}\|
$$

holds with probability exceeding $1 - N^{-11}$ for some universal constants $C_1 > 0$ and $C_2 > 0$ provided $M \geq \mu_c K \log^2 N$.

Proof. By triangle inequality, we have

$$
\frac{1}{p} \|G \mathcal{P}_\Omega (y) - G \hat{y}\| \\
\leq \left( C_1 \sqrt{\frac{\mu_c \kappa \log N}{M}} + C_2 \frac{\sigma}{\kappa \sigma K} \sqrt{\frac{N^2}{c \sigma M}}\right) \frac{1}{p} \|G \mathcal{P}_\Omega (y^*) - G \hat{y}\| + \frac{1}{p} \|H \mathcal{P}_\Omega (\varepsilon)\|.
$$

(42)

Then the proof is completed by combining Lemma 2 and Lemma 11.

Lemma 4. Denote $G^* = G \hat{y}$, $G = G y$, and recall that $G^0 = \frac{1}{p} \mathcal{P}_\Omega (y)$ and $G^{0,(l)} = \frac{1}{p} \mathcal{P}_{l-\Omega} (y) + \mathcal{P}_\Omega (\varepsilon)$ and their corresponding symmetric matrices are $\overline{G}^* = \begin{bmatrix} 0 & G^* \\ G^* & 0 \end{bmatrix}$, $\overline{G}^0 = \begin{bmatrix} 0 & G^0 \\ (G^0)^H & 0 \end{bmatrix}$ and $\overline{G}^{0,(l)} = \begin{bmatrix} 0 & G^{0,(l)} \\ (G^{0,(l)})^H & 0 \end{bmatrix}$.
\[ \begin{bmatrix} 0 & \mathbf{G}^{0,(l)} \\ \mathbf{G}^{0,(l)} & 0 \end{bmatrix}^H. \] Then for all \( 0 \leq l \leq N_1 + N_2 - 1 \), the following formula
\[ \| \mathbf{G}^{0,(l)} - \mathbf{G}^* \| \leq \| \mathbf{G}^1 - \mathbf{G}^* \| = \| \mathbf{G}^0 - \mathbf{G}^* \| \] holds.

**Proof.** According to the above matrices definitions, \( \mathbf{G}^{0,(l)} - \mathbf{G}^* \) is a submatrix of \( \mathbf{G}^1 - \mathbf{G}^* \), and so (43) holds.

**Lemma 5.** For any matrix \( \mathbf{A} \in \mathbb{C}^{N_1 \times K} \) and \( \mathbf{B} \in \mathbb{C}^{N_2 \times K} \), we have
\[ \| \frac{1}{p} \mathbf{G} \mathbf{P}_0 \mathbf{G}^H (\mathbf{AB}) - \frac{1}{M} \mathbf{G} \mathbf{G}^H (\mathbf{AB}) \| \leq C_8 \sqrt{\frac{NN_1 \log N}{M^2}} \| \mathbf{A} \|_{2,\infty} \| \mathbf{B} \|_{2,\infty} \] provided \( M \geq c_8 \log N \) for a universal constant \( C_8 > 0 \) with probability exceeding \( 1 - N^{-12} \).

**Proof.** Let
\[ \mathbf{Z}_{nm} := \frac{1}{p} \mathbf{G} \mathbf{P}_{nm} \mathbf{G}^H (\mathbf{AB}) - \frac{1}{M} \mathbf{G} \mathbf{G}^H (\mathbf{AB}) \] define the matrix family for \( n_m \in [N] \). Obviously we have \( \mathbb{E} [\mathbf{Z}_{nm}] = 0 \) and
\[ \| \mathbf{Z}_{nm} \| \leq 2 \max_{n_m \in [N]} \frac{1}{p} \| \mathbf{G} \mathbf{P}_{nm} \mathbf{G}^H (\mathbf{AB}) \| = \frac{2}{p} \| \mathbf{D} \mathbf{G} (\mathbf{AB}) \|_{\infty} \leq \frac{2}{p} \| \mathbf{AB} \|_{\infty} \leq \frac{2}{p} \| \mathbf{A} \|_{2,\infty} \| \mathbf{B} \|_{2,\infty}. \] Let \( \mathbf{G}_n \) denote a \( N_1 \times N_2 \) matrix whose entries are zero except that the \( n \)th skew-diagonal entries equal to \( \frac{1}{\sqrt{2n}} \) and \( r = \mathbf{G} (\mathbf{AB}) \). Then \( \mathbf{Z}_{nm} = \frac{1}{p} \mathbf{G} \mathbf{P}_{nm} r - \frac{1}{\sqrt{2}} \mathbf{G} \mathbf{r} \). So we have
\[ \sum_{m=1}^{M} \mathbb{E} [\mathbf{Z}_{nm} \mathbf{Z}^H_{nm}] = \sum_{m=1}^{M} \frac{1}{p^2} \mathbb{E} \left[ \left\| \mathbf{r}_{nm} \right\|_2^2 \mathbf{G} \mathbf{G}^H \mathbf{G}_n \right] = \frac{M}{p^2} \left( \frac{N-1}{N} \sum_{n=0}^{N-1} \left\| \mathbf{r}_{nm} \right\|_2^2 \mathbf{G} \mathbf{G}^H \mathbf{G}_n \right) \leq \frac{N}{M} \left\| \mathbf{G} \mathbf{G}^H (\mathbf{AB}) \right\|_{2,\infty}^2 \] where the first inequality follows from the fact \( \| \mathbf{C} - \mathbf{D} \| \leq \max \{ \| \mathbf{C} \|, \| \mathbf{D} \| \} \) for positive semidefinite matrices \( \mathbf{C} \) and \( \mathbf{D} \). Similarly we can get \( \sum_{m=1}^{M} \mathbb{E} [\mathbf{Z}_{nm}^H \mathbf{Z}_{nm}] \leq \sum_{m=1}^{M} \mathbb{E} [\mathbf{Z}_{nm} \mathbf{Z}^H_{nm}] \leq \frac{M}{N_2} \left\| \mathbf{AB} \right\|_{\infty}^2 \leq \frac{M}{N_2} \left\| \mathbf{A} \right\|_{2,\infty}^2 \left\| \mathbf{B} \right\|_{2,\infty}^2. \] By matrix Bernstein inequality [27] Theorem 6.1.1], if \( M \geq c_8 \log N \), then
\[ \left\| \sum_{m=1}^{M} \mathbf{Z}_{nm} \right\| \leq C_8 \sqrt{\frac{NN_1 \log N}{M^2}} \| \mathbf{A} \|_{2,\infty} \| \mathbf{B} \|_{2,\infty} \] holds with probability at least \( 1 - N^{-12} \) for some constant \( c_8 > 0 \).

**Lemma 6.** For all \( \mathbf{A} \in \mathbb{C}^{N_1 \times K}, \mathbf{B} \in \mathbb{C}^{N_2 \times K}, \mathbf{C} \in \mathbb{C}^{N_2 \times K} \) and \( \mathbf{D} \in \mathbb{C}^{N_2 \times K} \), we denote
\[ D \left( \mathbf{AC}^H, \mathbf{BD}^H \right) = \frac{1}{p} \mathbb{E} \langle \mathbf{G} \mathbf{P}_0 \mathbf{G}^H (\mathbf{AC})^H, \mathbf{BD}^H \rangle - \mathbb{E} \langle \mathbf{G} \mathbf{G}^H (\mathbf{AC})^H, \mathbf{BD}^H \rangle, \] then
\[ \left\| D \left( \mathbf{AC}^H, \mathbf{BD}^H \right) \right\| \leq C_8 \sqrt{\frac{NN_1 \log N}{M^2}} \min \left\{ \| \mathbf{A} \|_{2,\infty} \| \mathbf{C} \|, \| \mathbf{A} \| \| \mathbf{C} \|_{2,\infty} \right\} \cdot \min \left\{ \left\| \mathbf{D} \right\|_{2,\infty}, \left\| \mathbf{B} \right\|_{F}, \left\| \mathbf{D} \right\|_{2,\infty} \right\} \] holds if \( M \geq c_8 \log N \) for a universal constant \( C_8 > 0 \) with probability exceeding \( 1 - N^{-12} \).

**Proof.** Following from the definition of \( D \left( \mathbf{AC}^H, \mathbf{BD}^H \right) \) in (49), we have
\[ \| D \left( \mathbf{AC}^H, \mathbf{BD}^H \right) \| \leq \mathbb{E} \left[ \left\| \left( \frac{1}{p} \mathbf{G} \mathbf{P}_0 \mathbf{G}^H - \mathbf{G} \mathbf{G}^H \right) (\mathbf{AC})^H, \mathbf{BD}^H \right\| \right] \leq \| \frac{1}{p} \mathbf{G} \mathbf{P}_0 \mathbf{G}^H - \mathbf{G} \mathbf{G}^H \| \left\| \mathbf{AC}^H, \mathbf{BD}^H \right\| _F \leq C_8 \sqrt{\frac{NN_1 \log N}{M^2}} \| \mathbf{A} \|_{2,\infty} \| \mathbf{C} \|_{2,\infty} \| \mathbf{B} \| _F \| \mathbf{D} \| _F, \] where the second inequality follows from Lemma 5. Due to
\[ D \left( \mathbf{AC}^H, \mathbf{BD}^H \right) = D \left( \mathbf{BH} \mathbf{A}, \mathbf{DH} \mathbf{C} \right) = D \left( \mathbf{CH} \mathbf{D}, \mathbf{AH} \mathbf{B} \right) = D \left( \mathbf{DH} \mathbf{AC}, \mathbf{H}^T \mathbf{BD} \right) = D \left( \mathbf{DB} \mathbf{H} \mathbf{AC}, \mathbf{H} \right), \] the result (50) is obtained.

**Lemma 7.** (8 Lemma 3). Let \( \mathcal{T} \) be the subspace defined as follows:
\[ \mathcal{T} = \left\{ \mathbf{F}_1 \mathbf{V}^H + \mathbf{U} \mathbf{F}^H_2 \middle| \mathbf{F}_1 \in \mathbb{C}^{N_1 \times K}, \mathbf{F}_2 \in \mathbb{C}^{N_2 \times K} \right\}. \] Then
\[ \left\| \frac{1}{p} \mathbf{P}_\mathcal{T} \mathbf{G} \mathbf{P}_0 \mathbf{G}^H \mathbf{P}_\mathcal{T} - \mathbf{P}_\mathcal{T} \mathbf{G} \mathbf{G}^H \mathbf{P}_\mathcal{T} \right\| \leq 0.1 \] holds with probability exceeding \( 1 - N^{-12} \) provided that \( M \geq C_0 \mu c_8 K \log N \) for an absolute constant \( C_8 \).
VI. NUMERICAL SIMULATIONS

In this section, we conduct numerical simulations to compare the performance difference between PGD [10] and VGD. The generation way of experiment data keeps almost the same as [10] for fairly. Specifically, we generate a spectrally sparse signal $x^*$ of length $N = 1024$ with $K = 5$ frequency components. Each frequency $f_k$ is randomly generated from $[0, 1)$ without a minimum separation condition, and the amplitude of each complex coefficient $s_k$ with random phase is selected to be $1 + 10^{-0.5}b_k$ with $b_k$ satisfying the uniform distribution in $[0, 1]$. As for the damping factors, each $1/\tau_k$ is uniformly distributed on $[N/4, N/2]$. Then $M = 128$ samples are selected uniformly at random from $x^*$ and random noise that follows a complex normal distribution is added. We set $N_1 = 512$ and $N_2 = 513$ in both PGD and VGD. Algorithms are terminated if $\|x^{t+1} - x^\star\|_2 / \|x^\star\|_2 \leq 10^{-6}$, or maximally 9999 iterations are reached. The root mean squared error (RMSE) of $x^t$ is defined as $\|x^t - x^\star\|_2 / \|x^\star\|$ and the normalized mean squared error (NMSE) is calculated by $\|x^\text{rec} - x^\star\|_2^2 / \|x^\star\|_2^2$. The step sizes of both VGD and PGD are fixed on $10^{-5}.2$ For PGD, the parameters $\beta$ used in the projection are estimated from the initialization shown in [11].

![Fig. 2. Performance comparison between VGD and PGD under additive noise (See color figures in online version).](image)

We plot in Figure 2 the NMSE of the reconstructed signals by PGD and VGD under different signal to noise ratio (SNR) value. Each data point represents the average over 1000 random instances and the error bars mean the fluctuation range of NMSE in these tests under fixed SNR. Firstly, the average NMSE in dB of VGD scales inversely proportional to the SNR, which is consistent with Theorem 1. Secondly, it can be found that the NMSE of PGD shows large fluctuation when SNR $\geq 20\text{dB}$, while the performance of NMSE of VGD keeps relatively stable. It means that PGD fails to recover the signal while VGD succeeds in some tests.

![Fig. 3. Successful signal recovery rate versus the sampling ratio $p$ over 100 Monte Carlo trials with $K = 5$ when SNR = 60dB.](image)

Furthermore, the successful signal recovery rate versus the sampling ratio $p$ over 100 randomly generated instances when SNR = 60dB is reported in Figure 3. Successful signal recovery is declared if NMSE is less than $10^{-6}$. As can be seen, the phase transition of VGD always performs better than PGD’s, and is slightly bad than PGD-TrueProjParam in which we assume the incoherence parameter $\mu$ and the largest singular value $\sigma_1$ of $Gy^\star$ are known and set the projection parameter $\beta = \sqrt{\mu K/\sigma_1}$. It is reasonable because extra prior information is input into PGD-TrueProjParam.

![Fig. 4. The RMSE of PGD and VGD with respect to iteration count under different constant step size.](image)

To study what happens when PGD shows performance loss with high SNR in Figure 2 and Figure 3, we turn to focus on those tests where PGD fails. Taking one failed test with SNR = 60dB for example, we give the RMSE of PGD and VGD with respect to iteration count under different constant step size $\eta$ in this test, shown in Figure 4. The RMSE of
PGD converges to around $10^{-1}$, which is much larger than the value about $10^{-3}$ that VGD converges to for both $\eta = 10^{-4}$ and $\eta = 10^{-5}$. The reason behind this phenomenon is that in this test the projection operation of PGD works, that is to say, some rows of the iterates $\{Z^t\}$ in PGD exceed the predetermined $\beta$ in (11), and then these rows are trimmed to control their $l_2$-norm. This trim operation breaks the normal gradient descent trajectory, so it immensely degrades the algorithm performance.

When the projection operation of PGD works, that is to say, some rows of the iterates $\{Z^t\}$ in PGD exceed the predetermined $\beta$ in (11), and then these rows are trimmed to control their $l_2$-norm. This trim operation breaks the normal gradient descent trajectory, so it immensely degrades the algorithm performance.

To confirm our analysis in Section III-B, we report the row $l_2,\infty$ norm of the iterates $\{Z^t\}$ versus iteration count for both PGD and VGD in Figure 5 and Figure 6. When the predetermined $\beta$ is larger than $\|M\|_{2,\infty}$, the row $l_2,\infty$ norm of the iterates in PGD behaves the same as that in VGD as expected and eventually converges to the $l_2,\infty$ norm of the true solution $M$. But when $\beta < \|M\|_{2,\infty}$, the row $l_2,\infty$ norm of the iterates of PGD tends to $\beta$, while that of VGD converges to $\|M\|_{2,\infty}$. It means that the projection operation in PGD preserves the iterates of PGD going ahead towards the global optimum, which results in PGD failing in signal recovery.

VII. CONCLUSION

In this paper, VGD algorithm without projection operation based on low-rank Hankel noisy matrix completion is proposed for robust spectral compressed sensing. We study the convergence of VGD and demonstrate the required sample complexity and noise condition for robust recovery of spectrally sparse signal. Moreover, the superiority of VGD compared with PGD is analyzed and is confirmed by numerical simulations.

Notice that the recent work in [31] has resolved a fundamental limitation of the existing low-rank Hankel model and has proposed a low-rank double Hankel model that pushing the spectral poles to the unit circle, which can bring improvement for line spectral estimation. Hence, a possible further research direction is to study the application and theory of the nonconvex VGD in solving the double Hankel model.

REFERENCES

[1] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, “Beyond nyquist: Efficient sampling of sparse bandlimited signals,” IEEE transactions on information theory, vol. 56, no. 1, pp. 520–544, 2009.

[2] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin, “Sparsity and compressed sensing in radar imaging,” Proceedings of the IEEE, vol. 98, no. 6, pp. 1006–1020, 2010.

[3] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse mri: The application of compressed sensing for rapid mr imaging.” Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 58, no. 6, pp. 1182–1195, 2007.

[4] M. F. Duarte and R. G. Baraniuk, “Spectral compressive sensing.” Applied and Computational Harmonic Analysis, vol. 35, no. 1, pp. 111–129, 2013.

[5] Z. Yang, J. Li, P. Stoica, and L. Xie, “Sparse methods for direction-of-arrival estimation,” in Academic Press Library in Signal Processing, Volume 7. Elsevier, 2018, pp. 509–581.

[6] Y. Chi, L. L. Scharf, A. P Jesuski, and A. R. Calderbank, “Sensitivity to basis mismatch in compressed sensing.” IEEE Transactions on Signal Processing, vol. 59, no. 5, pp. 2182–2195, 2011.

[7] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, “Compressed sensing off the grid.” IEEE transactions on information theory, vol. 59, no. 11, pp. 7465–7490, 2013.

[8] Y. Chen and Y. Chi, “Robust spectral compressed sensing via structured matrix completion,” IEEE Transactions on Information Theory, vol. 60, no. 10, pp. 6576–6601, 2014.

[9] L. Cai, T. Wang, and K. Wei, “Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank hankel matrix completion.” Applied and Computational Harmonic Analysis, vol. 46, no. 1, pp. 94–121, 2019.

[10] ———, “Spectral compressed sensing via projected gradient descent.” SIAM Journal on Optimization, vol. 28, no. 3, pp. 2625–2653, 2018.

[11] Z. Yang and L. Xie, “On gridless sparse methods for line spectral estimation from complete and incomplete data,” IEEE Transactions on Signal Processing, vol. 63, no. 12, pp. 3139–3153, 2015.

[12] C. Zhou, Y. Gu, X. Fan, Z. Shi, G. Mao, and Y. D. Zhang, “Direction-of-arrival estimation for coprime array via virtual array interpolation.” IEEE Transactions on Signal Processing, vol. 66, no. 22, pp. 5956–5971, 2018.

[13] M. A. Davenport and J. Romberg, “An overview of low-rank matrix recovery from incomplete observations.” IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 4, pp. 608–622, 2016.

[14] G. Tang, B. N. Bhaskar, and B. Recht, “Near minimax line spectral estimation.” IEEE Transactions on Information Theory, vol. 61, no. 1, pp. 499–512, 2014.
Proof of Lemma 7 Denote $E := \mathcal{H}_{\Omega}(\varepsilon)$. Similar to the proof route in [33, Lemma 11], we use a truncation argument to deal with the unbounded entries in $E$. Specifically, let $\tilde{E}$ be the matrix with $\tilde{E}_{ij} = E_{ij}$ when $|E_{ij}| \leq \gamma \sigma \sqrt{\log N_2}$ for some constant $\gamma \geq 0$ otherwise $\tilde{E}_{ij} = 0$. To apply Lemma 8 on $\tilde{E}$, we can observe that $\delta_\ast = \min_{ij} |\tilde{E}_{ij}| \leq \gamma \sigma \sqrt{\log N_2}$, $\delta_1 \leq \sigma \sqrt{N_2 \rho}$ and $\delta_2 \leq \sigma \sqrt{N_1 \rho}$ from the definition of $E$. Then by Lemma 8 with $t = \sqrt{14c_7^2 \sigma \log N_2}$ and $N_1 \geq N_2$, we have
\[
P \left( \| \tilde{E} \| \geq 6\sigma \sqrt{N_2 \rho} + \sqrt{14c_7^2 \sigma \log N_2} \right) \leq N_2^{-13}.
\]

What’s more, we choose $\gamma$ large enough and use the tail properties of sub-Gaussian random variable to bound the remaining entries that satisfy $E_{ij} > \gamma \sigma \sqrt{\log N_2}$. Specifically,
\[
P \left( \tilde{E} \neq E \right) \leq P \left( \max_{ij} |E_{ij}| > \gamma \sigma \sqrt{\log N_2} \right) \leq N_2^{-13}.
\]

Finally we have
\[
P \left( \| E \| \geq 6\sigma \sqrt{N_2 \rho} + \sqrt{14c_7^2 \sigma \log N_2} \right) \\
\leq P \left( \tilde{E} \neq E \right) + P \left( \| E \| \geq 6\sigma \sqrt{N_2 \rho} + \sqrt{14c_7^2 \sigma \log N_2} \right) \\
\leq N_2^{-12}.
\]
(54)

Suppose $M \geq c_9 c_\ast \log^2 N_2$ for a sufficiently large constant $c_9 > 0$, we obtain
\[
\left\| \frac{1}{p} \mathcal{H}_{\Omega}(\varepsilon) \right\| = \left\| \frac{1}{p} E \right\| \leq C_0 \sigma \sqrt{N_2 / c_9 M}
\]
with probability at least $1 - N_2^{-12}$ for a universal constant $C_0 > 0$.}

Lemma 8 (Extension of [32, Corollary 3.12] to the rectangular matrices, Assume $N_1 \geq N_2$, and let $X$ be an $N_1 \times N_2$ matrix whose entries $X_{ij}$ are independent symmetric random variables bounded by $\delta_\ast$. Then there exists a universal constant $c$ such that
\[
P \left( \| X \| \geq 3 \left( \delta_1 + \delta_2 \right) + t \right) \leq N_2 \exp \left( -\frac{t^2}{c_\delta^2} \right),
\]
(56)

where $\delta_1 := \max_i \sqrt{\sum_j E[X_{ij}^2]}$, $\delta_2 := \max_j \sqrt{\sum_i E[X_{ij}^2]}$ and $\delta_\ast := \max_{ij} |X_{ij}|_{\infty}$.

Proof of Lemma 8 The proof is essentially identical to [32, Corollary 3.12], we omit it here for simplicity.

Lemma 9 For $A, B \in \mathbb{C}^{N \times K}$, $Q^{\ast} = \arg \min_{Q \in G(K)} \| A Q - B \|_F$ holds if and only if $(Q^{\ast})^H A^H B$ is conjugate symmetric and positive semidefinite.

Proof of Lemma 9 The proof is completed by extending the results in [34, Theorem 2] onto complex matrices.
Lemma 10 ([35] Theorem 3.2). Let \( C \in \mathbb{C}^{K \times K} \) be nonsingular. Then for any matrix \( E \leq \mathbb{C}^{K \times K} \) with \( \sigma_1 (E) \leq \sigma_1 (C) \) and any unitarily invariant norm \( \| \cdot \| \), we have
\[
\| \text{sgn} (C + E) - \text{sgn} (C) \| \leq \frac{2 \| E \|}{\sigma_K (C)}.
\]  

(57)

Lemma 11 (Extension of [17] Lemma 37) to complex matrices. Suppose \( X_1, X_2 \in \mathbb{C}^{(N_1+N_2) \times K} \) are two matrices such that
\[
\| X_1 - M \| \| M \| \leq \frac{\sigma_K}{2},
\]
\[
\| X_1 - X_2 \| \| M \| \leq \frac{1}{2} \sigma_K
\]
then the following inequalities
\[
\| X_1 Q_1 - X_2 Q_2 \| \leq 5 \kappa \| X_1 - X_2 \| ,
\]
\[
\| X_1 Q_1 - X_2 Q_2 \|_F \leq 5 \kappa \| X_1 - X_2 \|_F
\]
hold where \( Q_1 := \text{arg min}_{Q \in O(K)} \| X_1 Q - M \|_F \) and \( Q_2 := \text{arg min}_{Q \in O(K)} \| X_2 Q - M \|_F \).

APPENDIX B

Proof of Proposition 1

Proof of Proposition 1. Suppose that \( \nabla^2 f^* (Z) \) is the Hessian without noise of \( f (Z) \). Then for any \( W \in \mathbb{C}^{(N_1+N_2) \times K} \), we have
\[
\mathbb{E} \left\{ \text{vec} (W)^H \nabla^2 f^* (Z) \text{vec} (W) \right\} = 2 \mathbb{E} \left\{ \langle I - GG^H \rangle \left( Z U Z^H V + Z U W W^H V \right) \right\}
\]
\[
+ \| \langle I - GG^H \rangle \left( W U Z^H W + Z U W W^H V \right) \|_F^2
\]
\[
+ \frac{1}{2} \mathbb{E} \langle G P_{\Omega} G^H \rangle \left( Z U Z^H W + Z U W W^H V \right) \|_F^2
\]
\[
+ \frac{1}{2} \mathbb{E} \langle G P_{\Omega} G^H \rangle \left( Z U Z^H W + Z U W W^H V \right) \|_F^2
\]
\[
+ \frac{1}{4} \| W U Z U + Z U W U - Z V W V - W V Z V \|_F^2.
\]
(58)

Denote \( \Delta = \left[ \begin{array}{c} \Delta_U \\ \Delta_V \end{array} \right] \), \( \Delta_U := Z U - M U, \Delta_V := Z V - M V \), \( \Theta = \left[ \begin{array}{c} \Theta_U \\ \Theta_V \end{array} \right] \), \( \Theta_U := M U - C U \) and \( \Theta_V := M V - C V \), and we consider the expectation of
\[
\mathbb{E} \left\{ \text{vec} (W)^H \nabla^2 f^* (Z) \text{vec} (W) \right\},
\]
amely the population level, firstly.
\[
\mathbb{E} \left\{ \text{vec} (W)^H \nabla^2 f^* (Z) \text{vec} (W) \right\} = 2 \mathbb{E} \langle Z U Z^H W - g y^* , W U W^H V \rangle + \| W U Z^H W + Z U W W^H V \|_F^2
\]
\[
+ \frac{1}{2} \mathbb{E} \langle Z U Z^H W - g y^* , W U W^H V \rangle + \| W U Z^H W + Z U W W^H V \|_F^2
\]
\[
+ \frac{1}{4} \| W U Z U + Z U W U - Z V W V - W V Z V \|_F^2.
\]
(59)

where the equality follows from \( \mathbb{E} \left\{ \frac{1}{2} P_{\Omega} \right\} = I, GG^H \) is a projection operator that satisfies \( GG^H (gy^*) = gy^* \). According to [18] Appendix A that mainly uses \( M U M U = M U M V \) and the conjugate symmetry of \( Q^H (I - B M V) C \) by Lemma 6 we have
\[
\mathbb{E} \left\{ \langle Z U Z^H W - g y^* , W U W^H V \rangle + \| W U Z^H W + Z U W W^H V \|_F^2 \right\}
\]
\[
+ \| M U W^H V \|_F^2 + \| M U M V \|_F^2 + C^H (I - B M V) C \|
\]
(60)

where
\[
| \mathcal{L} |
\]
\[
\leq 9 \left( \| \Theta_U \| + \| \Theta_V \| \right) \left( \| C_U \| + \| C_V \| \right) \left( \| \Theta_U \| + \| \Theta_V \| \right)^2
\]
\[
+ \left( | \Delta_U | + | \Delta_V | \right) \left( \| M U \| + \| M V \| \right) \left( | \Delta_U | + | \Delta_V | \right)^2
\]
\[
\cdot \left( \| M U \| + \| M V \| \right) + \| C \| (\| \Theta \| + \| M \|).
\]

From the assumption, we have \( | \Theta_U | \leq \frac{1}{10} \sqrt{\sigma_1} \), \( | \Delta \|_\infty \leq \frac{1}{100} \sqrt{\sigma_1} \) and \( | \Delta \|_F \leq \frac{1}{100} \sqrt{\sigma_1} \). Moreover, we have \( \| M U \| \leq \| M V \| \leq \sqrt{\sigma_1} \) and \( | C \| \leq | \Theta | + | M | \). So we obtain
\[
| \mathcal{L} | \leq 9 \left( \frac{3}{400 \sigma_1} + \frac{1}{100} \right) \sigma_K \| W \|_F^2 \leq \frac{1}{100} \sigma_K \| W \|_F^2.
\]

(61)

Then we focus on the difference between the empirical level and the population level,
\[
\mathbb{E} \left\{ \text{vec} (W)^H \nabla^2 f^* (Z) \text{vec} (W) \right\}
\]
\[
- \mathbb{E} \left\{ \text{vec} (W)^H \nabla^2 f^* (Z) \text{vec} (W) \right\}
\]
\[
= \mathbb{E} \left\{ \langle G P_{\Omega} G^H \rangle \left( Z U Z^H W - M U M V \right) , W U W^H V \rangle
\]
\[
+ \frac{1}{p} \mathbb{E} \langle G P_{\Omega} G^H \rangle \left( W U Z^H W + Z U W W^H V \right) \|_F^2
\]
\[
- \frac{1}{2} \mathbb{E} \langle G^H \rangle \left( Z U Z^H W - M U M V \right) , W U W^H V \rangle
\]
\[
- \| g g^H \left( W U Z^H W + Z U W W^H V \right) \|_F^2
\]
\[
= S_1 + S_2 + S_3 + S_4.
\]

(59)

where
\[
S_1 = 2D \left( M U \Delta_U^H , W U W^H V \right) + D \left( \Delta_U M U , W U W^H V \right)
\]
\[
+ 2D \left( \Delta_U \Delta_U^H , W U W^H V \right) + 2D \left( W U M U , \Delta U W^H V \right)
\]
\[
+ 2D \left( W U \Delta_U^H , M U M V \right) + 2D \left( \Delta U W^H V , \Delta U W^H V \right),
\]
\[
S_2 = D \left( W U M U^H , W U M V \right) + D \left( M U W^H V , M U W^H V \right)
\]
\[
+ 2D \left( W U M U^H , M U M V \right) ,
\]
\[
S_3 = D \left( W U \Delta_U^H , W U \Delta_U^H \right) + D \left( \Delta U W^H V , \Delta U W^H V \right) ,
\]
\[
S_4 = 2D \left( W U M U^H , W U M V \right) + 2D \left( M U W^H V , \Delta U W^H V \right)
\]
and \( D (A C^H, B D^H) \) has defined in Lemma 6.
Each item in the first part $S_1$ all includes $W_U$ and $W_V$, so that we can control $S_1$ by Lemma 6 involving $\|W\|_F^2$. The second part $S_2$ contains those items in the subspace $L$ defined in Lemma 7. What's more, $S_2$ consists of those components that have two same variables for $D(\cdot, \cdot)$ in the remaining items and can be bounded by Lemma 6. Finally, $S_3$ will be controlled with the help of Lemma 6 and Lemma 7.

Specifically, for $S_1$, we have by invoking Lemma 6

$$|S_1| \leq 2CN\sqrt{\frac{NN_1 \log N}{M}} \left(2\|M_U\|_{2, \infty} \|\Delta V\|_{2, \infty} + 2\|M_V\|_{2, \infty}\right) \|W\|_F \|W\|_F.$$  

With the assumption of $\|\Delta\|_{2, \infty}$ in (19) and $\|M\|_{2, \infty} \leq \sqrt{\frac{\mu c_k}{N}} \sqrt{\sigma_1}$, we have

$$|S_1| \leq \frac{CS}{400} \|W\|_F^2,$$  

provided that $M \geq \mu c_k K \log N$.

For $S_2$, if $M \geq C_0 \mu c_k K \log N$, we have

$$|S_2| = \left|D \left(W_U M_V^H + M_U W_V^H, W_U M_V^H + M_U W_V^H\right)\right| \leq 0.1 \|W\|_F \|W\|_F,$$  

where the inequality follows from Lemma 7.

For $S_3$, we have

$$|S_3| \leq C \sqrt{\frac{NN_1 \log N}{M}} \left(\|\Delta V\|_{2, \infty} \|W\|_F + \|\Delta U\|_{2, \infty} \|W\|_F\right) \|W\|_F \leq \frac{CS}{800} \|W\|_F^2$$  

provided $M \geq \log N$.

For $S_4$, we have

$$|S_4| \leq \frac{C}{200} \sigma_k \|W\|_F^2.$$

provided $M \geq \log N$.

Finally, we prove $S_5$.

Specifically, for $S_5$, we have

$$|S_5| \leq C \sqrt{\frac{NN_1 \log N}{M}} \left(\|\Delta V\|_{2, \infty} \|W\|_F + \|\Delta U\|_{2, \infty} \|W\|_F\right) \|W\|_F \leq \frac{CS}{800} \|W\|_F^2$$  

provided $M \geq \log N$.

The proof is finished.

### Appendix C

**Proof of Proposition 2.**

Recall the definition of matrices in Lemma 4, it yields that $G^T$ has eigenvalue decomposition:

$$G^T = \frac{1}{\sqrt{2}} \begin{bmatrix} U & U \\ V & -V \end{bmatrix} \Sigma \begin{bmatrix} 0 & \Sigma^H \\ 0 & -\Sigma \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} U \\ V \end{bmatrix}^H.$$  

Its eigenvalues satisfy $\lambda_1(G^T) = \sigma_1(G^T)$, $\cdots$, $\lambda_K(G^T) = \sigma_K(G^T)$. Let $G^T = \begin{bmatrix} G^T_1 & \cdots & G^T_K \end{bmatrix}$, then

$$\lambda_{N_1+1}(G^T) = \cdots = \lambda_{N_1+N_2-K}(G^T) = -\sigma_1(G^T)$$

where

$$\lambda_{N_1+1}(G^T) = \cdots = \lambda_{N_1+N_2-K}(G^T) = -\sigma_1(G^T).$$

Moreover, $G^T$ has the following top-$K$ eigenvalue decomposition

$$\frac{1}{\sqrt{2}} \begin{bmatrix} U \\ V \end{bmatrix} \Sigma \frac{1}{\sqrt{2}} \begin{bmatrix} U \end{bmatrix}^H.$$  

(69)
due to $G^* = \mathcal{G}y^* = USV^H$. Similar result holds for $\overline{G}^0$ and $
olinebreak \overline{G}^{0,(l)}$.

According to Lemma 3 and Lemma 4 if $M \geq 6400C_1^2h_c^2K^5 log N$ and $\sigma \sqrt{\frac{N^2}{c_M}} \leq \frac{1}{20c_M} \sqrt{1/\kappa}$ hold, we have

$$\|G^{0,(l)} - \overline{G}^0\| \leq \|G^0 - \overline{G}^0\|$$

$$\leq \left( C_1 \sqrt{\frac{\mu_c K^3 log N}{M}} + \frac{C_2 \sigma}{\sigma_K} \sqrt{\frac{\alpha^2 N^2}{c_M}} \right) \|\mathcal{G}y\|$$

$$\leq \frac{1}{40 \sqrt{\kappa}} \sigma_K \leq \frac{1}{4} \sigma_K. \quad (70)$$

By Weyl’s inequality, we further have

$$\frac{3}{4} \sigma_{\text{K}} \leq \sigma_{\text{K}} (\Sigma^0) \leq \sigma_1 (\Sigma^0) \leq 2\sigma_1, \quad (71)$$

$$\frac{3}{4} \sigma_{\text{K}} \leq \sigma_{\text{K}} (\Sigma^{0,(l)}) \leq \sigma_1 (\Sigma^{0,(l)}) \leq 2\sigma_1, \quad (72)$$

where $\Sigma^0$ and $\Sigma^{0,(l)}$ are all $K \times K$ diagonal matrices that contain the top $K$ singular value of $G^0$ and $G^{0,(l)}$ respectively.

Denote $\tilde{M} = \begin{bmatrix} U \\ V \end{bmatrix}$, $\tilde{Z} = \begin{bmatrix} U^0 \\ V^0 \end{bmatrix}$ and $\tilde{Z}^{0,(l)} = \begin{bmatrix} U^{0,(l)} \\ V^{0,(l)} \end{bmatrix}$.

Besides, define $Q^0 := \min_{Q \in O(K)} \|\tilde{Z}^0 Q - \tilde{M}\|_F$, $Q^{0,(l)} := \min_{Q \in O(K)} \|\tilde{Z}^{0,(l)} Q - \tilde{M}\|_F$, $P := \min_{Q \in O(K)} \|\tilde{Z}^{0,(l)} Q - \tilde{Z}^0\|_F$. \quad (73)

1) To bound $\|Z^0 Q^0 - M\|$ in (27), we need the following decomposition

$$\|Z^0 Q^0 - M\|
= \|\tilde{Z}^0 (\Sigma^0)^{1/2} (Q^0 - \tilde{Q}) + \tilde{Z}^0 (\Sigma^0)^{1/2} Q - \tilde{Q} \Sigma^{1/2}\|
+ \left( \tilde{Z}^0 Q^0 - \tilde{M} \right) \Sigma^{1/2}
\leq \left( \left\|\Sigma^0 \right\|^{1/2} \right) \left\|Q^0 - \tilde{Q}\right\| + \left( \left\|\Sigma^0 \right\|^{1/2} \right) \left\|Q - \tilde{Q}\right\|^{1/2}
+ \left( \tilde{Z}^0 Q^0 - \tilde{M} \right) \left\|\Sigma^{1/2}\right\|.
\quad (76)$$

By applying Lemma 47, Lemma 46 and Lemma 45 with the above results (68)-(71), we obtain

$$\left\|Q^0 - \tilde{Q}\right\| \leq \frac{15 \kappa^{3/2}}{\sigma_K} \left\|G^0 - \overline{G}^0\right\|, \quad (77)$$

$$\left\|\Sigma^0 \right\|^{1/2} \left\|Q^0 - \tilde{Q}\right\|^{1/2} \leq \frac{15 \kappa}{\sqrt{\sigma_K}} \left\|G^0 - \overline{G}^0\right\|, \quad (78)$$

$$\left\|\tilde{Z}^0 Q^0 - \tilde{M}\right\| \leq \frac{3}{\sigma_K} \left\|G^0 - \overline{G}^0\right\|. \quad (79)$$

Taken them together and apply Lemma 3 Lemma 4 (70), (71) and (72), we have

$$\|Z^0 Q^0 - M\|
\leq 90 \left( C_1 \sqrt{\frac{\mu_c K^3 log N}{M}} + \frac{C_2 \sigma}{\sigma_K} \sqrt{\frac{\alpha^2 N^2}{c_M}} \right) \sqrt{\sigma_1}. \quad (80)$$

2) The second part is to bound $\left\|\left( \tilde{Z}^{0,(l)} Q^{0,(l)} - M \right)_{l'} \right\|_2$ in (28). Notice that the known key assumption is the incoherence property of $G^*$ in Definition 1 so the proof focuses on transferring the $l_2$-norm of the $l'$ row onto the matrix $\overline{G}^0$.

$$\left\|\left( \tilde{Z}^{0,(l)} Q^{0,(l)} - M \right)_{l'} \right\|_2
\leq \left\|\left( \tilde{Z}^{0,(l)} Q^{0,(l)} - \tilde{M} \right)_{l'} \right\|_2 + \left\|Z^{0,(l)} \left( Q^{0,(l)} - \tilde{Q}^{0,(l)} \right) \right\|_2.$$

For the first term, we have

$$\left\|\left( \tilde{Z}^{0,(l)} Q^{0,(l)} - \tilde{M} \right)_{l'} \right\|_2
= \left\|\left( \tilde{Z}^{0,(l)} Q^{0,(l)} \right)^{1/2} \left( Q^{0,(l)} - \tilde{Q}^{0,(l)} \right) \right\|_2
= \left\|\tilde{Z}^{0,(l)} \left( Q^{0,(l)} - \tilde{Q}^{0,(l)} \right) \right\|_2
\leq \left\|\tilde{M} \right\| \left\|\Sigma^{1/2} \right\| \left\|\tilde{Z}^{0,(l)} \tilde{M} \right\| \left\|\Sigma^{1/2} \right\|$$

$$\leq \frac{20 \kappa}{\sigma_K^{1/2}} \left\|\overline{G}^0 - G^{0,(l)}\right\| + \frac{3}{\sigma_K^{1/2}} \left\|G^0 - \overline{G}^0\right\|$$

where the first equality can be verified by using the corresponding matrix definition, the second inequality follows from Lemma 2, Lemma 46, [17] Lemma 45 and (72), and the final inequality uses (70). \quad (81)

For the second term, by using (81), we have

$$\left\|Z^{0,(l)}_{l'} \right\|_2 \leq \left\|Z^{0,(l)}_{l'} \tilde{Q}^{0,(l)} - M_{l'} \right\|_2 + \left\|M_{l'} \right\|_2$$

$$\leq 23 \left( C_1 \sqrt{\frac{\mu_c K^3 log N}{M}} + \frac{C_2 \sigma}{\sigma_K} \sqrt{\frac{\mu_c K^5 N}{M}} \right) \sqrt{\sigma_1}. \quad (82)$$

Then, we obtain

$$\left\|Z^{0,(l)} \left( Q^{0,(l)} - \tilde{Q}^{0,(l)} \right) \right\|_2 \leq \left\|Z^{0,(l)}_{l'} \right\|_2 \left\|Q^{0,(l)} - \tilde{Q}^{0,(l)} \right\|_2$$

$$\leq \left\|Z^{0,(l)}_{l'} \right\|_2 \cdot \frac{15 \kappa^{3/2}}{\sigma_K} \left\|G^0 - \overline{G}^{0,(l)}\right\|$$

$$\leq \frac{23 \left( C_1 \sqrt{\frac{\mu_c K^3 log N}{M}} + \frac{C_2 \sigma}{\sigma_K} \sqrt{\frac{\mu_c K^5 N}{M}} \right)}{\sqrt{\sigma_1}} \cdot \frac{15 \kappa^{3/2}}{\sigma_K}. \quad (83)$$

\quad (83)
\[ \sqrt{\sigma_1} \cdot 15 \frac{K^{3/2}}{\sigma_K} \left( C_1 \sqrt{\frac{\mu c_K \log N}{M}} + C_2 \frac{\sigma}{\sigma_K} \sqrt{\frac{N}{c_M}} \right) \left\| G^* y^* \right\| \]
\[ \leq 45 \left( C_1 \sqrt{\frac{\mu c_k K^2 \log N}{MN}} + C_2 \frac{\sigma}{\sigma_K} \sqrt{\frac{\mu c_K N}{M}} \right) \sqrt{\sigma_1}, \quad (83) \]
where the second inequality follows from [17] Lemma 47, and the third inequality follows from (82) and (70), and the final ones hold when \( M \geq 23^3 C_2^2 \mu c_k K \log N \) and \( \frac{\sigma}{\sigma_K} \sqrt{\frac{N}{c_M}} \leq \frac{1}{2} \binom{N}{c_M} \).

Taking (81) and (83) together we have
\[ \left\| \left( Z^{0,(i)} Q^{0,(i)} - M \right)_{l, l} \right\|_2 \]
\[ \leq 68 \left( C_1 \sqrt{\frac{\mu c_k K^2 \log N}{MN}} + C_2 \frac{\sigma}{\sigma_K} \sqrt{\frac{\mu c_K N}{M}} \right) \sqrt{\sigma_1} \]
for \( 0 \leq l \leq N_1 - 1 \). The proof is all the same for \( N_1 \leq l \leq N_1 + N_2 - 1 \).

3) Then we need to bound \( \left\| Z^{0,(i)} Q^{0,(i)} - Z^{0,(i)} T^{0,(i)} \right\|_F \) in (29). Recall the definition of \( P \) in (75) and \( T^{0,(i)} \) in (26), we have
\[ \left\| Z^{0,(i)} Q^{0,(i)} - Z^{0,(i)} T^{0,(i)} \right\|_F \leq \left\| Z^{0,(i)} - \bar{Z}^{0,(i)} \right\|_F \left\| P \left( \bar{Z}^{0,(i)} \right)^{1/2} - \left( \bar{Z}^{0,(i)} \right)^{1/2} P \right\|_F \]
\[ + \left\| \bar{Z}^{0,(i)} - \bar{Z}^{0,(i)} P \right\|_F \left\| \left( \bar{Z}^{0,(i)} \right)^{1/2} - \left( \bar{Z}^{0,(i)} \right)^{1/2} \right\|_F \]
\[ \leq \frac{4 \sqrt{\sigma_1}}{\sigma_K} \left\| \left( \bar{G}^{0,(i)} - \bar{G}^{0,(i)} \right) \bar{Z}^{0,(i)} \right\|_F \leq \frac{19 \kappa}{\sqrt{\sigma_K}} \left\| \left( \bar{G}^{0,(i)} - \bar{G}^{0,(i)} \right) \bar{Z}^{0,(i)} \right\|_F, \quad (84) \]
where the second inequality follows from (71) and triangle inequality, and the third inequality follows from Davis-Kahan SinΘ theorem [23] [29] and [17] Lemma 46].

Therefore, what’s left is to give the upper bound of \( \left\| \left( \bar{G}^{0,(i)} - \bar{G}^{0,(i)} \right) \bar{Z}^{0,(i)} \right\|_F \). Without loss of generality, we assume \( 0 \leq l \leq N_1 - 1 \). Denote
\[ R_1 := \sum_{i=0}^{N_2-1} \left( \frac{1}{p} \delta_{l,i} - 1 \right) \left( \bar{G} \right)_{l,N_{1+i}, i} \left( \bar{Z} \right)_{N_{1+i}, i}, \]
\[ \sum_{i=0}^{N_2-1} \frac{1}{p} \delta_{l,i} \left( \bar{Z} \right)_{l,N_{1+i}, 0} \left( \bar{Z} \right)_{N_{1+i}, i}, \in C^{1 \times K}, \quad (85) \]
\[ R_2 := \sum_{i=0}^{N_2-1} \left( \frac{1}{p} \delta_{l,i} - 1 \right) \left( \bar{G} \right)_{N_{1+i}, l, i} \left( \bar{Z} \right)_{N_{1+i}, l}, \]
\[ \sum_{i=0}^{N_2-1} \frac{1}{p} \delta_{l,i} \left( \bar{Z} \right)_{N_{1+i}, l, 0} \left( \bar{Z} \right)_{N_{1+i}, l}, \]
\[ \quad (86) \]
where \( \delta_{l,i} = 1 \) if \( l + i \in \Omega \), otherwise \( \delta_{l,i} = 0 \). \( e_i \) is the standard basis of \( \mathbb{R}^{N_2} \) and \( \frac{\bar{H} e}{\sigma} := \left[ \begin{array}{c} 0 \\ \frac{H e}{\bar{H}} \end{array} \right] \). Recall the definition of \( \bar{G} \) and \( \bar{G}^{0,(i)} \), we have
\[ \left\| \left( \bar{G}^{0,(i)} - \bar{G}^{0,(i)} \right) \bar{Z}^{0,(i)} \right\|_F \leq 2 \left\| R_1 \right\|_2 + \left\| R_2 \right\|_2 \left\| \bar{Z}^{0,(i)} \right\|_2. \quad (87) \]

We try to control \( R_1 \) firstly. Thanks to the apply of leave-one-out technique that makes \( \bar{Z}^{0,(i)}_{l,N_{1+i}, i} \), independent with \( \delta_{l,i} \), so the finite sequences \( \left\{ r_{1,i}^{(1)} \right\} \) and \( \left\{ r_{1,i}^{(2)} \right\} \) have statistical independence, and \( E \left( r_{1,i}^{(1)} \right) = 0 \). Firstly for \( r_{1,i}^{(1)} \), by invoking Lemma [2] we have
\[ \left\| r_{1,i}^{(1)} \right\|_2 \leq \frac{1}{p} \left\| \bar{G} \right\|_F \left\| \bar{Z}^{0,(i)} \right\|_2 \leq \frac{\mu c_K}{p N} \sigma_1 \left\| \bar{Z}^{0,(i)} \right\|_2, \]
and
\[ E \left( \sum_{i=0}^{N_2-1} \left\| r_{1,i}^{(1)} \right\|_2 \right) \leq \frac{1}{p} \left\| \bar{G} \right\|_F^2 \left\| \bar{Z}^{0,(i)} \right\|_2^2 \leq \frac{\mu c_K}{p N} \sigma_1^2 \left\| \bar{Z}^{0,(i)} \right\|_2^2. \]

Then by matrix Bernstein inequality [27 Theorem 6.1.1], we can show that there exists a constant \( C_3 > 0 \) such that
\[ \left\| r_{1,i}^{(2)} \right\|_2 \leq \frac{1}{p} \left\| \delta_{l,i} \left( \bar{H} e \right)_{l,N_{1+i}, i} \right\| \left\| \bar{Z}^{0,(i)} \right\|_2 \leq \frac{1}{p} c_1 \sigma \left\| \bar{Z}^{0,(i)} \right\|_2, \quad (88) \]
for a constant \( c_1 > 0 \) where \( \left\| \cdot \right\|_2 \) is the sub-exponential norm. Moreover,
\[ v_1 := \left\| E \left[ \sum_{i=0}^{N_2-1} r_{1,i}^{(2)} r_{1,i}^{(2) H} \right] \right\|_F \]
\[ \leq c_2 \sigma^2 \frac{2}{p} \left\| \bar{Z}^{0,(i)} \right\|_2^2 \leq \frac{c_2 N^2 \sigma^2}{p} \left\| \bar{Z}^{0,(i)} \right\|_2^2 \]
holds for a constant \( c_2 > 0 \). Similar results hold for \( v_2 := \left\| E \left[ \sum_{i=0}^{N_2-1} r_{1,i}^{(2)} r_{1,i}^{(2) H} \right] \right\|_F \). According to matrix Bernstein inequality in [27 Theorem 6.1.1], we obtain that with probability exceeding \( 1 - N^{-11} \) for a universal constant \( C_{32} > 0 \),
\[ \left\| \sum_{i=0}^{N_2-1} r_{1,i}^{(2)} \right\|_2 \leq \sqrt{\max \{ v_1, v_2 \} \log N} + \left\| r_{1,i}^{(2)} \right\|_2 \log N \]
\[ \leq c_2 \sigma \frac{\sqrt{N \log N}}{p} \left\| \bar{Z}^{0,(i)} \right\|_2 + c_1 \sigma \frac{1}{p} \log N \left\| \bar{Z}^{0,(i)} \right\|_2. \]
where the third inequality holds if \( M \geq \log N \). By combining (83) and (89), we have the bound
\[
\|{\mathcal R}_1\|_2 \leq C_{31}\left( \sqrt{\frac{\mu c_k \log N}{M}} + \frac{\mu c_k \log N}{M} \right) \sigma_1
\]
\[\quad + C_{32} \sqrt{\frac{N^2 \log N}{M}} \|\tilde{Z}^{0,(l)}\|_{2,\infty}, \tag{90}\]
with probability at least \( 1 - N^{-10} \).

Similarly, we can bound \( \|{\mathcal R}_2\|_2 \). We omit the process for simplicity and finally obtain
\[
\|{\mathcal R}_2\|_2 \leq C_{41}\left( \sqrt{\frac{\mu c_k \log N}{M}} + \frac{\mu c_k \log N}{M} \right) \sigma_1
\]
\[\quad + C_{42} \sqrt{\frac{N^2 \log N}{M}} \|\tilde{Z}^{0,(l)}\|_{2,\infty}, \tag{91}\]
with probability exceeding \( 1 - N^{-10} \) for some universal constants \( C_{41} > 0 \) and \( C_{42} > 0 \).

So the remaining thing is to bound \( \|\tilde{Z}^{0,(l)}\|_{2,\infty} \). Following the route in [18 Claim 19] and [17 Claim (187)], we give the proof briefly. Firstly we decompose \( \|\tilde{Z}^{0,(l)}\|_{2,\infty} \) as follows
\[
\|\tilde{Z}^{0,(l)}\|_{2,\infty} \leq \|\tilde{Z}^{0,(l)}\|_{2,\infty} \, \text{sgn} \left( \frac{\tilde{Z}^{0,(l)}}{\tilde{Z}^{0,(l),\text{zero}}} \right) + \|\tilde{Z}^{0,(l),\text{zero}}\|_{2,\infty}, \tag{92}\]
where \( \tilde{Z}^{0,(l),\text{zero}} \) is the leading \( K \) eigenvectors of \( \tilde{G}^{0,(l)} \) that is the matrix gotten by zeroing out the \( l \)-th row and column of \( \tilde{G}^{0,(l)} \). For the first part, with the help of [20 Lemma 4 and Lemma 14] we have
\[
\|\tilde{Z}^{0,(l),\text{zero}}\|_{2,\infty} \leq 4 \|\tilde{M}\|_{2,\infty} + 8\kappa \|\tilde{Z}^{0}\|_{2,\infty}
\]
\[\leq 4 \|\tilde{M}\|_{2,\infty} + 8\kappa C_5 \left( \sqrt{\sum_{i=0}^{K-1} \left( \frac{\tilde{G}^{i}}{\|\tilde{G}^{i}\|_{2,\infty}} \right)^{2}} \right) \]
\[\leq 4 \left( 4 + 8\kappa C_5^2 \right) \sqrt{\frac{\mu c_k K}{N}} + 8\kappa C_5 \sqrt{\frac{\mu^2 c_k^2 K^2 \kappa^2}{pN^2}} \tag{93}\]
where the third inequality follows from Lemma 3 and 5, and the last inequality holds if \( M \geq \mu c_k^2 C_k^2 K^2 \). Moreover, by Davis-Kahan Sin\(\Theta\) theorem and Weyl's inequality, we have
\[
\|\tilde{Z}^{0,(l)}\|_{2,\infty} \leq \frac{2\sqrt{2}}{\sigma_K} \sqrt{\frac{\|\tilde{G}^{i}\|_{2,\infty}}{2}}, \tag{94}\]
Combining the above estimates (93) and (94) yields
\[
\|\tilde{Z}^{0,(l)}\|_{2,\infty} \leq \left( 4 + 2\sqrt{2} + 12C_5^2 \right) \sqrt{\frac{\mu c_k K}{N}} \tag{95}\]
for a constant \( C_5 > 0 \).

Eventually, putting (83), (87), (90), (91) and (95) together and if \( M \geq \mu c_k \log N \), we have
\[
\|Z^0 Q^0 - Z^{0,(l)} T^{0,(l)}\|_F \leq 912 \left( C_6 \sqrt{\frac{\mu c_k^2 K^2 \log N}{M}} + \frac{C_7 \sigma}{\sigma_K} \sqrt{\frac{\mu^2 c_k^2 K N \log N}{M}} \right) \tag{96}\]
with probability exceeding \( 1 - N^{-9} \) with some absolute constants \( C_6 > 0 \) and \( C_7 > 0 \) for \( 0 \leq l \leq N_1 - 1 \). When \( N_1 \leq l \leq N_1 + N_2 - 1 \), the same conclusion can be obtained similarly.

4) Finally, we need bound \( \|Z^0 Q^0 - M\|_{2,\infty} \) in (30). Because the whole process is similar to the proof of (37) in Proposition 3, we omit it for simplicity.

\[\text{APPENDIX D} \]

\[\text{PROOF OF PROPOSITION 3} \]

\[\text{Proof of Proposition 3} \] 1) For proving \( \|Z^{t+1} Q^{t+1} - M\|_2 \), we need use (34) and (37). We construct an auxiliary iteration defined as follows:
\[
\tilde{Z}^{t+1}_{U} := Z_{U}^{t+1} - \eta \left( \mathbb{I} - \tilde{G}^{H} \right) \left( Z_{U}^{t+1} (Z_{V}^{t+1})^{H} \right) M_{V}
\]
\[\quad + \frac{\eta^2}{p} P_{t} \tilde{G}^{H} \left( Z_{U}^{t+1} (Z_{V}^{t+1})^{H} - \tilde{G} \right) M_{V}
\]
\[\quad + \frac{\eta}{2} M_{U} (Q^{t})^{H} (Z_{U}^{t+1} (Z_{V}^{t+1})^{H} - Z_{V}^{t}) Z_{U}^{t} - (Z_{V}^{t})^{H} Z_{V}^{t} Q^{t}, \]
\[
\tilde{Z}^{t+1}_{V} := Z_{V}^{t+1} - \eta \left( \mathbb{I} - \tilde{G}^{H} \right) \left( Z_{U}^{t+1} (Z_{V}^{t+1})^{H} \right) M_{U}
\]
\[\quad + \frac{\eta^2}{p} P_{t} \tilde{G}^{H} \left( Z_{U}^{t+1} (Z_{V}^{t+1})^{H} - \tilde{G} \right) M_{U}
\]
\[\quad + \frac{\eta}{2} M_{V} (Q^{t})^{H} (Z_{V}^{t+1} (Z_{U}^{t+1})^{H} - Z_{V}^{t}) Z_{V}^{t} - (Z_{U}^{t})^{H} Z_{U}^{t} Q^{t}. \]
Moreover, denote
\[
\tilde{Z}^{t+1}_{U} := Z_{U}^{t+1} - \eta \left( \mathbb{I} - \tilde{G}^{H} \right) \left( Z_{U}^{t+1} (Z_{V}^{t+1})^{H} \right) M_{V}
\]
\[\quad - \eta \tilde{G}^{H} \left( Z_{U}^{t+1} (Z_{V}^{t+1})^{H} - M_{U} M_{V}^{H} \right) M_{V}
\]
\[\quad - \frac{\eta}{2} M_{U} (Q^{t})^{H} (Z_{U}^{t+1} (Z_{V}^{t+1})^{H} - Z_{V}^{t}) Z_{U}^{t} - (Z_{V}^{t})^{H} Z_{V}^{t} Q^{t}, \]
\[
\tilde{Z}^{t+1}_{V} := Z_{V}^{t+1} - \eta \left( \mathbb{I} - \tilde{G}^{H} \right) \left( Z_{U}^{t+1} (Z_{V}^{t+1})^{H} \right) M_{U}
\]
\[\quad - \eta \tilde{G}^{H} \left( Z_{U}^{t+1} (Z_{V}^{t+1})^{H} - M_{U} M_{V}^{H} \right) M_{U}
\]
\[\quad - \frac{\eta}{2} M_{V} (Q^{t})^{H} (Z_{V}^{t+1} (Z_{U}^{t+1})^{H} - Z_{V}^{t}) Z_{V}^{t} - (Z_{U}^{t})^{H} Z_{U}^{t} Q^{t}. \]
and \( \tilde{Z}^{t+1} = \left[ \tilde{Z}^{t+1}_{U} \right] \) as well as \( \mathcal{E} \tilde{Z}^{t+1} = \left[ \mathcal{E} \tilde{Z}^{t+1}_U \right] \). Then we have the decomposition
\[
\|Z^{t+1}Q^t - M\| \leq \|Z^{t+1}Q^t - \tilde{Z}^{t+1}\| + \|\tilde{Z}^{t+1} - M\|.
\]
(96)

To bound \( \|\tilde{Z}^{t+1} - \mathcal{E} \tilde{Z}^{t+1}\| \), we first denote \( \Delta_U := Z_U^t Q^t - M_U \), \( \Delta'_U : = Z'_U Q^t - M_V \) and \( \Delta' : = \Delta_U / \Delta'_U \).

Due to \( Q^t \in O(K) \), we have
\[
\Delta_U (Z'_U)^H + M_U M_U^H = \Delta_U (Z'_U)^H + M_U (\Delta'_U)^H + \Delta_U (\Delta'_U)^H.
\]
(97)

Then
\[
\|\tilde{Z}^{t+1} - \mathcal{E} \tilde{Z}^{t+1}\| \leq 2\eta\|M_U\| \left| \frac{1}{\rho} \mathcal{G} \mathcal{P}_0 \mathcal{G}^H \left( Z_U (Z'_U)^H - M_U M_U^H \right) \right| - \mathcal{G}^H^H \left( Z_U (Z'_U)^H - M_U M_U^H \right) - 2\eta\|M_U\| \left| \frac{1}{\rho} \mathcal{H} \mathcal{P}_0 (\epsilon) \right|.
\]
(98)

where the first inequality follows from \( \|M_U\| = \|M_V\| \), \( \mathcal{G} \mathcal{Y} = \mathcal{G} \mathcal{Y}^* + \mathcal{H} \mathcal{E} \) and \( \mathcal{G} \mathcal{P}_0 \mathcal{G}^H (\mathcal{H} \mathcal{E}) = \mathcal{H} \mathcal{P}_0 (\epsilon) \), the second inequality follows from (97), the third inequality follows from Lemma 5 the fourth inequality uses Lemma 11 as well as \( \|M_U\|_{2,\infty} \), \( \|M_V\|_{2,\infty} \leq \frac{\mu c \log N}{\epsilon} \), and holds due to \( \|\Delta'\|_{2,\infty} \leq \frac{\mu c K \log N}{\epsilon} \sqrt{\frac{\log N}{c_2 \sigma^2}} \) when \( M \geq 4C_1^2 \mu c K^{1/2} N \log N \) and \( \frac{1}{\sigma K} \sqrt{\frac{N}{c_2 \sigma^2 \sqrt{\sigma^2}}} \leq \frac{1}{2c_2 \sqrt{k}} \), and the final inequality uses (37) again and holds when \( M \geq \mu^2 c^2 K^{2/2} N \log N \).

(2) To bound \( \|\mathcal{E} \tilde{Z}^{t+1} - M\| \), we have
\[
\|Q^t\|_G \left( (Z_U^t)^H Z_U^t - (Z_V^t)^H Z_V^t \right) Q^t = (\Delta_U^t)^H \Delta_U^t + (\Delta'_U)^H M_U + M_U^H \Delta_U - (\Delta'_U)^H M_V - M_V^H \Delta_V.
\]
(99)

Moreover, we obtain
\[
\|\mathcal{E} \tilde{Z}^{t+1} - M\| = \left| \left[ \Delta_U^t - \eta \Delta_U^t M_U^H M_V - \eta M_U^H \Delta_U^t + \gamma_1 \right] \right| - \left| \left[ \Delta_U^t - \eta \Delta_U^t M_U^H M_V - \eta M_U^H \Delta_U^t + \gamma_2 \right] \right| \leq \frac{1}{2} \left| |I - \eta M_U^H M_V| \|\Delta'\| \right| + \frac{1}{2} \left| |I - \eta \left[ M_U M_U^H \right] \|\Delta'\| \right| + \gamma_1 \leq \frac{1}{2} \left( 1 - 2\eta K N \right) \left| \|\Delta'\| \right| + \frac{1}{2} \left| \|\Delta'_U\| \right| + 4\eta \left| \|\Delta'_U\| \right|^{\frac{3}{2}} \frac{\sigma^2}{\sigma_1} \leq \left( 1 - 2\eta K N \right) \left| \|\Delta'\| \right| + \frac{1}{2} \left| \|\Delta'_U\| \right| + 4\eta \left| \|\Delta'_U\| \right| \frac{3}{2} \frac{\sigma^2}{\sigma_1} \leq \left( 1 - 0.9\eta K N \right) \left| \|\Delta'\| \right|,
\]
(100)

where the fact \( \mathcal{G}^H \left( M_U M_U^H \right) = M_U M_U^H \), (37, 39) and the conjugate symmetry of \( \left( \Delta'_U \right)^H M_U + \left( \Delta'_U \right)^H M_V \) by Lemma 9 are used consecutively to get the first equality, \( \gamma_1 \) and \( \gamma_2 \) containing at least two \( \Delta_U^t \)'s and \( \Delta'_U \)'s are denoted as follows:
\[
\gamma_1 := -\eta \Delta_U^t \left( \Delta'_U \right)^H M_V - \frac{\eta}{2} M_U \left( \Delta'_U \right)^H \Delta_U + \frac{\eta}{2} M_U \left( \Delta'_U \right)^H \Delta_U,
\]
\[
\gamma_2 := -\eta \Delta_U \left( \Delta'_U \right)^H M_U - \frac{\eta}{2} M_U \left( \Delta'_U \right)^H \Delta_U + \frac{\eta}{2} M_U \left( \Delta'_U \right)^H \Delta_U,
\]
the first inequality follows from \( M_U M_U^H = M_U^H M_U \), the second inequality follows from \( \|M_U M_U^H\| = \|M_U^H M_U\| \).

(3) To bound \( \|Z^{t+1}Q^t - \tilde{Z}^{t+1}\| \), we utilize \( \mathcal{L}_2 \) Lemma 37 (restated as Lemma 11) in our paper) which gives the bounds on two matrices after “aligning” them with \( M \) to deal with the unwelcome \( Q^{t+1} \). Specifically, we let \( X_1 = Z^{t+1} Q^t \) and \( X_2 = \tilde{Z}^{t+1} \) in Lemma 11 and then focus on verifying they satisfy the lemma’s required condition.

Firstly, we need to show \( Q_1 = \left( Q^t \right)^H Q^{t+1} + I \). The former obviously holds from the definition of \( Q^{t+1} \) and the latter can be verified through proving that \( M_U^H \tilde{Z}^{t+1} \) is conjugate symmetric and positive semidefinite (PSD) according to Lemma 9. To be specific, conjugate symmetric is an
immediate consequence according to the matrices definition and for PSD, we have
\[
\begin{align*}
&\|M^H \tilde{Z}^{+1} - M^H M\| \leq \|M\| \|\tilde{Z}^{+1} - M\| \\
&\leq 2\sqrt{\sigma_1} \left( \|\tilde{Z}^{+1} - E\tilde{Z}^{+1}\| + \|E\tilde{Z}^{+1} - M\| \right) \\
&\leq 2\sqrt{\sigma_1} \left( (6C_8C_1 - 0.9C_1) \eta \sigma_K + C_1 \right) \rho \sqrt{\frac{\mu c s^5 K \log N}{c s M}} \\
&+ \frac{1}{p} \mathcal{G} \mathcal{P}_1 \mathcal{G}^H \left( \mathcal{U} (\mathcal{A}_U^H)^H \right) - \mathcal{G} \mathcal{G}^H \left( \mathcal{U} (\mathcal{A}_U^H)^H \right) \\
&+ \|\mathcal{U}_M \mathcal{U}^T\| + \|\mathcal{M}_U \mathcal{A}_U^H\| + \|\mathcal{A}_U (\mathcal{A}_U^H)^H\| \\
&\leq C_8 \sqrt{\frac{N N_1 \log N}{M}} \left( \|\mathcal{U}^H\|_{2,\infty} \|\mathcal{M}_V\|_{2,\infty} \\
&+ \|\mathcal{U}^H\|_{2,\infty} \|\mathcal{M}_U\|_{2,\infty} + \|\mathcal{U}^H\|_{2,\infty} \|\mathcal{V}\|_{2,\infty} \right) \\
&+ \|\mathcal{U}_M \mathcal{U}^T\| + \|\mathcal{M}_U \mathcal{A}_U^H\| + \|\mathcal{A}_U (\mathcal{A}_U^H)^H\|,
\end{align*}
\]
where the second inequality follows from Lemma 5.
For the second part, with (99) we have
\[
\begin{align*}
\|Q^H (Z_U^H)^H Z_U - (Z_U^H)^H Z_V^T Q\| &\leq \|\Delta_U\|^2 + \|\Delta_V\|^2 + 2 \|\mathcal{M}_U\| \|\Delta_U\| + 2 \|\mathcal{M}_V\| \|\Delta_V\| \\
\|
\end{align*}
\]
\[
\leq \rho^{t+1} \left( C_1 \sqrt{\frac{\mu_c \kappa^5 K \log N}{M} + C_2 \frac{\sigma}{\kappa K} \sqrt{\frac{\kappa^5 N^2}{c_M} \sigma_1}} \right),
\]
where (34) and \( \rho = 1 - 0.01 \eta \sigma_K \) are used.

2) For proving \( \| Z^{t+1}(l, Q)^{t+1}(l, 0) - M \|_{l_2} \), we need use \( \text{Lemma } 11 \), \( \text{Lemma } 5 \), and \( \text{Lemma } 6 \). Assume \( 0 \leq l \leq N_1 - 1 \), we have
\[
\left( Z^{t+1}(l, Q)^{t+1}(l, 0) - M \right)_{l_2} = \left( Z^{t+1}(l, Q)^{t+1}(l, 0) - (M_U)_{l_2} \right) \\
= \left( Z^{t+1}(l, Q)^{t+1}(l, 0) - (M_U)_{l_2} \right).
\]

Then it turns to bound \( \| \Delta^{t}(l, Q)^{t+1}(l, 0) \|_{l_2} \). Since
\[
\| \Delta^{t}(l, Q)^{t+1}(l, 0) \|_{l_2} \leq \| Z^{t+1}(l, Q)^{t+1}(l, 0) - M \|_{l_2}.
\]
we invoke Lemma 11 to deal with \( \| Z^{t+1}(l, Q)^{t+1}(l, 0) - Z_q Q^t \|_{l_2} \).

Let \( X_1 = Z_q Q^t \) and \( X_2 = Z_q T^{t}(l) \), we have \( \| X_1 - X_2 \|_{l_2} \leq \frac{\omega}{\kappa K} \sqrt{\frac{N^2}{c_M}} \leq \frac{1}{16} \leq \frac{\omega}{\kappa K} $$ holds via \( \text{Lemma } 11 \), \( \text{Lemma } 5 \) provided \( \frac{\omega}{\kappa K} \geq 16C_2 \mu_c \kappa^2 K \log N \) and \( \frac{\omega}{\kappa K} \geq \frac{\sqrt{N^2}}{c_M} \). Meanwhile, it's obvious to verify that
\[
Q_1 = I \text{ and } Q_2 = \left( T^{t}(l) \right)^H Q^t.
\]

Therefore, according to \( \text{Lemma } 5 \) and \( \text{Lemma } 11 \), we have
\[
\text{Lemma } 5 \leq 5 \kappa \| Z^{t+1}(l, Q)^{t+1}(l, 0) - Z_q Q^t \|_{l_2} + \| Z_q Q^t - M \|_{l_2}.
\]

where the first inequality follows from Lemma 11 and the second inequality holds by invoking \( \text{Lemma } 5 \) as well as due to \( M \geq 25C_2 \mu_c \kappa^2 K \log N \) and \( \frac{\omega}{\kappa K} \geq \frac{\sqrt{N^2}}{c_M} \). Therefore, according to \( \text{Lemma } 5 \) and \( \text{Lemma } 11 \), we have
\[
\text{Lemma } 5 \leq 5 \kappa \| Z^{t+1}(l, Q)^{t+1}(l, 0) - Z_q Q^t \|_{l_2} + \| Z_q Q^t - M \|_{l_2}.
\]

where the first inequality holds if when \( M \geq 16C_2 \mu_c \kappa^2 K \log N \) and \( \frac{\omega}{\kappa K} \geq \frac{\sqrt{N^2}}{c_M} \), the second and third inequality follow from \( M \geq 900C_2 \mu_c \kappa^2 K \log N \) and \( \frac{\omega}{\kappa K} \geq \frac{\sqrt{N^2}}{c_M} \) and the final inequality follows when \( M \geq 4C_2 \mu_c \kappa^2 K \log N \), \( \frac{\omega}{\kappa K} = \frac{\sqrt{N^2}}{c_M} \) and \( \kappa \leq \frac{\omega}{800C_2 \kappa^2 K} \).

Then we are to control \( \| F_2 \| \). Notice \( \text{Lemma } 11 \) and \( \text{Lemma } 16 \), we have
\[
\| F_2 \| \leq \| F_1 + (M_U)_{l_2} \| \| Q^t \|^{-1} \| Q^{t+1} - I \|.
\]

where the first inequality holds if when \( M \geq 16C_2 \mu_c \kappa^2 K \log N \) and \( \frac{\omega}{\kappa K} \geq \frac{\sqrt{N^2}}{c_M} \). Second and third inequality follow from \( M \geq 900C_2 \mu_c \kappa^2 K \log N \) and \( \frac{\omega}{\kappa K} \geq \frac{\sqrt{N^2}}{c_M} \) and the final inequality follows when \( M \geq 4C_2 \mu_c \kappa^2 K \log N \), \( \frac{\omega}{\kappa K} = \frac{\sqrt{N^2}}{c_M} \) and \( \kappa \leq \frac{\omega}{800C_2 \kappa^2 K} \).
where the final inequality follows from Claim 1

Claim 1. Recall the definition of $Q^{t,(l)}$ in (32), and let $0 \leq l \leq N_1 - 1$. Under the assumption in Proposition 3, we have

$$
\left\| \left( Q^{t,(l)} \right)^{-1} Q^{t+1,(l)} - I \right\| 
\leq \frac{C_1 \rho^l \sqrt{\frac{\mu_c \kappa^3 K \log N}{M} + \frac{C_2 \sigma}{\sqrt{\kappa^3 N^2 c_1 M}}} + \frac{C_2 \sigma}{\sqrt{\kappa^3 N^2 c_2 M}}}{M} \frac{\sigma^2}{\kappa^2},
$$

(118)

Finally, we focus on bound $\|F_3\|_2$.

$$
\|F_3\|_2 \leq \left\| \left( Q^{t,(l)} \right)^T \left( Z^{t,(l)}_U \right)^T \left( Z^{t,(l)}_V \right)^T \left( Z^{t,(l)}_V \right)^T \left( Z^{t,(l)}_U \right)^T \right\|_2
\leq \frac{\eta}{2} \left\| \left( Z^{t,(l)}_U \right)^T \right\|_2 \left\| \left( \Delta^{t,(l)}_U \right)^T \right\|_2 + \frac{\eta}{2} \left\| \left( \Delta^{t,(l)}_V \right)^T \right\|_2 + 2 \left\| \left( \Delta^{t,(l)}_V \right)^T \left( \Delta^{t,(l)}_V \right) \right\|_2 + 2 \left\| \left( \Delta^{t,(l)}_U \right)^T \left( \Delta^{t,(l)}_U \right) \right\|_2
\leq \left\| M_U \right\|_2 + 2 \left\| \Delta^{t,(l)}_V \right\|_2 + 2 \left\| \Delta^{t,(l)}_U \right\|_2
= 6\eta \sqrt{\frac{\mu_c K}{N} \sigma_1 \left\| \Delta^{t,(l)}_V \right\|_2},
$$

(119)

where the first inequality is obtained by (12) and the second inequality follows from (105) and the final inequality holds due to

$$
\left\| (M_U)_{l_r} \right\|_2 \leq \frac{\mu_c K}{N} \sqrt{\kappa} \|
$$

if

$$
M \geq 4C_1^2 \mu_c \kappa^7 K \log N \quad \text{and} \quad \frac{\sigma}{\kappa} \sqrt{\frac{N^2 c_1 M}{C_1^2 \mu_c \kappa^7 K \log N}} \leq \frac{1}{2C_1 \sqrt{\kappa}},
$$

as well as

$$
\left\| \Delta^{t,(l)}_V \right\|_2 \leq \sqrt{\kappa} \|
$$

if

$$
M \geq 16C^2_1 \mu_c \kappa^7 K \log N
$$

and

$$
\frac{\sigma}{\kappa} \sqrt{\frac{N^2 c_1 M}{C_1^2 \mu_c \kappa^7 K \log N}} \leq \frac{1}{2C_1 \sqrt{\kappa}}.
$$

Combining the above bounds on $\|F_1\|_2$, $\|F_2\|_2$ and $\|F_3\|_2$, i.e., (116), (117) and (119) respectively, we arrive at

$$
\left\| \left( Z^{t+1,(l)} \right)^{-1} Z^{t+1,(l)} - M \right\|_2 \leq \left\| F_1 \right\|_2 + \left\| F_2 \right\|_2 + \left\| F_3 \right\|_2
\leq \left( 1 - 0.6\eta \sigma_K \right) \left\| \Delta^{t,(l)}_V \right\|_2 + 2\eta \sigma_K \sqrt{\frac{\mu_c K}{N} \kappa \left\| \Delta^{t,(l)}_V \right\|_2}
+ 2\eta \sigma_K \left[ C_1 \rho^l \sqrt{\frac{\mu_c \kappa^2 K^2 \log N}{M N}} + C_2 \sigma \frac{\sqrt{\mu_c K^2 N}}{\kappa} \right] \left\| \Delta^{t,(l)}_V \right\|_2
+ 6\eta \sqrt{\frac{\mu_c K}{N} \sigma_1 \left\| \Delta^{t,(l)}_V \right\|_2}
\leq \left( C_1 \rho^l + 1 \right) \sqrt{\frac{\mu_c \kappa^2 K^2 \log N}{M N}} + C_2 \sigma \frac{\sqrt{\mu_c K^2 N}}{\kappa} \sqrt{\frac{\mu_c K}{N} \sigma_1}
$$

(120)

with $\rho = 1 - 0.01\eta \sigma_K$ for some proper positive constants $C_1$ and $C_2$. Besides, the proof and conclusion for $N_1 \leq l < N_1 + N_2 - 1$ stay nearly the same.

3) For proving $\left\| Z^{t+1} Q^{t+1} - Z^{t+1} T^{t+1} \right\|_F$, we need use (33), (36) and (37). According to the definition of $T^{t+1,(l)}$ in (33), we have

$$
\left\| Z^{t+1} Q^{t+1} - Z^{t+1} T^{t+1} \right\|_F \leq \left\| Z^{t+1} Q^{t} - Z^{t+1} T^{t} \right\|_F = \left\| \mathcal{E}_1 + \mathcal{E}_2 \right\|_F, \quad (121)
$$

where

$$
\mathcal{E}_1 = \left( Z^t - \eta \nabla f (Z^t) \right) Q^t - \left( Z^{t+1} Q^{t+1} - Z^{t+1} T^{t+1} \right) \right\|_F
$$

and

$$
\mathcal{E}_2 = \left( Z^t - \eta \nabla f (Z^t) \right) Q^t - \left( Z^{t+1} T^{t+1} \right) \right\|_F
$$

Notice that $Z^{t+1} = Z^t - \eta \nabla f (Z^t)$, $Z^{t+1} = Z^t - \eta \nabla h (Z^t)$ and the difference between $\nabla f (\cdot)$ and $\nabla h (\cdot)$ in (29) and (16).

For $\mathcal{E}_1$, we denote $p (\theta) = \theta \cdot (Z^{t,(l)} T^{t,(l)} - Z^t Q^t) + Z^t Q^t$ and then have

$$
\left\| \mathcal{E}_1 \right\|_F = \left\| Z^{t+1} Q^t - Z^t Q^t - \eta \left[ \nabla f \left( Z^{t+1} T^{t,(l)} \right) - \nabla f (Z^{t+1} Q^t) \right] \right\|_F^2
$$

$$
\leq \left\| \left[ I - \eta \int_0^1 \nabla^2 f (p (\theta)) \, d\theta \right] \cdot \nabla f (Z^{t+1} T^{t,(l)} - Z^t Q^t) \right\|_2^2
$$

$$
\left\| Z^{t+1} Q^t - Z^t Q^t - \eta \nabla^2 f (p (\theta)) \cdot \nabla f (Z^{t+1} T^{t,(l)} - Z^t Q^t) \right\|_2^2
$$

$$
\left\| Z^{t+1} Q^t - Z^t Q^t \right\|_2^2 \cdot \left\| \nabla^2 f (p (\theta)) \right\|_2^2 \cdot \left\| Z^{t+1} T^{t,(l)} - Z^t Q^t \right\|_2^2
$$

(122)

To utilize the benign geometric property of the Hessian in RIC as shown from Proposition 1, we let

$$
M \geq 3200^2 C_1^2 C_2^2 \mu_c^2 K^2 \kappa^1 K \log N \quad \text{and} \quad \frac{\sigma}{\kappa} \sqrt{\frac{N^2 c_1 M}{C_1^2 \mu_c \kappa^7 K \log N}} \leq \frac{1}{2C_1 \sqrt{\kappa}},
$$

then

$$
\left\| Z^{t+1} Q^t - M \right\|_2 \leq \frac{1}{1600 K \sqrt{N_1 + N_2}} \sqrt{\sigma_1},
$$

(123)

Hence, there holds $\left\| p (\theta) - M \right\|_2 \leq \frac{1}{800 K \sqrt{N_1 + N_2}} \sqrt{\sigma_1}$ for any $0 \leq \theta \leq 1$. Moreover, we can get

$$
\left\| Z^{t+1} Q^t - Z^t Q^t \right\|_2 \leq \left( 1 - 0.03 \sigma_K \right) \left\| Z^{t+1} T^{t,(l)} - Z^t Q^t \right\|_F
$$

(124)

where the last inequality holds due to $\eta \leq \frac{\sigma}{M}$.
For \( \mathcal{E}_{21} \) and \( \mathcal{E}_{22} \), we have

\[
\left\| \mathcal{E}_{21} \right\|_F \leq \left\| \mathcal{E}_{21} \right\|_F + \left\| \mathcal{E}_{22} \right\|_F \leq \sqrt{N_2 \sum_{i=0}^{N_2-1} \left( \frac{1}{p} \delta_{t,i} - 1 \right) \left( Z_U^{(t)} (Z_V^{(t)})^H - \mathcal{G} y^* \right)^H_i, i} \left( Z_V^{(t)} \right)_i, \right\|_2^2 + \left( \frac{1}{p} \mathcal{H}_{\mathcal{P}_1} (\epsilon) \right) \left\| Z_V^{(t)} \right\|_2^2, \right\|_2^2 \leq 2 \left( C_1 \rho^t \sqrt{\frac{\mu^2 c_2 \kappa^2 K^2 \log N}{M N}} + C_2 \frac{\sigma}{\sigma_K} \sqrt{\frac{\mu K N}{M}} \right) \sqrt{\sigma_1} \leq \sqrt{\frac{\mu c_2 K N}{\sigma_1}}, \tag{126}
\]

where the second inequality follows from the entries value of \( \mathcal{E}_{21} \) and \( \mathcal{E}_{22} \) according to the definition of \( \mathcal{P}_1 \), and \( \mathcal{P}_{-t} \), and \( \delta_{t,i} = 1 \) if \( l + i \in \Omega \), otherwise \( \delta_{t,i} = 0 \).

Before bounding \( \mathcal{W}_1 \) and \( \mathcal{W}_2 \), we firstly denote \( \Xi^{(t)} := Z_U^{(t)} T^{(t)} - M_U, \Xi^{(t)} := Z_V^{(t)} T^{(t)} - M_V \) and \( \Xi^{(t)} := \Xi^{(t)}_U \Xi^{(t)}_V \). Moreover, we have

\[
\left\| \Xi^{(t)} \right\|_{2,\infty} \leq \left\| Z^T \mathcal{Q}^t - Z^T \mathcal{Q}^{(t)} \right\|_F + \left\| Z^T \mathcal{Q}^t - M \right\|_{2,\infty} \leq 2 \left( C_3 \rho^t \sqrt{\frac{\mu^2 c_2 \kappa^2 K^2 \log N}{M N}} + C_4 \frac{\sigma}{\sigma_K} \sqrt{\frac{\mu K N}{M}} \right) \sqrt{\sigma_1} \leq \sqrt{\frac{\mu c_2 K N}{\sigma_1}}, \tag{127}
\]

where the second inequality follows from \( \Xi^{(t)}_U \) and \( \Xi^{(t)}_V \), and the final one holds when \( M \geq 16 C_7^2 \mu c_2 \kappa^2 K \log N \) and \( \frac{\sigma}{\sigma_K} \leq \frac{1}{4 c_2 \sqrt{\kappa}} \).

Meanwhile, from (36) and (37) we have

\[
\left\| \Xi^{(t)}_U \right\|_{2,\infty} \leq \left\| Z^T \mathcal{Q}^t - Z^T \mathcal{Q}^{(t)} \right\|_F + \left\| Z^T \mathcal{Q}^t - M \right\|_{2,\infty} \leq 2 \left( C_6 \rho^t \sqrt{\frac{\mu c_2 \kappa^2 K \log N}{M}} + C_7 \frac{\sigma}{\sigma_K} \sqrt{\frac{\kappa^2 N^2 \mu}{c_2 M}} \right) \sqrt{\sigma_1}, \tag{128}
\]

where the last inequality follows from \( \frac{\mu c_2 K N \log N}{\mu c_2 K N \log N} \leq \frac{M}{N} \leq 1 \).

Then for the first part \( \mathcal{W}_1 \), we let \( \mathcal{W}_1 := \sum_{i=0}^{N_2-1} \mathcal{W}_1 \), then \( \mathcal{W}_1 \)'s are independent and \( \mathbb{E} \mathcal{W}_1 = 0 \). What’s more,

\[
\left\| \mathcal{W}_1 \right\|_2^2 \leq \frac{1}{p} \left\| Z_U^{(t)} \mathcal{Q}^{(t)} - \mathcal{G} y^* \right\| \left\| Z_V^{(t)} \right\|_{2,\infty} \leq \frac{1}{p} \left( \left\| \Xi^{(t)}_U \right\|_{2,\infty} \right) \left\| M_V \right\|_{2,\infty} + \left\| M_U \right\|_{2,\infty} \left\| \Xi^{(t)}_V \right\|_{2,\infty} + \left\| \Xi^{(t)}_U \right\|_{2,\infty} \left\| \Xi^{(t)}_V \right\|_{2,\infty} \left\| Z_V^{(t)} \right\|_{2,\infty} \right\| \left\| Z_V^{(t)} \right\|_{2,\infty} \leq 4 \sigma \kappa \frac{C_6 \rho^t \sqrt{\mu K N}}{\sigma_1} \sqrt{\sigma_1}, \tag{129}
\]

holds with probability at least \( 1 - (N_1 + N_2)^{-1} \), where the second inequality follows from \( M \geq 6 c_1^2 \frac{\mu c_2 \kappa^2 K \log N}{M^2} \), and the last inequality holds since \( M \geq 3200 \frac{C_7^2 \mu c_2 \kappa^2 K \log N}{N} \).

For the second part, we have

\[
\left\| Z_V^{(t)} \right\|_{2,\infty} \leq 2 \sqrt{\frac{\mu c_2 K N}{\sigma_1}} \sqrt{\sigma_1}, \tag{129}
\]

from Lemma 1 and \( \left\| Z^{(t)}_V \right\|_{2,\infty} \leq 2 \sqrt{\frac{\mu c_2 K N}{\sigma_1}} \sqrt{\sigma_1} \).
For the third part \( W_2 \), it is identical to \( X_2 \) in (140). According to (141), then we have

\[
\|W_2\| \cdot \left\| \left( Z_{U}^{(l)} \right)_{1,2} \right\|_2 \leq 16C_{11} \left( C_{1} \rho \sqrt{1 + \frac{c_{2}^{2} \sigma_{K}^{2} K^{2} \log^{2} N}{M^{2}}} \right) + C_{2} \sigma_{K} \sqrt{\frac{\mu c_{2}^{2} K^{2} \log^{2} N}{M^{2}}} \sigma_{1} \leq 0.01\sigma_{K} \left( C_{1} \rho \sqrt{1 + \frac{c_{2}^{2} \sigma_{K}^{2} K^{2} \log N}{M^{2}}} \right) + C_{2} \sigma_{K} \sqrt{\frac{\mu c_{2}^{2} K^{2} \log N}{M}} \sigma_{1},
\]

where the last inequality follows from \( \left\| Z_{V}^{(l)} \right\|_{2, \infty, \infty} \leq 2\sqrt{\frac{\mu c_{2}^{2} K^{2}}{\rho_{0} \sigma_{1}}} \) and \( M \geq c_{1}^{2} \mu c_{2} K^{4} K \log N \).

Hence, combining (125), (128), (129) and (130) together, we have

\[
\left\| \left[ E_{22} \right] \right\| = \left\| \left[ E_{2} \right] \right\| \leq 0.02\sigma_{K} \left( C_{1} \rho \sqrt{1 + \frac{c_{2}^{2} \sigma_{K}^{2} K^{2} \log N}{M^{2}}} \right) + C_{2} \sigma_{K} \sqrt{\frac{\mu c_{2}^{2} K^{2} \log N}{M}} \sigma_{1},
\]

with suitable positive constants \( C_{1} \) and \( C_{2} \).

Therefore, from (121), (124) and (131), we obtain

\[
\left\| \left( Z_{U}^{(l+1)} - Z_{U}^{(l+1,1)} \right) T_{l+1,1} \right\|_F \leq \left\| E_{1} \right\|_F + \eta \left\| \left[ E_{22} \right] \right\|_F \leq (1 - 0.03\eta)\sigma_{K} \left\| Z_{U}^{(l)} \right\|_F \cdot \left\| T_{l+1,1} \right\|_F - \frac{1}{2} q_{K} \sigma_{K} \sqrt{1 + \frac{c_{2}^{2} \sigma_{K}^{2} K^{2} \log N}{M^{2}}} \sigma_{1} \leq \left( C_{1} \rho \sqrt{1 + \frac{c_{2}^{2} \sigma_{K}^{2} K^{2} \log N}{M^{2}}} \right) + C_{2} \sigma_{K} \sqrt{\frac{\mu c_{2}^{2} K^{2} \log N}{M}} \sqrt{\sigma_{1}}.
\]

for \( 0 \leq l \leq N_{1} - l \) with suitable positive constants \( C_{1} \) and \( C_{2} \) where \( \rho = 1 - 0.01\eta \sigma_{K} \) and (36) is used. The same conclusion and proof hold for \( N_{1} - l \leq N_{2} - 1 \).

4) For proving \( \left\| Z_{U}^{(l+1)} \right\|_{2, \infty} \), we need use (34), (35) and (36). For any \( 0 \leq l \leq N_{1} + N_{2} - 1 \), we have

\[
\left\| Z_{U}^{(l+1)} \right\|_{2, \infty} \leq \left\| Z_{U}^{(l+1)} - M \right\|_{2, \infty} \leq \frac{1}{2} q_{K} \sigma_{K} \sqrt{1 + \frac{c_{2}^{2} \sigma_{K}^{2} K^{2} \log N}{M^{2}}} \leq \frac{1}{8C_{2} \sqrt{\epsilon}}.
\]

Moreover, from (36) we have

\[
\left\| Z_{U}^{(l+1)} - Z_{U}^{(l+1,1)} \right\| \leq \frac{1}{4} q_{K} \sigma_{K} \sqrt{1 + \frac{c_{2}^{2} \sigma_{K}^{2} K^{2} \log N}{M^{2}}} \leq \frac{1}{8C_{2} \sqrt{\epsilon}}.\]
\(-\frac{9}{2} \mathbf{M}_U (\mathbf{Q}^{t(l)})^H \left( (\mathbf{Z}_U^{t(l)})^H \mathbf{Z}_U^{t(l)} - (\mathbf{Z}_V^{t(l)})^H \mathbf{Z}_V^{t(l)} \right) \mathbf{Q}^{t(l)},

\begin{align*}
\mathbf{Z}_U^{t(l)} &= \mathbf{Z}_U^{t(l)} - \eta \left( \mathbf{Z}_U^{t(l)} \left( \mathbf{Z}_V^{t(l)} \right)^H - \mathbf{G} y^* \right) \mathbf{M}_V \\
\mathbf{Z}_V^{t(l)} &= \mathbf{Z}_V^{t(l)} - \eta \left( \mathbf{Z}_V^{t(l)} \left( \mathbf{Z}_U^{t(l)} \right)^H - \mathbf{G} y^* \right) \mathbf{M}_U \\
\mathbf{Z}_U^{t(l)} &= \mathbf{Z}_U^{t(l)} - \eta \left( \mathbf{Z}_U^{t(l)} \left( \mathbf{Z}_V^{t(l)} \right)^H - \mathbf{G} y^* \right)^H \mathbf{M}_U \\
\mathbf{Z}_V^{t(l)} &= \mathbf{Z}_V^{t(l)} - \eta \left( \mathbf{Z}_V^{t(l)} \left( \mathbf{Z}_U^{t(l)} \right)^H - \mathbf{G} y^* \right)^H \mathbf{M}_U \\
&= \frac{9}{2} \mathbf{M}_U (\mathbf{Q}^{t(l)})^H \left( (\mathbf{Z}_U^{t(l)})^H \mathbf{Z}_U^{t(l)} - (\mathbf{Z}_V^{t(l)})^H \mathbf{Z}_V^{t(l)} \right) \mathbf{Q}^{t(l)}.
\end{align*}

To apply Lemma 10 we let \( \mathbf{C} = \mathbf{M}^H \mathbf{Z}^{t+1(l)} \) and \( \mathbf{E} = \mathbf{M} \mathbf{Z}^{t+1(l)} - \mathbf{Z}^{t+1(l)} \). It's easy to verify that \( \text{sgn}(\mathbf{C} + \mathbf{E}) = \arg \min_Q \| \mathbf{Z}^{t+1(l)} \mathbf{Q}^{t(l)} - \mathbf{M} \|_F = (\mathbf{Q}^{t(l)})^{-1} \mathbf{Z}^{t+1(l)} \) by (132). Then we need to prove \( \text{sgn}(\mathbf{C}) = I \), i.e., \( \mathbf{C} \) is positive definite. The whole process is similar to the operation around (101). The conjugate symmetry of \( \mathbf{M}^H \mathbf{Z}^{t+1(l)} \) can be confirmed via the corresponding matrices definition. To show all eigenvalues of \( \mathbf{C} \) are positive, we have

\[
\| \mathbf{Z}^{t+1(l)} - \mathbf{M} \| \leq \| \mathbf{Z}^{t+1(l)} - \mathbf{E} \mathbf{Z}^{t+1(l)} \| + \| \mathbf{E} \mathbf{Z}^{t+1(l)} - \mathbf{M} \|. 
\]

Putting (135), (136), (137), (126) and (115) together, we have

\[
\| \mathbf{Z}^{t+1(l)} - \mathbf{M} \| 
\leq 6 \eta \mathbf{C}_S \sqrt{\frac{\mu_c K N \log N}{M}} \| \mathbf{Z}^{t(l)} \|_{\infty} + 2 \eta C_0 \sqrt{\frac{N^2}{c_s M}} \sqrt{\sigma_1} + (1 - \eta \sigma_K) \| \mathbf{A}^{t(l)} \| + 4 \eta \| \mathbf{A}^{t(l)} \|^2 \frac{\sigma_1}{\sqrt{\sigma_1}} \\
+ \left( 1 - \eta \sigma_K \right) \| \mathbf{A}^{t(l)} \| + 4 \eta \| \mathbf{A}^{t(l)} \|^2 \frac{\sigma_1}{\sqrt{\sigma_1}} \leq \frac{1}{6} \sigma_K,
\]

where the second inequality holds if \( \mathbf{M} \geq 144 \sigma_2 \mu_c^2 \mu_c K^2 K^2 \log N \), and due to \( \| \mathbf{A}^{t(l)} \| \leq \frac{1}{6} \sqrt{\sigma_1} \) when \( \mathbf{M} \geq 1024 \sigma_2 \mu_c^2 \mu_c K^2 K^2 \log N \), and \( \frac{\sigma_2}{\sigma_K} \sqrt{\frac{N^2}{c_s}} \leq \frac{1}{32 \sigma_2 \sqrt{\sigma_1}} \), and the final one follows from (115) \( \leq \frac{1}{32 \sigma_2 \sqrt{\sigma_1}} \) when \( \mathbf{M} \geq 1024 \sigma_2 \mu_c^2 \mu_c K^2 K^2 \log N \). Then we have \( \| \mathbf{M}^H \mathbf{Z}^{t+1(l)} - \mathbf{M} \| \leq \| \mathbf{Z}^{t+1(l)} - \mathbf{M} \| \leq \frac{1}{6} \sigma_K \).

With \( \lambda_K(\mathbf{M}^H \mathbf{M}) = \lambda_K(2 \mathbf{M}_U^H \mathbf{M}_U) = 2 \sigma_K \), and Weyl's inequality, we obtain \( \lambda_K(\mathbf{M}^H \mathbf{Z}^{t+1(l)} - \mathbf{M}) \geq \frac{1}{6} \sigma_K \), and so \( \mathbf{C} = \mathbf{M}^H \mathbf{Z}^{t+1(l)} \) is positive definite.

Then the remaining thing is to bound \( \| \mathbf{E} \| = \| \mathbf{M}^H (\mathbf{Z}^{t+1(l)} - \mathbf{Z}^{t+1(l)} \mathbf{Q}^{t(l)} - \mathbf{Z}^{t+1(l)} \mathbf{Q}^{t(l)}) \| \). Let us consider its top half part firstly,

\[
\| \mathbf{Z}^{t+1(l)} - \mathbf{Z}^{t+1(l)} \mathbf{Q}^{t(l)} \| 
\leq 6 \eta \mathbf{C}_S \sqrt{\frac{\mu_c K N \log N}{M}} \| \mathbf{Z}^{t(l)} \|_{\infty} + 2 \eta C_0 \sqrt{\frac{N^2}{c_s M}} \sqrt{\sigma_1},
\]

where the second inequality follows from the fact that the spectral norm of a submatrix is no more than the whole matrix, the third inequality uses the transformation like (97) and Lemma 5, and the final inequality holds via (126) and Lemma 1.
There exists similar equality for $\|\tilde{Z}_V^{t+1,I} - Z_V^{t+1,I} Q_{t,I}\|$. Putting them together, we have

$$\|\tilde{Z}_V^{t+1,I} - Z_V^{t+1,I} Q_{t,I}\|$$

$$\leq \eta \|X_1\| \|\Delta_{t,I}\| + \eta \|X_2\| \|\Delta_{t,I}\| + \eta \|X_3\|$$

$$+ \eta \left\| \frac{1}{p} \mathcal{H}_P (\varepsilon) \right\| \|\Delta_{t,I}\|$$

$$= \eta \left\| (I - \mathcal{G} \mathcal{G}^H) \left( Z_U^{t,I} (Z_V^{t,I})^H\right) \right\| \|\Delta_{t,I}\|$$

$$+ \frac{1}{p} \mathcal{H}_P \mathcal{G} \mathcal{H} \left( \left( Z_U^{t,I} (Z_V^{t,I})^H - \mathcal{G} y^*\right) \right) \|\Delta_{t,I}\|$$

$$+ \frac{1}{p} \mathcal{P}_I \mathcal{G} \mathcal{H} \left( \left( Z_U^{t,I} (Z_V^{t,I})^H - \mathcal{G} y^*\right) \right) \|\Delta_{t,I}\|$$

$$+ \|Q_{t,I}\|^2 \left( \left( Z_U^{t,I} (Z_V^{t,I})^H - \mathcal{G} y^*\right) \right) \|\Delta_{t,I}\|$$

$$+ \|Q_{t,I}\|^2 \left( \left( Z_U^{t,I} (Z_V^{t,I})^H - \mathcal{G} y^*\right) \right) \|\Delta_{t,I}\|$$

$$+ \eta \left\| \frac{1}{p} \mathcal{H}_P (\varepsilon) \right\| \|\Delta_{t,I}\|.$$ (138)

For $X_1$, we have

$$\|X_1\|$$

$$\leq \frac{1}{p} \mathcal{P}_I \mathcal{G} \mathcal{H} \left( \left( Z_U^{t,I} (Z_V^{t,I})^H - \mathcal{G} y^*\right) \right)$$

$$\leq \frac{1}{p} \mathcal{P}_I \mathcal{G} \mathcal{H} \left( \left( Z_U^{t,I} (Z_V^{t,I})^H - \mathcal{G} y^*\right) \right)$$

$$\leq C_8 \sqrt{NN_1 \log N} \left( \|M\|_{2,\infty} + \|M_{U}\|_{2,\infty} \right)$$

$$\cdot \left\| \mathcal{P}_I \mathcal{H} \right\|_{2,\infty} + \left\| \mathcal{P}_I \mathcal{H} \right\|_{2,\infty}$$

$$+ \left\| \mathcal{P}_I \mathcal{H} \right\|_{2,\infty} + \left\| \mathcal{P}_I \mathcal{H} \right\|_{2,\infty}$$

$$+ \left\| \mathcal{P}_I \mathcal{H} \right\|_{2,\infty} + \left\| \mathcal{P}_I \mathcal{H} \right\|_{2,\infty}$$

$$\leq 3C_8 \sqrt{\mu \kappa N \log N} \left( \|M\|_{2,\infty} + \|M_{U}\|_{2,\infty} \right)$$

$$+ 3 \sqrt{\sigma_1} \|\Delta_{t,I}\|.$$ (139)

where the first inequality uses the fact that $\mathcal{G} \mathcal{G}^H \mathcal{G} y^* = \mathcal{G} y^*$. The second one follows from similar decompositions of (77) and Lemma [5], the third one uses (115), (120) and $\|\Delta_{t,I}\| \leq \sqrt{\sigma_1}$ since $M \geq 16C_2^2 \mu \kappa N \log N$ and $\frac{\sigma_1}{\sqrt{\sigma_1}} \leq \frac{1}{\sqrt{\sigma_1}} \leq \frac{1}{\sqrt{\sigma_1}}$, and the last inequality holds with (126) and (115) since $M \geq 36C_2^2 \mu \kappa N^2 \log N$.

For $X_2$, we can notice that only the $l$-row exists the non-zero entries. We can rewrite as follows:

$$X_2 = \sum_{j=0}^{N_2-1} \left( \frac{1}{p} \delta_{i,j} - 1 \right) \left( Z_U^{t,I} (Z_V^{t,I})^H - \mathcal{G} y^*\right) e_j^I e_j^I$$

$$:= \sum_{j=0}^{N_2-1} w_j.$$ (140)
\[
\leq 2 \sqrt{\sigma_1 \eta} \left( \|X_1\| + \|X_2\| + \|X_3\| + \left\| \frac{1}{p} H(P_{\Omega}(\varepsilon)) \right\| \right) \|\Delta^{t:(l)}\|
\]
\[
\leq 2\eta \left[ \widetilde{C}_1 \rho^t \frac{\mu \sigma^4 K \log N}{M} + \widetilde{C}_2 \frac{\sigma}{\sigma_K} \sqrt{\frac{\kappa^3 N^2}{c_* M}} \right] \sigma_1^2
\]
\[
\leq \sigma_K, \quad (143)
\]

where \( \widetilde{C}_1 = (19 + C_{11}) C_1, \widetilde{C}_2 = (19 + C_{11}) C_2 + C_0, \)
the second inequality holds due to \( \|\Delta^{t:(l)}\| \leq \frac{1}{\alpha} \sqrt{\sigma_1} \) since
\( M \geq 4C_1^2 \mu \sigma^4 K \log N \) and \( \frac{\sigma}{\sigma_K} \sqrt{\frac{N^2}{c_* M}} \leq \frac{1}{\sqrt{\sigma_1}} \), and
the last inequality follows from \( M \geq 16\widetilde{C}_1^2 \mu \sigma^4 K \log N, \)
\( \frac{\sigma}{\sigma_K} \sqrt{\frac{N^2}{c_* M}} \leq \frac{1}{4C_2^2 \sigma_1} \) and \( \eta \leq \frac{\sigma_K}{400\sigma_1}. \)

According to the previous result \( \lambda_K \left( M^H \widetilde{Z}^{t+1:(l)} \right) \geq \frac{1}{2} \sigma_K \) and (143), the condition \( \|E\| \leq \lambda_K (C) \) holds. Therefore, by Lemma 10, we have
\[
\left\| \left( Q^{t:(l)} \right)^{-1} Q^{t+1:(l)} - I \right\| = \|\text{sgn} (C + E) - \text{sgn} (C)\|
\]
\[
\leq \frac{\left\| M^H (Z^{t+1:(l)} Q^{t:(l)} - \widetilde{Z}^{t+1:(l)}) \right\|}{\lambda_K \left( M^H \widetilde{Z}^{t+1:(l)} \right)}
\]
\[
\leq \frac{12}{11} \eta \left[ \widetilde{C}_1 \rho^t \frac{\mu \sigma^4 K \log N}{M} + \widetilde{C}_2 \frac{\sigma}{\sigma_K} \sqrt{\frac{\kappa^3 N^2}{c_* M}} \right] \sigma_1^2
\]
\[
\leq \frac{12}{11} \eta \left[ \widehat{C}_1 \rho^t \frac{\mu \sigma^4 K \log N}{M} + \widehat{C}_2 \frac{\sigma}{\sigma_K} \sqrt{\frac{\kappa^3 N^2}{c_* M}} \right] \sigma_1^2
\]
\[
\leq \frac{12}{11} \eta \left[ \widehat{C}_1 \rho^t \frac{\mu \sigma^4 K \log N}{M} + \widehat{C}_2 \frac{\sigma}{\sigma_K} \sqrt{\frac{\kappa^3 N^2}{c_* M}} \right] \sigma_1^2
\]
\[
(144)
\]
where the last line uses the middle result of (143).