Spectrum Truncation Power Iteration for Agnostic Matrix Phase Retrieval

Lewis Liu, Songtao Lu, Tuo Zhao, and Zhaoran Wang

***Abstract*—Agnostic matrix phase retrieval (AMPR) is a general low-rank matrix recovery problem given a set of noisy high-dimensional data samples. To be specific, AMPR is targeting at recovering an r-rank matrix $\mathbf{M}^* \in \mathbb{R}^{d_1 \times d_2}$ as the parametric component from n instantiations/samples of a semi-parametric model $y = f(\langle \mathbf{M}^*, \mathbf{X} \rangle, \epsilon)$, where the predictor matrix is denoted as $\mathbf{X} \in \mathbb{R}^{d_1 \times d_2}$, link function $f(\cdot, \epsilon)$ is agnostic under some mild distribution assumptions on $\mathbf{X}$, and ϵ represents the noise. In this paper, we formulate AMPR as a rank-restricted largest eigenvalue problem by applying the second-order Stein's identity and propose a new spectrum truncation power iteration (STPower) method to obtain the desired matrix efficiently. Also, we show a favorable rank recovery result by adopting the STPower method, *i.e.*, a near-optimal statistical convergence rate under some relatively general model assumption from a wide range of applications. Extensive simulations verify our theoretical analysis and showcase the strength of STPower compared with the other existing counterparts.**

***Index Terms*—Spectrum Truncation Power (STPower), agnostic matrix phase retrieval (AMPR), first- and second-order Stein's identity, eigenvalue problem**

I. Introduction

In many large-scale machine learning and signal processing scenarios, it is required to describe the relationship between a given response $y \in \mathbb{R}$ and corresponding data $\mathbf{X} \in \mathbb{R}^{d_1 \times d_2}$ by estimating the potential parametric components $\mathbf{M}^* \in \mathbb{R}^{d_1 \times d_2}$ of an underlying learning model. The following two facts motivate us to recover parameter $\mathbf{M}^*$ in the *agnostic matrix phase retrieval* (AMPR) problem: *i*) the correlated information among entries in $\mathbf{X}$ results in correlated columns (rows) of $\mathbf{M}^*$ [1]–[6]; *ii*) the knowledge of the nonparametric function parts of the model is scarce. To be more specific, the AMPR problem can be formulated as

$$y = f(\langle \mathbf{M}^*, \mathbf{X} \rangle, \epsilon), \text{ subject to } \quad \operatorname{rank}(\mathbf{M}^*) = r^* \leq r_0, \text{ (I.1)}$$

and $\|\mathbf{M}^*\|_F = 1$, where $\mathbf{X}, \mathbf{M}^* \in \mathbb{R}^{d_1 \times d_2}$, $\mathbf{X}$ follows a general distribution with conditions specified later (mostly sub-Gaussian chosen by the observer), ϵ denotes the zero-mean random noise, and $f(\cdot, \epsilon)$ is the nonparametric link function. Note that $f(\cdot, \epsilon)$ might be misspecified with the ground truth, *e.g.*, in some image signal measurements, it is hard to specify whether the output quantity is the absolute value or square of the inner product between data $\mathbf{X}$ and parameter $\mathbf{M}^*$. In this model, it is postulated that $\|\mathbf{M}^*\|_F = 1$ for identifiability as its Frobenius norm can always be incorporated into the nonparametric component. In addition, following [7], we only require that there exists a mapping $\psi : u \mapsto \mathbb{E}[f(U, \epsilon)|U = u]$ with respect to $U = \langle \mathbf{M}^*, \mathbf{X} \rangle$ so that $\mathbb{E}D^2\psi(U) > 0$, where D^2 is the second-order distributional derivative [8]. In practice, matrix $\mathbf{M}^*$ following (I.1) can be estimated under foregoing conditions via sufficient samples ($d_1, d_2 \gg r^*$).

As a special case of this model, we are able to obtain a widely used flexible form of single index model (SIM) [9]–[11] with the s^*-sparse constraint by replacing the matrix $\mathbf{M}^*, \mathbf{X}$ by vectors $\boldsymbol{\beta}^*, \mathbf{x} \in \mathbb{R}^d$ as follows,

$$y = f(\mathbf{x}^T\boldsymbol{\beta}^*, \epsilon), \text{ subject to } \quad |\operatorname{supp}(\boldsymbol{\beta}^*)| = s^* \leq s, \quad \text{(I.2)}$$

with $\boldsymbol{\beta}^*$ recovered from noisy high-dimensional environments. The definition above shows the developmental aspect of our idea from SIM. More specifically, by setting the link function $f(\cdot, \epsilon)$ to $|\cdot|^2 : \mathbb{R} \to \mathbb{R}$, we can attain the original generalized phase retrieval (PR) problem which recovers signals $\boldsymbol{\beta}^*$ of interest from n moduli $|\langle \boldsymbol{\beta}^*, \mathbf{x}_i \rangle|, i = 1, 2, ..., n$ measured by n *i.i.d.* measurement vectors $\mathbf{x}_i$ with some distribution assumption [12]. On the other hand, as a line of work [13]–[15] has developed algorithms and provided theoretical guarantees of recovering a low-rank matrix from linear measurements, generalizing such settings to agnostic link functions is another key initial inspiration for our viewpoints. In this paper, we will focus on model (I.1) with real inputs and measurements, while the generalization of this model to the complex cases is straightforward by dealing with the real and complex parts of the variables separately.

A. Motivation

Motivating applications of retrieving low-rank matrix parameter $\mathbf{M}^*$ without any specific form of the link function arise across a wide range of diverse problems. For example, in low-rank matrix sensing problem [16], it is of interest to estimate $\mathbf{M}^*$ of the lowest rank from n observations $y_i = \operatorname{Tr}(\mathbf{M}^*\mathbf{X}_i) = \langle \mathbf{M}^*, \mathbf{X}_i \rangle, \forall i \in n$. For the widespread generalized linear models [17], [18], the underlying signal matrix is recovered through nonlinear link functions.

Therefore, it is desired to design an effective method for matrix phase retrieval (MPR) with both computational efficiency and statistical guarantees under general link functions (*i.e.*, agnostic nonparametric components). Related methods for traditional phase retrieval and SIM have been well developed, where the corresponding theoretical guarantees

Lewis (Miaofeng) Liu is with MILA & DIRO, Université de Montréal, Quebec, Canada.

Songtao Lu is with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA. (email: songtao@ibm.com)

Tuo Zhao is with the School of Industrial and Systems Engineering, Georgia Tech, Atlanta, GA 30332, USA.

Zhaoran Wang is with the Department of Industrial Engineering and Management Sciences, Northwestern University.

heavily depend on specific structures of the problems, e.g., assumptions/constraints on $f(\cdot,\epsilon)$ and $\mathbf{X}$. In the vector cases, previous works can be extended to solve agnostic phase retrieval problem in the case that the estimated vector is sparse. A direct way to estimate $\boldsymbol{\beta}^*$ is using the nonlinear least squares regression [19] in a nonconvex optimization regime. Deriving optimal estimator via convex relaxation was proposed in [20] under a sub-Gaussian assumption, while the strategy is only valid for a specific form of $f(\cdot,\epsilon)$. Another line of gradient descent based work follows seminal insights as Wirtinger Flow (WF) in [21]–[23], where thresholded and truncated Wirtinger Flow are applied and it has been shown that both of them can achieve near-optimal statistical accuracy respectively.

However, these works are unable to be applied to matrices directly. Although a low-rank matrix recovery problem via nonlinear link functions is considered in [18], the approach is confined within some specific properties of $f(\cdot,\epsilon)$ such as differentiability and monotonicity. The applications of other gradient-based methods, e.g., factorized gradient descent (FGD), multiple invocations of exact singular value decompositions (SVDs) [24] are also limited in practice due to the similar issues. With respect to a recent setting for low-rank matrix retrieval, a more generic formulation was considered in [12] by projecting the magnitude of each column in $\mathbf{M}^*$ separately and the corresponding algorithm is able to recover the ground truth in a near-optimal rate, but the link function considered in this work is specified. Therefore, the results are hard to be extended to general/agnostic cases.

To summarize, the existing works have either a lack of general assumptions on $f(\cdot,\epsilon)$ or an absence of generalizing the sparsity of vectors to the sparsity of the spectrum (*i.e.*, low rank) of matrices with provable guarantees. To overcome these defects, a spectrum truncation power (STPower) algorithm is proposed and its corresponding convergence behaviors are quantified in this paper. To the best of our knowledge, it is the first algorithm that solves AMPR with a linear convergence rate for the optimization error of problem (I.1) and near-optimal statistical rate.

B. Related Work

There is a considerably extensive works of studying phase retrieval model and SIM in the areas of machine learning [7], applied mathematics [16], [19], signal processing [12], [20], [25], etc. Here, we only survey the works that are most related. Recently, compared with the low-dimensional and non-sparse phase retrieval problems studied in [21], [26], [27], high-dimensional sparse signal recovery has sparkled significant attention in massive data processing. By adopting existing optimization methods, recent works [28], [29] incorporate the sparsity of the lifted matrix $\boldsymbol{\beta}^*\boldsymbol{\beta}^{*\top}$ to design ℓ_1-regularized phase retrieval by semidefinite programming and obtain reasonable good solutions by applying iterative efficient algorithms. For example, AltMinPhase [30] first performs the spectral initialization and then uses the alternating minimization algorithm to solve the sparse PR problem.

One class of the most popular algorithms proposed in [23], [31]–[33] basically is to truncate both the gradient flow by keeping the first k largest absolute values at each iterate so that an effective descent direction can be found on the sparse support. For example, thresholded wirtinger flow (TWF) [22] restricts the value of updated estimator by a threshold function based on the WF method. The main idea of these algorithms is just to keep the sparsity of the target vector. Similar truncation operations have been extensively developed in [34], where the truncation methods for solving PR problems are analyzed in [22] and [23] respectively.

Another line of works in the SIM related literature [35]–[37] shows that the least squares with ℓ_1-regularization can solve problem (I.2) under condition $\mathrm{Cov}(f(U,\epsilon),U^2)\neq 0$ with an excessive risk bounds, where $\mathrm{Cov}(\cdot,\cdot)$ denotes the correlation. Combining the U-process loss function, the method proposed in [38] can also solve this problem in high-dimensional settings. However, none of these previous works considered the low-rank property of the matrix form PR.

Low-rank phase retrieval [12], [39] and low-rank matrix sensing [14] are two matrix-based estimation problems which are also relevant to our work in the sense of matrix variables. The problem of low-rank phase retrieval is as follows: given a set of n measurements in the form $y_{i,j}:=|\mathbf{x}_{i,j}^\top\mathbf{m}_j|^2$ $(i=1,2,...,n,\ j=1,...,d_2)$ for each column $\mathbf{m}_j$ of $\mathbf{M}^*$ separately, the goal is to recover $\mathbf{x}$ by some non-convex algorithms such as QR factorization and alternative TWF methods. While, matrix sensing seeks to recover rank-r^* matrix $\mathbf{M}^*$ from measurements by the form of $y_i=\langle\mathbf{M}^*,\mathbf{X}_i\rangle$. Note that it serves as a special case of our model when $f(z,\epsilon)\equiv z$. A more general matrix sensing problem is nonlinear affine rank minimization, which could be recently solved by a variant of projected gradient descent methods (*a.k.a.* MAPLE) in [18] under the assumption that the link function must be differentiable and monotonic. To further strengthen the generality of our considered model, a comparison of this work with other most related existing works is summarized in Table I. Note that in the matrix multi-index model (MIM) [40], such as sufficient dimensionality reduction, the link function is a multivariate function with respect to measurements regardless of noise, *i.e.*, $y=f(\cdot,\ldots,\cdot,\epsilon)$, which is out of the scope of this work as shown in (I.1).

C. Main Contributions

In this paper, the proposed STPower algorithm is greatly inspired by power iteration methods [41], to which our truncation on the singular values is novel and different. To the best of our knowledge, there is no existing work that incorporates the low-rank formulation into the agnostic SIM regime and STPower is the first spectrum truncation iterative scheme that solves problem (I.1) with provable theoretical convergence rate guarantees. Main contributions of this work are highlighted as follows:

1) Leveraging the low-rank structure of the underlying PR model, we propose a new spectrum truncation algorithm to solve AMPR problem by applying computationally efficient power iteration.

TABLE I: Comparison of problem settings with other related works.

Method	$f(u)$	$\mathbf{X}$	Measurement [1]
LRPR [12]	$\|u\|^2$	Gaussian	$\langle \mathbf{x}_{i,k}, \mathbf{m}_k^* \rangle$ (column-wise) $i \in [n]$, $k \in [d_2]$
MAPLE [18]	$0 < f'(u) < C$, differentiable monotonic	sub-Gaussian	$\langle \mathbf{X}_i, \mathbf{M}^* \rangle$ $i \in [n]$
TPower [34]	$\|u\|^2$	Guassian	$\langle \mathbf{x}_i, \mathbf{m}^* \rangle$ $i \in [n]$
STPower (this work)	second-order distributionally differentiable	sub-Gaussian	$\langle \mathbf{X}_i, \mathbf{M}^* \rangle$ $i \in [n]$

We have $\mathbf{M}^* := [\mathbf{m}_1^*, \mathbf{m}_2^*, ..., \mathbf{m}_{d_2}^*] \in \mathbb{R}^{d_1 \times d_2}$, where $\mathbf{m}_k^*$ denotes the k-th column of $\mathbf{M}^*$. We denote a vector by a bold lower case letter (e.g., $\mathbf{x}$), and a matrix by a bold upper case letter (e.g., $\mathbf{X}$).

2) Under very mild assumptions on measurement matrix $\mathbf{X}$ and unknown function $f(\cdot, \epsilon)$, our estimator is shown to achieve a near-optimal statistical rate with a linear convergence rate for optimization error, implying the strong low-rank recovery guarantees. Multiple experiments verify our theoretical findings.

3) The considered AMPR model is very general, which can be applied to recover the desired variables in a wide class of PR problems, especially for the case where the prior knowledge of the nonparametric model is inaccurate or partially unknown.

D. Notation

In the following, we outline most of the commonly used notations for convenience of discussion. If not specified, $\text{vec}(\mathbf{M})$ denotes a vector obtained by concatenating the columns of $\mathbf{M}$, otherwise it can also be realized by concatenating the rows and transposing. We define:

$$\lambda_{\max}(\mathbf{A}, r) = \max_{\mathbf{M} \in \mathbb{R}^{d_1 \times d_2}} \text{vec}(\mathbf{M})^\top \mathbf{A}\, \text{vec}(\mathbf{M}), \tag{I.3a}$$

$$\text{subject to } \|\mathbf{M}\|_F = 1,\ \text{rank}(\mathbf{M}) \le r, \tag{I.3b}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of matrices. Let $\mathcal{S}^p = \{\mathbf{A} \in \mathbb{R}^{p \times p} | \mathbf{A} = \mathbf{A}^\top\}$ stand for the set of symmetric matrices. For any $\mathbf{A} \in \mathcal{S}^p$, we denote its eigenvalues by $\lambda_{\min}(\mathbf{A}) = \lambda_p(\mathbf{A}) \le \ldots \le \lambda_1(\mathbf{A}) = \lambda_{\max}(\mathbf{A})$. We denote by $\rho(\mathbf{A})$ the spectral norm of $\mathbf{A}$, and define $\rho(\mathbf{A}, r) = \max\{|\lambda_{\max}(\mathbf{A}, r)|, |\lambda_{\min}(\mathbf{A}, r)|\}$. In addition, the set $\{1, ..., n\}$ is denoted by $[n]$. Let $\text{basis}(\mathbf{M}) := \{i_1, i_2, ..., i_r\}$, where columns indexed by $\{i_1, i_2, ..., i_r\} \subseteq [d_2]$ form a set of maximum independent vectors in $\mathbb{R}^{d_1}$ and $\text{rank}(\mathbf{M}) = r$. Note that there may be multiple choices of basis($\mathbf{M}$) for a fixed $\mathbf{M}$ and we can take any of them for a specific need. Without loss of generality, we also postulate that $d_1 \le d_2$, otherwise we can transpose $\mathbf{M}$ and $\mathbf{X}$ without altering the results. Moreover, we denote $\mathbf{A}_\mathcal{B}$ as the restriction of $\mathbf{A}$ on the rows and columns induced by basis index set $\mathcal{B}$ ($|\mathcal{B}| = r$) of matrix $\mathbf{M}$, that is, r columns indexed by $\mathcal{B}$ hold the $\text{vec}(\mathbf{M})$'s rd_2 elements (where we suppose $d_1 \le d_2$), to which the indexes for the columns and rows of $\mathbf{A}$ are corresponding to each other. Given a rank restriction of $\mathbf{M}$ by an index set $\mathcal{B}$, we define

$$\mathbf{M}(\mathcal{B}) := \operatorname*{argmax}_{\mathbf{M} \in \mathbb{R}^{d_1 \times d_2}} \text{vec}(\mathbf{M})^\top \mathbf{A}\, \text{vec}(\mathbf{M}), \tag{I.4a}$$

$$\text{subject to } \|\mathbf{M}\|_F = 1,\ \text{basis}(\mathbf{M}) \subseteq \mathcal{B}. \tag{I.4b}$$

Given two sequences of random variables $\{X_n\}$ and $\{Y_n\}$, We denote by $X_n = \mathcal{O}_p(Y_n)$ that the sequence of values X_n/Y_n is stochastically bounded, *i.e.*, for any $\epsilon > 0$, there exists a finite $M > 0$ and a finite $N > 0$ such that, $P(|X_n/Y_n| > M) < \epsilon, \forall n > N$. There will be some sets of interested matrices, denoted by $\mathcal{S}_{\mathbf{M}^{d_1 \times d_2}} = \{\mathbf{M} \in \mathbb{R}^{d_1 \times d_2} \mid \|\mathbf{M}\|_F = 1\}$ and $\mathcal{B}_{\mathbf{M}(r)} = \{\mathbf{M} \in \mathbb{R}^{d_1 \times d_2} \mid \text{rank}(\mathbf{M}) \le r\}$. Other notations will be introduced when they are used.

II. Models and Algorithms

In this section, we will formulate the MPR problem (I.1) into an estimation of the largest eigenvalue of a empirical expectation of a random matrix under some rank restriction. After that, we will present the proposed spectrum truncation iterative algorithm in details.

A. Eigenvector Estimator for AMPR

Now we first clarify the motivation for our estimators of matrix parameter $\mathbf{M}^*$. Then we will derive the formal formulation of AMPR and present the assumptions of the link function $f(\cdot, \epsilon)$ and the generality of applicable settings.

The main idea of developing this estimator is inspired by the second-order Stein's identity [42]. For completeness, we illustrate the first-order Stein's identity with constraints on the first-order derivative [36] for SIMs as follows.

1) First-Order Stein's Identity for SIMs:

Proposition II.1 (First-order Stein's Identity [43])**.** Suppose that $\mathbf{X} \in \mathbb{R}^{d_1 \times d_2}$ is a real-valued random matrix with differentiable probability density function $g : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}$. For a continuous function $h : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}$ with existing $\mathbb{E}[\nabla h(\mathbf{X})]$, the following identity holds:

$$\mathbb{E}[h(\mathbf{X})S(\mathbf{X})] = \mathbb{E}[\nabla h(\mathbf{X})], \tag{II.1}$$

where $S(\mathbf{X}) = -\nabla g(\mathbf{X})/g(\mathbf{X})$ serves as the score function of $g(\cdot)$.

When the above proposition is applied to the AMPR model, we can obtain a direct estimator $\mathbf{M}^*$ by the chain rule. Letting $h(\mathbf{X}) = f(\langle \mathbf{M}^*, \mathbf{X} \rangle, \epsilon)$ in II.1, we have

$$\mathbb{E}(yS(\mathbf{X})) = \mathbb{E}[\partial_{x_1} f(\langle \mathbf{M}^*, \mathbf{X} \rangle, \epsilon)]\mathbf{M}^*, \tag{II.2}$$

where ∂_{x_1} denotes the partial derivative of function $f(x_1, x_2)$ with respect to variable x_1. In this way, we are able to directly estimate $\mathbf{M}^*$ by $\mathbb{E}(yS(\mathbf{X}))$ as we have set $\|\mathbf{M}^*\|_F =$

1 for scaling invariance. Unfortunately, it is required that $\mathbb{E}[\partial_{x_1} f(\langle \mathbf{M}^*, \mathbf{X}\rangle, \epsilon)] \neq 0$, which keeps the application of the identity off a wide range of models, for example, where $f(\cdot, \epsilon)$ is quadratic in the first variable for the phase retrieval problem. Thus, we are motivated to check the second-order information in $f(\cdot, \epsilon)$ for a milder assumption.

2) Second-Order Stein's Identity for SIMs:

Proposition II.2 (Second-Order Stein's Identity [42])**.** Let the probability density function $g(\cdot)$ of $\mathbf{X}$ be twice differentiable. Then for any second-order distributionally differentiable function $h : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}$ with existing $\mathbb{E}[D^2 h(\mathbf{X})]$, it follows that

$$\mathbb{E}[h(\mathbf{X})T(\mathbf{X})] = \mathbb{E}[D^2 h(\mathbf{X})], \tag{II.3}$$

where the second-order score function $T : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^{d_1 d_2 \times d_1 d_2}$ is defined as $T(\mathbf{X}) = \nabla^2 g(\mathbf{X})/g(\mathbf{X})$.

For our MPR model, setting $h(\mathbf{X}) = f(\langle \mathbf{M}^*, \mathbf{X}\rangle, \epsilon)$ in (II.3), we have

$$\mathbb{E}[yT(\mathbf{X})] = \mathbb{E}[f(\langle \mathbf{M}^*, \mathbf{X}\rangle, \epsilon) \cdot T(\mathbf{X})] \tag{II.4}$$
$$= 2\mathbb{E}[\partial^2_{x_1} f(\langle \mathbf{M}^*, \mathbf{X}\rangle, \epsilon)] \operatorname{vec}(\mathbf{M}^*) \operatorname{vec}(\mathbf{M}^*)^T, \tag{II.5}$$

where $\partial^2_{x_1}$ refers to the second-order partial derivative of function $f(x_1, x_2)$ with respect to variable x_1. Hence we can extract $\mathbf{M}^*$ from the second-order Stein's identity in a variety of situations when the first-order identity staggers. Explicitly in this context, $\mathbb{E}[\partial^2_{x_1} f(\langle \mathbf{M}^*, \mathbf{X}\rangle, \epsilon)]$ is supposed to be nonzero. Without loss of generality, it suffices to assume $\mathbb{E}[\partial^2_{x_1} f(\langle \mathbf{M}^*, \mathbf{X}\rangle, \epsilon)] > 0$, otherwise the sign of y can be flipped and $f(\cdot, \epsilon)$ is changed to $-f(\cdot, \epsilon)$. Such restricted function $f(\cdot, \epsilon)$ is called a second-order link function in [44]. Intuitively, as the information of $\operatorname{vec}(\mathbf{M}^*)$ is included in the second-order cross moments, we will call $\mathbf{A}^* = \mathbb{E}[yT(\mathbf{X})]$ the *second-order link matrix*. When $\operatorname{vec}(\mathbf{X}) \sim \mathcal{N}(0, \mathbf{I}_{d_1 d_2})$, we have

$$2\mathbb{E}[\partial^2_{x_1} f(\langle \mathbf{M}^*, \mathbf{X}\rangle, \epsilon)] \operatorname{vec}(\mathbf{M}^*) \operatorname{vec}(\mathbf{M}^*)^T \tag{II.6}$$
$$\overset{(a)}{=} \mathbb{E}[f(\langle \mathbf{M}^*, \mathbf{X}\rangle, \epsilon)(\operatorname{vec}(\mathbf{X}) \operatorname{vec}(\mathbf{X})^\top - \mathbf{I}_{d_1 d_2})] \tag{II.7}$$
$$\overset{(b)}{=} \mathbb{E}[y(\operatorname{vec}(\mathbf{X}) \operatorname{vec}(\mathbf{X})^\top - \mathbf{I}_{d_1 d_2})] \tag{II.8}$$

where in (a) we utilize the density function of Gaussian distribution, and (b) is true due to the definition of the model, i.e., $y = f(\langle \mathbf{M}^*, \mathbf{X}\rangle, \epsilon)$. As the measurement matrix $\mathbf{X}$ can usually be generated by the observer, such *i.i.d.* Gaussian distribution is widely applied for a concise form of (II.4). Motivated by (II.4), we need to obtain the leading eigenvector of the sampled version of $\mathbf{A}^*$: $\mathbf{A} = 1/n \sum_{i=1}^n [y_i T_i(\mathbf{X})]$, which is a vectorization of the solution. Then, we decompose $\mathbf{A}$ by $\mathbf{A} = \mathbf{A}^* + \mathbf{E}$, where $\mathbf{E}$ is a random perturbation matrix due to a finite number of empirical samples. To recover $\mathbf{M}^*$ (*i.e.*, $\operatorname{vec}(\mathbf{M}^*)$.) from the noisy observation $\mathbf{A}$ when the perturbation $\mathbf{E}$ is relatively small, we need to enforce the rank of $\mathbf{M}$ through the estimation procedure under a tunable preset level r_0 ($r_0 \geq r^*$ implicitly). Using this idea, we formulate the estimation problem in the following new form.

Definition II.3 (the largest r_0-rank eigenvalue problem)**.**

$$\bar{\mathbf{M}}^* = \underset{\mathbf{M} \in \mathbb{R}^{d_1 \times d_2}}{\operatorname{argmax}} \operatorname{vec}(\mathbf{M})^\top \mathbb{E}[yT(\mathbf{X})] \operatorname{vec}(\mathbf{M}), \tag{II.9a}$$
$$\text{subject to } \|\mathbf{M}\|_F = 1,\ \operatorname{rank}(\mathbf{M}) \leq r_0. \tag{II.9b}$$

Definition II.3 enables us to solve the AMPR problem (I.1) by a rank-restricted large eigenvalue problem (II.9). Furthermore, we remark that the information of the unknown link function f is encoded in the coefficient of $\operatorname{vec}(\bar{\mathbf{M}}^*) \operatorname{vec}(\bar{\mathbf{M}}^*)^T$ in (II.5) as a scaling factor. Hence due to normalization $\|\mathbf{M}\|_F = 1$, our estimation does not depend on f anymore. It is possible to solve (II.9) by exhaustively enumerating subsets of $\{1, ..., d_1\}$ or $\{1, ..., d_2\}$ to obtain $r_0 \times r_0$ principle matrix $\mathbf{A}_{r_0}$'s and pick up the largest $\lambda_{\max}(\mathbf{A}_{r_0})$. However, this is intractable due to expensive computational cost. Inspired by [23], [34], we propose a *spectrum truncation power iteration* to guarantee a low-rank structure of the matrix efficiently.

B. Spectrum Truncation Power Iteration

We now present the iterative procedure by leveraging the standard power iteration method for rank-restricted eigenvalue problems. The details of implementing the proposed algorithm is shown in Algorithm 1 formally, which produces a sequence of intermediate at most r_0-rank parameter matrix $\mathbf{M}^{(0)}, \mathbf{M}^{(1)}, ...$ We split the procedure into three main steps: 1) at time step t, the vectorized iterate $\operatorname{vec}(\mathbf{M}^{(t)})$ of $\mathbf{M}^{(t)}$ is multiplied by $\mathbf{A}$; 2) the spectrum of $\mathbf{M}^{(t+0.5)}$ is truncated by holding out the surplus non-zero singular values so that the matrix is approximately below rank r_0; 3) the truncated matrix is normalized by its Frobenius norm. The truncation operator is defined as the following.

Definition II.4 (Spectrum Truncation Operator)**.** Given a matrix $\mathbf{M} \in \mathbb{R}^{d_1 \times d_2}$ of rank r and its singular value decomposition in the form: $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$, where $\mathbf{U} \in \mathbb{R}^{d_1 \times d_1}$ and $\mathbf{V} \in \mathbb{R}^{d_2 \times d_2}$ are unitary, $\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_r & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{d_1 \times d_2}$ and $\mathbf{\Sigma}_r = \operatorname{diag}(\sigma_1, ..., \sigma_r) \in \mathbb{R}^{r \times r}$. Here, $\sigma_1 \geq ... \geq \sigma_r > 0$ are the ordered positive singular values of $\mathbf{M}$. Then the spectrum truncation operator $\mathcal{T}_{r_0}$ gives: $\mathcal{T}_{r_0}(\mathbf{M}) = \mathbf{U}\mathbf{\Sigma}'\mathbf{V}^*$, where $\mathbf{\Sigma}' = \begin{bmatrix} \mathbf{\Sigma}_{r_0} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{d_1 \times d_2}$ and $\mathbf{\Sigma}_{r_0} = \operatorname{diag}(\sigma_1, ..., \sigma_{r_0}) \in \mathbb{R}^{r_0 \times r_0}$.

Through the truncation, we obtain another immediate iterate $\widehat{\mathbf{M}}^{(t+1)}$ as a low-rank projection of $\mathbf{M}^{(t+0.5)}$. Note that when the rank of $\mathbf{M}^{(t+0.5)}$ is no more than r_0, all the positive singular values of $\mathbf{M}^{(t+0.5)}$ are kept intact.

In case of scarce knowledge of the target rank, the truncation level is to be tuned, which is proved to be feasible by Theorem III.2. The computational complexity of the proposed algorithm at each iteration is $\mathcal{O}(r_0(d_1 d_2)^2)$(where $d_1 \leq d_2$) with the fast low-rank principal component analysis (PCA) implementation [45]. Compared to former truncated power methods for sparse vectors, we save the labor on selecting the truncation set as the SVD result has arranged the r_0 largest singular values in the top left corner of $\mathbf{\Sigma}$. Another benefit brought by power methods is its computationally efficient

Algorithm 1 Spectrum Truncation Power (STPower) Iteration

Input: second-order link matrix $\mathbf{A} \in \mathbb{S}^{d_1 d_2}$, initial matrix $\mathbf{M}^{(0)} \in \mathbb{R}^{d_1 \times d_2}$
Output: Estimation of $\mathbf{M}^*$: $\overline{\mathbf{M}} \in \mathbb{R}^{d_1 \times d_2}$
Parameters: rank restriction parameter r_0, accuracy for convergence verification η
$t \leftarrow 0$.
Repeat
  Power Iteration:

$$\mathrm{vec}(\mathbf{M}^{(t+0.5)}) = \mathbf{A}\,\mathrm{vec}(\mathbf{M}^{(t)})/\|\mathbf{A}\,\mathrm{vec}(\mathbf{M}^{(t)})\| \tag{II.10}$$

  Spectrum Truncation: $\widehat{\mathbf{M}}^{(t+1)} \leftarrow \mathcal{T}_{r_0}(\mathbf{M}^{(t+0.5)})$
  Normalization: $\mathbf{M}^{(t+1)} \leftarrow \widehat{\mathbf{M}}^{(t+1)}/\|\widehat{\mathbf{M}}^{(t+1)}\|_F$
  Update: $t \leftarrow t+1$
Until $\|\mathbf{M}^{(t)} - \mathbf{M}^{(t-1)}\|_F < \eta$
Output: $\overline{\mathbf{M}} \leftarrow \mathbf{M}^{(t)}$

parallel implementation on multiple computing resources, e.g., large-scale clusters. As a result, the estimation process can be considerably accelerated in such practical scenarios. In addition, due to the normalization step in our algorithm, we can only recover the target matrix up to the Frobenius norm, which is similar to the recovery of unknown signals up to magnitudes in [46] without knowledge of link functions.

III. Theoretical Guarantees

In this section, we will present the main theoretical convergence results of STPower, including the linear convergence rate with respect to optimization error under Gaussian measurements (note that a general distribution is also allowed for the measurement matrix, e.g., sub-Gaussian). The results will give a clear demonstration of the low-rank recovery performance of STPower with applications to AMPR models. In addition, theoretical insights help refine the details in our algorithm implementation.

A. Convergence Guarantees

Before showing the theorems, we first define the gap between the largest and the other eigenvalues: $\delta\lambda = \lambda_{\max}(\mathbf{A}^*) - \max_{j>1}|\lambda_j(\mathbf{A}^*)|$. Also, the eigenvector $\mathbf{v}(\lambda_{\max})$ of the largest eigenvalue $\lambda = \lambda_{\max}(\mathbf{A}^*)$ corresponds to low-rank matrix $\mathbf{M}^*$ with rank r^*. We need the further assumption below.

Assumption III.1. We assume the eigenvalues of $\mathbf{A}$ are non-degenerate.

Due to the randomness in obtaining $\mathbf{A}$, such an assumption is reasonable. The following result quantifies the strong low rank recovery performance and the convergence rate of the STPower iterative algorithm.

Theorem III.2. Under Assumption III.1, assume $\rho(\mathbf{E}, r) < \frac{\delta\lambda}{2}$. Let $r_0 \geq r^*$ and $r = 2r_0 + r^*$. We define

$$\delta(r) := \frac{\sqrt{2}\rho(\mathbf{E}, r)}{\sqrt{\rho(\mathbf{E}, r)^2 + (\delta\lambda - 2\rho(\mathbf{E}, r))^2}}, \tag{III.1}$$

$$\gamma(r) := \frac{\lambda - \delta\lambda + \rho(\mathbf{E}, r)}{\lambda - \rho(\mathbf{E}, r)} < 1. \tag{III.2}$$

If the initial matrix iterate $\mathbf{M}^{(0)}$ admits $\|\mathbf{M}^{(0)}\|_F = 1$, the target matrix $\mathbf{M}^*$ satisfy $|\langle \mathbf{M}^{(0)}, \mathbf{M}^* \rangle| \geq \delta(r) + \omega$, and $\omega \in (0, 1)$ such that

$$\beta = \sqrt{\left(1 + \frac{2\sqrt{r^*}}{\sqrt{r_0 - r^*}}\right) \cdot \left(1 - \frac{(1-\gamma(r)^2)\omega(1+\omega)}{2}\right)} < 1 \tag{III.3}$$

serving as a coefficient of contraction mapping, then we have either $\|\mathbf{M}^{(0)} - \mathbf{M}^*\|_F < 2\delta(r)/(1-\beta)$ or

$$\|\mathbf{M}^{(t)} - \mathbf{M}^*\|_F \leq \beta^t \|\mathbf{M}^{(0)} - \mathbf{M}^*\|_F + \frac{2\delta(r)}{1-\beta},\ \forall t \geq 0. \tag{III.4}$$

Proof: See Appendix A for a detailed proof. ∎

Theorem III.2 states the essential relation between a linear convergence behavior of STPower and corresponding statistical error, which are related to quantity $\rho(\mathbf{E}, r)$. It can be observed that when the rank restricted norm $\rho(\mathbf{E}, r)$ falls below a half of the eigen-gap (the gap between the second largest eigenvalue in absolute value and the largest eigenvalue of $\mathbf{A}$), it holds that $\gamma(r) < 1$ and $\delta(r) = \mathcal{O}(\rho(\mathbf{E}, r))$, where we ignore the constants in the discussion. Note that for any $r_0 > r^*$, if $\gamma(r)$ is sufficiently small then we can meet the requirement on β defined in (III.3) by a sufficiently small ω of the order $\mathcal{O}((r^*/r_0)^{1/2})$. In particular, we have the following remark characterizing the dependency of r_0 on r^* under the impact of the eigenvalue gap of the second-order link marix $\mathbf{A}^*$ and the sample size n, while combining with Theorem III.6 on the sample complexity.

Remark III.3. For $c \in (0, 1)$, when

$$n = \Omega_p\left(\left(\frac{1+c}{\delta\lambda + (1+c)\lambda}\right)^2 r \max(d_1, d_2) \log\max(d_1, d_2)\right),$$

we have

$$\rho(\mathbf{E}, r) < \frac{\delta\lambda + (1+c)\lambda}{1+c}, \quad \text{and} \quad \gamma(r) < c. \tag{III.5}$$

According to the definition of β in (III.3), if we also set

$$r_0 > \left(1 + \left(\frac{2}{1-c^2} - 2\right)^2\right) \cdot r^*, \tag{III.6}$$

then, there exists an $\omega \in (0, 1)$ such that we have β defined in Theorem III.2 to satisfy $\beta < 1$. Here we assume the constants in notations Ω_p and $\mathcal{O}_p$ to be 1 for simplicity.

If the rank restricted norm is small enough, the initialization restriction will be easy to achieve, and experiments show that even a random initialization with *i.i.d.* Gaussian entries works. Otherwise, we can regulate r_0 to be greater first and keep reducing r_0 by rerunning the algorithm with an initial value from the result of the last run. In the next paragraph, we will elaborate this issue in more details.

Suppose that $\max_{i,j}|\mathbf{M}^*_{i,j}|$ is sufficiently large and $(k, l) = \mathrm{argmax}_{i,j}|\mathbf{M}^*_{i,j}|$, then we initialize $\mathbf{M}^{(0)} = \mathbf{G}_{k,l}$, where $\mathbf{G}_{k,l}$ is the matrix of 0's except a 1 at (k, l). In this way $|\langle \mathbf{M}^{(0)}, \mathbf{M}^* \rangle| = |\mathbf{M}^*_{k,l}|$ is sufficiently large such that $|\langle \mathbf{M}^{(0)}, \mathbf{M}^* \rangle| \geq C'\rho(\mathbf{E}, r)$ holds with $r_0 = \mathcal{O}(r^*)$ for some constant C'. However, if we fail to initially make $|\langle \mathbf{M}^{(0)}, \mathbf{M}^* \rangle|$ large enough, we can regulate r_0 to be greater for a larger

$\rho(\mathbf{E}, r)$ to meet the initial condition of $\mathbf{M}^{(0)}$, which gives a larger $\rho(\mathbf{E}, r)$ and $\mathbf{M}^{(t)}$ may not converge to $\mathbf{M}^*$ accurately. Fortunately, as long as $|\langle \mathbf{M}^{(t)}, \mathbf{M}^* \rangle|$ converges to a value that is not too small (may be much larger than $|\langle \mathbf{M}^{(0)}, \mathbf{M}^* \rangle|$), we may reduce r_0 and rerun the algorithm with a r_0-rank truncation of $\mathbf{M}^{(t)}$ as the initial vector. We can use two stages for the initialization, *e.g.*, 1) we can run STPower with $r_0 = d_1 d_2$ by random initialization where STPower reduces to power method which gives the dominant eigenvector $\mathrm{vec}(\mathbf{M})$ of $\mathbf{A}$, 2) r_0 is reduced to r_0', following Algorithm 1, we have a new initial value $\mathbf{M}^{(0)} := \mathcal{T}_{r_0'}(\mathbf{M})/\|\mathcal{T}_{r_0'}(\mathbf{M})\|$, then Lemma A.5 (which will be presented later) implies $\langle \mathbf{M}^{(0)}, \mathbf{M}^* \rangle \geq \delta(r)$.

After the conditions of initialization and rank restricted norm are satisfied, it follows that $\|\mathbf{M}^{(t)} - \mathbf{M}^*\|_F$ converges geometrically until $\|\mathbf{M}^{(t)} - \mathbf{M}^*\|_F = \mathcal{O}(\rho(\mathbf{E}, r))$, which indicates $\beta^t \|\mathbf{M}^{(0)} - \mathbf{M}^*\|_F \to 0$ due to the accumulative contraction leveraged by the power iteration. Such implication also inspires us to give Lemma A.3 in Appendix, which sharpens the results in terms of replacing the full matrix norm by a smaller $\rho(\mathbf{E}, r)$. Similar result for the vector case was also shown in [34] by using the eigenvalue perturbation theory [47]. We remark that our simple yet effective rank restricted norm actually encodes more structural information of the problem.

B. Concentration Property

Next, we proceed to present the concentration properties of replacing the population matrix $\mathbf{A}^*$ with its sampled version $\mathbf{A}$ below. The results are based on the following mild assumption. Recall that y is the scalar response of the data generating model, $T(\mathbf{X})$ is the second-order score function defined in Section II.2. We denote by $T(\mathbf{X})_{ij}$ the element at the i-th row and j-th column of the matrix $T(\mathbf{X}) \in \mathbb{R}^{d_1 d_2 \times d_1 d_2}$.

Assumption III.4. Let σ be the largest singular value of the covariance matrix $yT(\mathbf{X})$. Then there exists a constant $K \in \mathbb{R}$ such that for all $i, j \in [d_1 d_2]$, $|yT(\mathbf{X})_{ij}| \leq K$ and $\sigma \leq K$.

Note that such practical condition, which is also applicable to a large class of sub-Gaussian sampling matrices [48], is mild since there is no any assumption on the distribution of $\mathbf{X}$. The following lemma is a new crucial cover number result in the normalized rank-restricted matrix space.

Lemma III.5 (Cover Number in the Rank Restricted Matrix Space)**.** For the set $\mathcal{S}_{\mathbf{M}}^{d_1 \times d_2} \cap \mathcal{B}_{\mathbf{M}}(r)$ with the metric induced by the Frobenius norm, an ϵ-net $\mathcal{N}(\epsilon, r, d_1, d_2)$ is a subset such that for every point $\mathbf{M} \in \mathcal{S}_{\mathbf{M}}^{d_1 \times d_2} \cap \mathcal{B}_{\mathbf{M}}(r)$, there exists a point $\mathbf{M}_\epsilon \in \mathcal{N}(\epsilon, r, d_1, d_2)$ satisfying $\|\mathbf{M} - \mathbf{M}_\epsilon\|_F \leq \epsilon$. We denote the minimal cardinality of an ϵ-net of $\mathcal{S}_{\mathbf{M}}^{d_1 \times d_2} \cap \mathcal{B}_{\mathbf{M}}(r)$ as $N(\epsilon, r, d_1, d_2)$. Considering the independence of the columns, we have a bound of $N(\epsilon, r, d_1, d_2)$, i.e.,

$$N(\epsilon, r, d_1, d_2) \leq \left(\frac{2}{\epsilon} + 1\right)^{r d_1}. \tag{III.7}$$

In addition, for $0 \leq \epsilon < 1$ and $\mathbf{S}$ being a symmetric $d_1 d_2 \times d_1 d_2$ matrix, the following inequality also holds.

$$\max_{\mathbf{M}_1 \in \mathcal{N}(\epsilon, r, d_1, d_2)} |\,\mathrm{vec}(\mathbf{M}_1)^\top \mathbf{S}\,\mathrm{vec}(\mathbf{M}_1)| \geq (1 - 2\epsilon) \max_{\mathbf{M}_2 \in \mathcal{S}_{\mathbf{M}}^{d_1 \times d_2} \cap \mathcal{B}_{\mathbf{M}}(r)} |\,\mathrm{vec}(\mathbf{M}_2)^\top \mathbf{S}\,\mathrm{vec}(\mathbf{M}_2)|.$$

Proof: See Appendix B for a detailed proof. ∎

In the following, we will provide the concentration bound given a sufficiently large n.

C. Sample Complexity

The idea of showing this result is to stochastically bound the rank restricted norm of perturbation matrix $\mathbf{E}$ with problem dimensions r, d_1, d_2, and n.

Theorem III.6. In the AMPR model, under Assumption III.4, we have the perturbation formulation of matrix $\mathbf{A} \triangleq \mathbf{A}^* + \mathbf{E}$, where $\mathbf{A}^* = \mathbb{E}[y(\mathrm{vec}(\mathbf{X})\,\mathrm{vec}(\mathbf{X})^T - \mathbf{I})]$ and $\mathbf{A} = \frac{1}{n}\sum_{i=1}^n [y_i(\mathrm{vec}(\mathbf{X}_i)\,\mathrm{vec}(\mathbf{X}_i)_i^T) - \mathbf{I}]$ with $\{y_i, \mathbf{X}_i\}$ being independent samples drawn from $y = f(\langle \mathbf{M}^*, \mathbf{X} \rangle)$, and satisfying Assumption III.4. Then for sufficiently large r, d_1, d_2 and $n > \frac{r \max(d_1, d_2) \log \max(d_1, d_2)}{K}$ we have

$$\rho(\mathbf{A} - \mathbf{A}^*, r) = \mathcal{O}_p\left(\sqrt{\frac{r \max(d_1, d_2) \log \max(d_1, d_2)}{n}}\right), \tag{III.8}$$

where K is the some constant given in Assumption III.4.

Proof: See Appendix C for a detailed proof. ∎

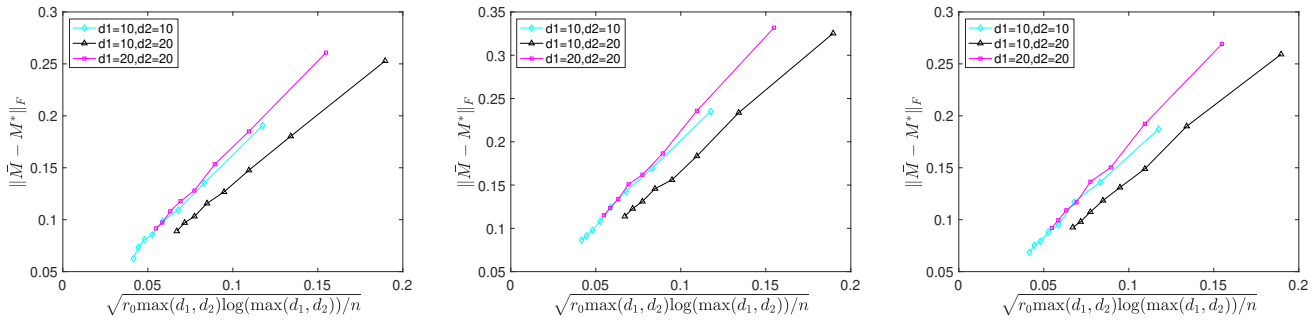
The other intermediate lemmas leading to the main results in details are presented in the appendix. Note that there are $r \max(d_1, d_2)$ independent variables of a low-rank matrix. The difference between the obtained rate and the lower bound of low rank matrix sensing [49] is up to $\mathcal{O}(\sqrt{\log \max(d_1, d_2)})$, therefore we claim that our algorithm achieves a near-optimal minmax statistical rate.

Actually, compared with Assumption III.4, much milder assumptions, i.e., $\|y_i T(\mathbf{X}_i)\|_2 \leq C$ almost surely and $\|\sum_i \mathbb{E}\mathbf{X}_i^2\|_2 \leq \sigma^2$ for some constants C and σ (which can be found in [50]), are enough to show the statistical rate (III.8) by applying the Bernstein inequality in Lemma A.6 to estimate a rough tail bound for $\rho(\mathbf{A} - \mathbf{A}^*, r)$.

In addition, our theoretical analysis further reveals the connection between the structure of the largest eigenvalue problems and the first- and second-order Stein's identities for a wide range of distributions for measurements in AMPR.

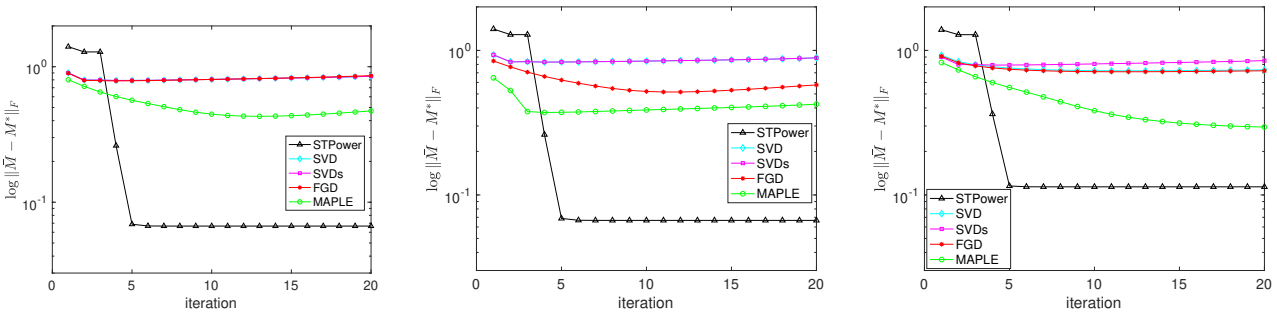
IV. Numerical Experiments

In this section, we show numerical results for three typical AMPR models and verify the theoretical finite-sample statistical error on the simulated data, as well as for a comparison to several matrix recovery methods with similar yet slightly different settings. STPower can deal with different link functions which are second-order differentiable under distributional derivatives.



$f_1(x,\epsilon)=x^2+\epsilon$ $f_2(x,\epsilon)=|x|+\epsilon$ $f_3(x,\epsilon)=x^2+\sin(x)+\epsilon$

Fig. 1: Frobenius distances between the true parameter $\mathbf{M}^*$ and the estimation $\overline{\mathbf{M}}$ in AMPR with link function in one of $f_1(\cdot,\epsilon)$, $f_2(\cdot,\epsilon)$, and $f_3(\cdot,\epsilon)$ in case of $d_1=d_2=10$, $d_1=10,d_2=20$, $d_1=d_2=20$, $r^*=3$, and varying sample size n.



$d_1=20,d_2=20,r^*=3$ $d_1=30,d_2=30,r^*=3$ $d_1=50,d_2=50,r^*=3$

Fig. 2: Comparison of different matrix recovery methods via convergence curves under logarithmic Frobenius distances measure. The link function is $f(x,\epsilon)=x^2+\sin(x)+\epsilon$, under different dimensions.

A. Experiments Setup

In this work, we test the convergence property of the proposed algorithm on the following three neat representative functions:

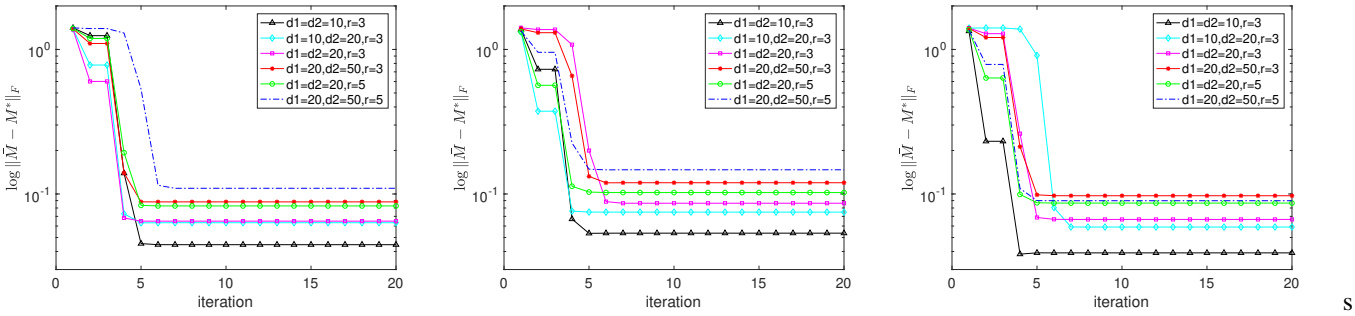
$$f_1(x,\epsilon)=x^2+\epsilon, f_2(x,\epsilon)=|x|+\epsilon, f_3(x,\epsilon)=x^2+\sin(x)+\epsilon,$$

where $f_1(x,\epsilon)$ accounts for the noisy phase retrieval model as traditional vector scenarios except for a rank restriction on the matrix $\mathbf{M}$, the absolute value in $f_2(x,\epsilon)$ keeps away the function from commonly recognized differentiable categories, $f_3(x,\epsilon)$ can be regarded as a robust extension of $f_1(x,\epsilon)$. We will concentrate on *i.i.d.* standard Gaussian design for each entry of $\mathbf{X}$, i.e., $\mathbf{X}_{ij}\sim\mathcal{N}(0,1)$, for a clearer interpretation of the convergence performance behaviour and the near-optimal statistical rate. The convergence behaviors of STPower under more complicated settings would be studied as the future work.

Since our theory requires the number of samples $n \gtrsim r\max(d_1,d_2)\log\max(d_1,d_2)/K, r \ll d_1,d_2$, we fix $d_1=d_2=20$, $r^*=3$ while varying n in an uniform range with a step-size of 1000. To generate a random r^*-rank $d_1\times d_2$ matrix $\mathbf{M}^*$ for the data model, we use two random matrices $\mathbf{M}_1\in\mathbb{R}^{d_1\times r^*}$ and $\mathbf{M}_2\in\mathbb{R}^{r^*\times d_2}$ with *i.i.d.* standard Gaussian entries, then let $\mathbf{M}'=\mathbf{M}_1\mathbf{M}_2\in\mathbb{R}^{d_1\times d_2}$ and normalize $\mathbf{M}'$ to obtain $\mathbf{M}^*$. We simply initialize the iterate $\mathbf{M}^{(0)}$ by the standard Gaussian random values. For guarantee of a good initial matrix, we use the warm-start strategy illustrated in section III.2. We also generate another independent second-order link matrix $\mathbf{A}_{val}$ to fine-tune r_0. Moreover, we utilize the Frobenius distance between the final output $\overline{\mathbf{M}}$ after a run of STPower and the aforementioned $\mathbf{M}^*$, *i.e.*, $\|\overline{\mathbf{M}}-\mathbf{M}^*\|_F$ as the measurement of the estimation error. For the convergence condition, we set $\eta=10^{-6}$. In addition, the truncation operation in our realization adopts a top-r_0 singular value decomposition for efficiency. Throughout the experiments, in a loop of n, we randomly draw n *i.i.d.* samples for $\mathbf{X}$ while computing the responses $\{y_i\}_{i=1}^n$ via $\mathbf{M}^*$ and $\epsilon\sim\mathcal{N}(0,1)$ according to (I.1) to obtain $\mathbf{A}$. Considering the randomness of the noise, we repeat the above procedure for $T=50$ times independently for each fixed n to quantify the average error $e_n=1/T\sum_{t=1}^T\|\overline{\mathbf{M}}_t-\mathbf{M}^*\|_F$ as the single result.

B. Numerical Results

We plot the Frobenius distance (estimation error) against statistical convergence rate $\sqrt{r\max(d_1,d_2)\log\max(d_1,d_2)/n}$ in Figure 1(a)–1(c) for each second-order link function. It demonstrates that the Frobenius estimation error is approximately tightly bounded within the same order of a linear function of the statistical rate, which justifies our main theoretical results. For a further



$f_1(x,\epsilon)=x^2+\epsilon$ $f_2(x,\epsilon)=|x|+\epsilon$ $f_3(x,\epsilon)=x^2+\sin(x)+\epsilon$

Fig. 3: Convergence curves under logarithmic Frobenius distances measure in AMPR with link function in one of $f_1(\cdot,\epsilon)$, $f_2(\cdot,\epsilon)$, and $f_3(\cdot,\epsilon)$ in case of diverse d_1, d_2, and r^* values.

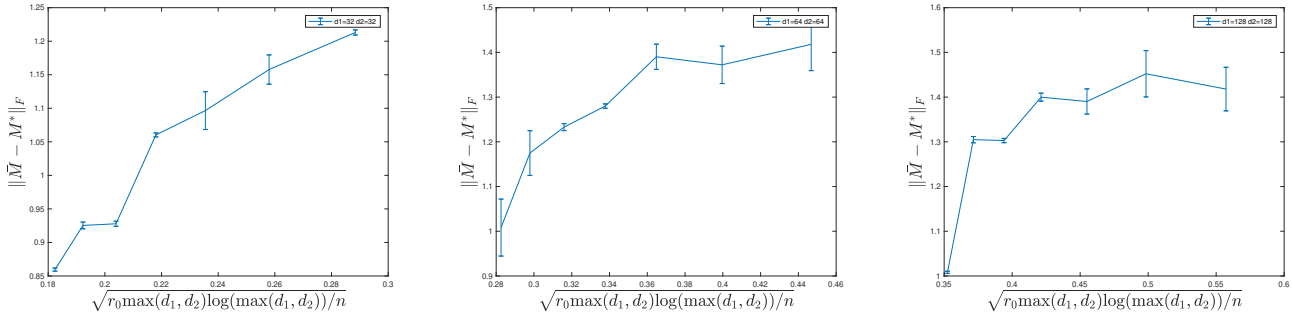


Fig. 4: Frobenius distances between the true parameter $\mathbf{M}^*$ and the estimation $\overline{\mathbf{M}}$ in AMPR with link function $f_1(\cdot,\epsilon)$, large dimensions of $d_1=d_2=32$, $d_1=d_2=64$, $d_1=d_2=128$, $r^*=3$, and varying sample size n.

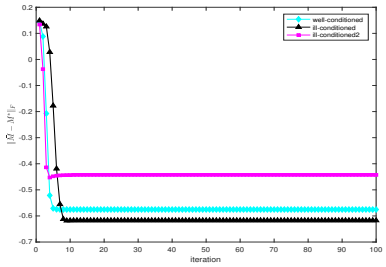


Fig. 5: Convergence curves under logarithmic Frobenius distances measure with link function $f_1(\cdot,\epsilon)$ and different condition numbers.

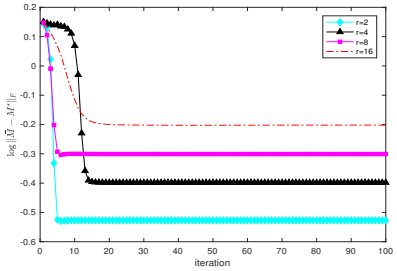


Fig. 6: Convergence curves under logarithmic Frobenius distances measure with link function $f_1(\cdot,\epsilon)$, $d_1=d_2=32$ and different r^* values.

demonstration of the convergence rate of the algorithm, we also conduct additional simulations for convergence curves in comparison to other low-rank matrix estimation algorithms - SVD, factorized gradient descent (FGD), multiple invocations of exact singular value decompositions (SVDs) [24], and MAPLE [18] for link function $f(x,\epsilon)=x^2+\sin(x)+\epsilon$, as $f_2(x,\epsilon)$ can hardly be solved in the their paradigm due to special smoothness constraints on link functions. In addition, these baselines are gradient-based truncation approaches that pose strong limitation (*e.g.*, differentiability) in link functions.

The curves averaged over 10 independent trials of running the algorithms show the relation between error in logarithm versus the number of iterations in Figure 2(a)–2(c). We can observe that STPower converges more rapidly and achieves a lower error in this setting compared with the baselines. Among the baselines, MAPLE shows the best convergence performance but still reaches a bad solution. Moreover, we find that the performance of some gradient-based truncation methods are not stable. The reason might be partially due to the difficulty of finding an optimal optimal step-size. In return, it shows another benefit enjoyed by power methods that less hyperparameters are needed to be controlled for

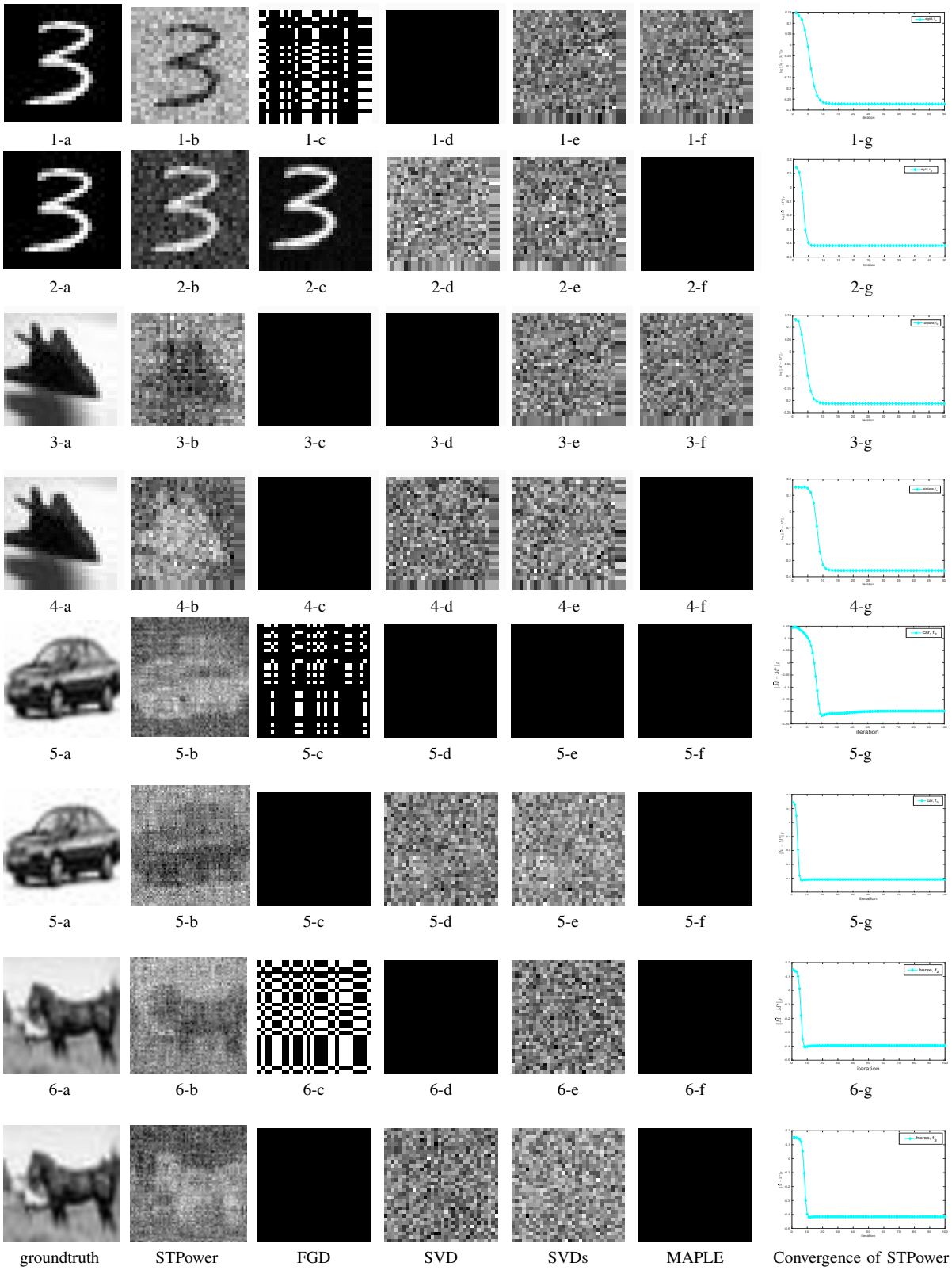
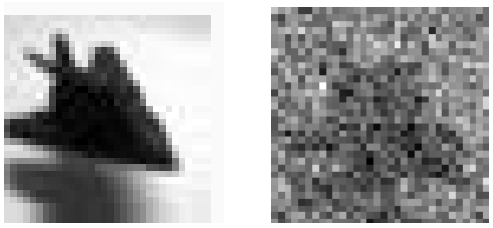


Fig. 7: Image recovery results by STPower compared with other gradient-based methods. We select four images as examples with low-rank structure from MNIST and CIFAR10 dataset, respectively. For each original image, the first row corresponds to recovery under the link function f_2, and the second row corresponds to recovery under f_3. The first column provides the groundtruth image to recover, and the column 2-6 represent the image recovery results obtained by STPower, FGD, SVD, SVDs, and MAPLE, and the last column shows the convergence behaviors of STPower in the four cases. Note that in some of the recovered images, the values of the pixels are flipped compared with the original one. This is because the phase can only be recovered up to a difference of π.

TABLE II: Comparison of reconstruction errors with other related works.

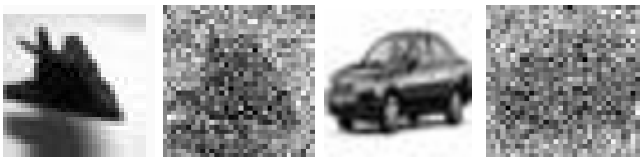
Reconstruction Error	STPower	FGD	SVD	SVDs	MAPLE
car, f_2	0.588	1.376	1.457	1.457	1.457
car, f_3	0.363	1.457	1.491	1.491	1.457
horse, f_2	0.380	1.434	1.462	1.205	1.462
horse, f_3	0.372	1.462	1.453	1.429	1.462

We measure the quantitative performance of the algorithms using the reconstruction error, i.e., $\|\bar{\mathbf{M}} - \mathbf{M}^*\|_F$, where the Frobenius norm of each matrix, i.e., $\bar{\mathbf{M}}$ and $\mathbf{M}^*$ has been normalized to 1.



groundtruth recovered image

Fig. 8: An example of image recovery from Fourier measurements by applying STPower.



groundtruth airplane recovered airplane groundtruth car recovered car

Fig. 9: Examples of the image recovery results from sparse high-dimensional images.

implementing the algorithm. Toward this end, we have verified the performance of STPower with respect to both computational efficiency and statistical accuracy.

In addition, we provide some further simulation results to support the notable convergence performance and wide applicability of our algorithm. A bunch of convergence curves with different d_1, d_2, r^* settings plotted in Figure 4(a)–4(c). We can conclude that the optimization error of STPower converges rapidly at a geometric rate until the statistical error is achieved. With the increase of matrix dimensions and the true rank, the statistical error rises, which verifies the gap caused by rank restricted norm presented on the right-hand side of (III.4) and statistical rate in Theorem III.6. Such results further indicate that our algorithm inherits both the fast convergence property of the power method and effectiveness of a proper matrix estimator.

To show the more numerical performance of STPower compared with the other state-of-the-art methods in the cases with larger dimensions and different condition numbers of the second-order link matrix $\mathbf{A}$, we perform additional experiments as below. We first added the comparison of the STPower in well-conditioned and ill-conditioned cases under Gaussian measurements, wherein the well-conditioned case we use the *i.i.d.* Gaussian entries, which results in a diagonal $\mathbf{A}^*$ with identical singular values, i.e., the variances of the normal distribution. Hence the condition number is 1. In the ill-conditioned case, we modified the variance of some entry of $\mathbf{X}$ to be 0.1 and two entries of $\mathbf{X}$ to be 1.5 and 0.15, to generate "ill-conditioned" and "ill-conditioned2" cases respectively in Figure 5. It can be observed from the figure that STPower, in either case, shows a linear convergence behavior at the beginning, and then the convergence curve becomes flat after a number of iterations. Since the modified variances are different, it can be seen that the final statistical error achieved by STPower in well-conditioned and ill-conditioned cases are dissimilar as well. Better error induced by "ill-conditioned" follows from a larger eigengap and smaller variance, while the worse performance of "ill-conditioned2" results from a larger variance that outweighs the larger eigengap.

For large scale experiments, we lay out the results of the scaling effects of statistical errors in Figure 4 with dimensions of 32×32 , 64×64 , and 128×128. We can extend to larger thousands of dimensions by blockwise computation and parallelization based on power iteration when having access to more computational resources. Furthermore, we add more real data experiments to demonstrate the effectiveness of STPower for general/unknown link functions in Figure 7. We select four images as examples with a low-rank structure from MNIST and CIFAR10 datasets, respectively. Also, we adopt Gaussian measurements with link function f_2 for the first two rows and link function f_3 for the last two rows to generate the responses of the AMPR model. Our recovery result in the second column reveals that the proposed STPower is able to recover the digits successfully and the airplane approximately while the existing works fail to recover the ground truth. The reason is that the previous works lie on the smoothness/differentiability and monotonicity of the link function. Specifically, for two considered AMPR models, link function f_2, i.e., absolute value, does not satisfy these conditions, and link function f_3 only meets these conditions over a limited domain. Therefore, other gradient-based methods either diverge in the pixel values (We clip the grayscale pixel values at 0 and 255.) or only have some chance to recover, e.g., Figure 7(2-b), when the measurements lie in the region satisfying the aforementioned conditions. The overall results illustrate that our proposed model agnostic PR solved by STPower is more general than the ones considered in [18]. To make the experiments more informative and extensive, we select two more recovery results of more complicated images from the CIFAR10 dataset, which consist of a horse and a car with more texture, respectively. The new experiments share the same settings as before. Such results further demonstrate that our algorithm works relatively well without the differentiability and monotonicity of the link function. Also, with the reconstruction errors provided in Table II, we confirm that our method provides quantitatively fair solutions to the agnostic matrix PR problem.

In high-dimensional settings, our method may require more storage. In real-world data, by using the sparsity feature

of images, we are able to directly apply our algorithm via the operations for sparse matrices. In Figure 9 we provide example results for sparse images of dimension 256×256 under the link function f_3. While extending our method to a distributed scenario by leveraging the efficient matrix-vector multiplication of the power iteration would be an interesting future direction.

Furthermore, for practical consideration, we also attempt to run our algorithm for Fourier measurements [25] with Gaussian noise, and provide an example result in Figure 8. Specifically, we adopt (II.6) to calculate the second-order link matrix but replace the response y with the response of Fourier transformation measurements with added Gaussian noise. Without theoretical guarantees, our method works slightly well in the sense that the boundary of the object could be found. Also, it can be observed that when the variance of the added Gaussian noise is large, STPower will perform better. It is worth noting that in practical PR problems, Fourier measurements would be used instead of the random measurements considered in [34], [46]. However, the methods developed from randomized PR with strong theoretical guarantees are still widely applied in practice. Similarly, we hope that our work sheds light on the practical algorithm design for PR, and serves as a theoretically sound choice even in real-world problems, especially for low-rank matrix signals where very few approaches are explored.

V. Concluding Remark

In this paper, we considered a new class of AMPR model, which is a generalization of traditional PR and SIMs to low-rank matrix scenarios under relatively mild assumptions on the second-order link function $f(\cdot, \epsilon)$ and $\mathbf{X}$. Specifically, we cast the AMPR model into a rank restricted largest eigenvalue problem with a second-order link matrix by Stein's identity. A novel algorithm STPower was proposed to solve the constrained optimization problem with a near-optimal optimal statistical rate. We justified the effectiveness of our algorithm with sufficient technical analysis and numerical results. To conclude, we list the advantages of our algorithm and model to justify the meaningfulness of our work below.

- We require no specific information of the link function in the matrix PR model, compared with existing works that inquire the gradient of the link function or assume a quadratic model.
- We pose less restrictions on the link function, compared with related works on matrix PR. For example, the most related works in Table 1.
- Our power-iteration based method has fewer hyperparameters to tune, for example, the stepsize and penalty coefficients, when compared with existing gradient based methods. Hence STPower is more implementable in practice.
- In addition, since STPower is based on power iteration method, it can be parallelized on large clusters for large-scale applications.

Additionally, in popular deep learning models composed by cascades of index models, our model can be applied to develop effective estimators for weight matrices with low-rank restriction, where the activation function corresponds to the link function $f(\cdot, \epsilon)$ in the literature. This is a prospective future work.

Appendix

In this section, we give the details of proofs used for the main theories. Our proofs utilize several technical tools including the perturbation theory of the symmetric eigenvalue problem, the convergence analysis for the vanilla power method, the error analysis of spectrum truncation operation, the cover number in the matrix space, and concentration inequalities.

A. Proof of Theorem III.2

The proof involves several supporting lemmas. In the following sections, we denote $\mathbf{A}_{\mathcal{B}}$ as the principal submatrix of $\mathbf{A}$ with rows and columns indexed by the indexes in $\mathrm{vec}(\mathbf{M})$ corresponding to the elements in columns indicated in set $\mathcal{B}$. If needed, we will also denote by $A_{\mathcal{B}}$ as the restriction of $\mathbf{A}$ on such rows and columns. First, we have the lemma giving how the error changes under the vanilla power iteration method applied to the matrix case below. Note that $\mathbf{A} = \mathbf{A}^* + \mathbf{E}$.

Lemma A.1. Let $\mathrm{vec}(\mathbf{N})$ derived from vectorizing the matrix $\mathbf{N} \in \mathbb{R}^{d_1 \times d_2}$ be the eigenvector of the largest eigenvalue of a $d_1 d_2 \times d_1 d_2$ symmetric matrix $\mathbf{A}$ in absolute value, and suppose that $\gamma < 1$ is the ratio of the second largest to largest eigenvalue by absolute values. Then for any $\mathbf{M} \in \mathbb{R}^{d_1 \times d_2}$ such that $\|\mathbf{M}\|_F = 1$ and $\langle \mathbf{M}, \mathbf{N} \rangle > 0$, letting $\mathrm{vec}(\mathbf{M}') = \mathbf{A}\,\mathrm{vec}(\mathbf{M})/\|\mathbf{A}\,\mathrm{vec}(\mathbf{M})\|$, we have

$$\|\mathbf{M}' - \mathbf{N}\|_F^2 \le [1 - \frac{1}{2}(1-\gamma^2)(1+\langle \mathbf{M}, \mathbf{N} \rangle)\langle \mathbf{M}, \mathbf{N} \rangle]\|\mathbf{M} - \mathbf{N}\|_F^2. \tag{A.1}$$

Proof: Without loss of generality, we may assume that the largest eigenvalue in absolute value $\lambda_1(\mathbf{A}) = 1$, and $|\lambda_j(\mathbf{A})| \le \gamma$ for $j > 1$. $\mathrm{vec}(\mathbf{M})$ can be decomposed into

$$\mathrm{vec}(\mathbf{M}) = p\,\mathrm{vec}(\mathbf{N}) + q\mathbf{n}', \tag{A.2}$$

where $\mathbf{n}' \in \mathbb{R}^{d_1 d_2}$ satisfying $\mathrm{vec}(\mathbf{N})^\top \mathbf{n}' = 0$, $\|\mathbf{N}\|_F = \|\mathbf{n}'\|_2 = 1$, and $p^2 + q^2 = 1$, which is followed by $p = \langle \mathbf{M}, \mathbf{N} \rangle$. Let $\boldsymbol{b}' = \mathbf{A}\mathbf{n}'$, then $\|\boldsymbol{b}'\|_2 \le \gamma$ and $\mathrm{vec}(\mathbf{N})^\top \boldsymbol{b}' = 0$. Then by left multiplying $\mathbf{A}$ to (A.2) we have $\mathbf{A}\,\mathrm{vec}(\mathbf{M}) = p\,\mathrm{vec}(\mathbf{N}) + q\boldsymbol{b}'$, it follows that

$$\begin{aligned} |\langle \mathbf{M}', \mathbf{N} \rangle| &= \frac{|\,\mathrm{vec}(\mathbf{N})^\top \mathbf{A}\,\mathrm{vec}(\mathbf{M})|}{\|\mathbf{A}\,\mathrm{vec}(\mathbf{M}\|} \qquad (A.3) \\ &= \frac{|p\lambda_1(\mathbf{A}) + q\,\mathrm{vec}(\mathbf{N})^\top \mathbf{A}\mathbf{n}'|}{\sqrt{p^2 + q^2\|\boldsymbol{b}'\|^2}} = \frac{|p + q\,\mathrm{vec}(\mathbf{N})^\top \boldsymbol{b}'|}{\sqrt{p^2 + q^2\|\boldsymbol{b}'\|^2}} \\ &= \frac{|p|}{\sqrt{p^2 + q^2\|\boldsymbol{b}'\|^2}} \ge \frac{|p|}{\sqrt{p^2 + q^2\gamma^2}} \\ &= \frac{|\langle \mathbf{M}, \mathbf{N} \rangle|}{\sqrt{1 - (1-\gamma^2)(1 - \langle \mathbf{M}, \mathbf{N} \rangle^2)}} \\ &\ge |\langle \mathbf{M}, \mathbf{N} \rangle|[1 + (1-\gamma^2)(1 - \langle \mathbf{M}, \mathbf{N} \rangle^2)/2]. \qquad (A.4) \end{aligned}$$

The last inequality is derived from $1/\sqrt{1-t} \ge 1 + t/2$ for $t \in [0, 1)$. For simplicity and without loss of generality, we postulate that $\langle \mathbf{M}', \mathbf{N} \rangle \ge 0$, otherwise the signs in the proof

can be simply properly changed. Then we are able to compute the norm

$$\begin{aligned}\|\mathbf{M}'-\mathbf{N}\|_F^2 &= 2-2|\langle \mathbf{M}',\mathbf{N}\rangle| \qquad \text{(A.5)}\\ &\le 2-2|\langle \mathbf{M},\mathbf{N}\rangle|[1+(1-\gamma^2)(1+\langle \mathbf{M},\mathbf{N}\rangle)(1-\langle \mathbf{M},\mathbf{N}\rangle)/2]\\ &= [1-(1-\gamma^2)(1+\langle \mathbf{M},\mathbf{N}\rangle)\langle \mathbf{M},\mathbf{N}\rangle/2](2-2\langle \mathbf{M},\mathbf{N}\rangle)\\ &= [1-(1-\gamma^2)(1+\langle \mathbf{M},\mathbf{N}\rangle)\langle \mathbf{M},\mathbf{N}\rangle/2]\|\mathbf{M}-\mathbf{N}\|_F^2,\end{aligned}$$

which shows the desired bound. ∎

The following standard result shows the perturbation theory of the symmetric eigenvalue problem.

Lemma A.2. If $\mathbf{P}$ and $\mathbf{P}+\mathbf{U}$ are $p\times p$ symmetric matrices, then $\forall 1\le k\le p$,

$$\lambda_k(\mathbf{P})+\lambda_p(\mathbf{U})\le\lambda_k(\mathbf{P}+\mathbf{U})\le\lambda_k(\mathbf{P})+\lambda_1(\mathbf{U}), \quad \text{(A.6)}$$

where $\lambda_k(\mathbf{P})$ stands for the k-th largest eigenvalue of $\mathbf{P}$.

The detailed proof of Lemma A.2 can be found in [47]. By leveraging this lemma, we can show the error bound between the true parameter matrix and a relaxed one below.

Lemma A.3. Suppose that the rank of $d_1\times d_2$ matrix $\mathbf{M}$ is r. Let $\mathcal{B}$ ($\mathcal{B}\neq\varnothing$) be the set of column indexes of the maximum linearly independent columns in $\mathbf{M}$ ($\mathrm{rank}(\mathbf{M})=r, |\mathcal{B}|=r$) and $\mathrm{basis}(\mathbf{M}^*)\subseteq\mathcal{B}$. If $\rho(\mathbf{E},r)\le\delta\lambda/2$, then the ratio of the second largest (in absolute value) to the largest eigenvalue of matrix $\mathbf{A}$ is no more than $\gamma(r)=\frac{\lambda_1(\mathbf{A}^*)-\delta\lambda+\rho(\mathbf{E},r)}{\lambda_1(\mathbf{A}^*)-\rho(\mathbf{E},r)}$. Furthermore,

$$\|\mathbf{M}^*-\mathbf{M}(\mathcal{B})\|_F\le\delta(r):=\frac{\sqrt{2}\rho(\mathbf{E},r)}{\sqrt{\rho(\mathbf{E},r)^2+(\delta\lambda-2\rho(\mathbf{E},r))^2}}. \quad \text{(A.7)}$$

Proof: Replacing $\mathbf{P}$, $\mathbf{U}$ with $\mathbf{A}^*_{\mathcal{B}}$ and $\mathbf{E}_{\mathcal{B}}$ respectively, we have

$$\begin{aligned}\lambda_1(\mathbf{A}_{\mathcal{B}})&\ge\lambda_1(\mathbf{A}^*_{\mathcal{B}})+\lambda_p(\mathbf{E}_{\mathcal{B}})\ge\lambda_1(\mathbf{A}^*_{\mathcal{B}})-\rho(\mathbf{E}_{\mathcal{B}})\\ &\ge\lambda_1(\mathbf{A}^*)-\rho(\mathbf{E},r), \quad \text{(A.8)}\end{aligned}$$

where the last inequality comes from the definition of $\rho(\mathbf{E},r)$ and the fact that $\mathbf{E}_{\mathcal{B}}$ is a restriction of $\mathbf{E}$ on $\mathcal{B}$. Furthermore, $\forall j\ge 2$,

$$|\lambda_j(\mathbf{A}_{\mathcal{B}})|\le|\lambda_j(\mathbf{A}^*_{\mathcal{B}})|+\rho(\mathbf{E}_{\mathcal{B}})\le\lambda_1(\mathbf{A}^*)-\delta\lambda+\rho(\mathbf{E},r),$$

which demonstrates the first statement of the lemma.

We decompose the largest eigenvector of $\mathbf{A}_{\mathcal{B}}$, i.e., $\mathrm{vec}(\mathbf{M}(\mathcal{B}))$, with $\mathbf{M}(\mathcal{B})$ of low rank no greater than r, into the following two orthogonal directions,

$$\mathrm{vec}(\mathbf{M}(\mathcal{B}))=s\,\mathrm{vec}(\mathbf{M}^*)+t\mathbf{m}', \quad \text{(A.9)}$$

where $\|\mathbf{M}^*\|_F=\|\mathbf{m}'\|=1$, $\mathrm{vec}(\mathbf{M}^*)^\top\mathbf{m}'=0$ and $s^2+t^2=1$, with eigenvalue $\lambda'\ge\lambda_1(\mathbf{A}^*)-\rho(\mathbf{E},r)$. It follows that

$$s\mathbf{A}_{\mathcal{B}}\,\mathrm{vec}(\mathbf{M}^*)+t\mathbf{A}_{\mathcal{B}}\mathbf{m}'=\lambda'(s\,\mathrm{vec}(\mathbf{M}^*)+t\mathbf{m}').$$

Multiplying by $\mathbf{m}'^\top$, we have

$$s\mathbf{m}'^\top\mathbf{A}_{\mathcal{B}}\,\mathrm{vec}(\mathbf{M}^*)+t\mathbf{m}'^\top\mathbf{A}_{\mathcal{B}}\mathbf{m}'=\lambda' t,$$

that is,

$$|t|\le|s|\frac{|\mathbf{m}'^\top\mathbf{A}_{\mathcal{B}}\,\mathrm{vec}(\mathbf{M}^*)|}{\lambda'-\mathbf{m}'^\top\mathbf{A}_{\mathcal{B}}\mathbf{m}'}=|s|\frac{|\mathbf{m}'^\top\mathbf{E}_{\mathcal{B}}\,\mathrm{vec}(\mathbf{M}^*)|}{\lambda'-\mathbf{m}'^\top\mathbf{A}_{\mathcal{B}}\mathbf{m}'}, \quad \text{(A.10)}$$

where the last equality follows from $\mathbf{m}'^\top\mathbf{A}^*_{\mathcal{B}}\,\mathrm{vec}(\mathbf{M}^*)=\lambda_1(\mathbf{A}^*)\mathbf{m}'^\top\,\mathrm{vec}(\mathbf{M}^*)=0$, which is supported by the assumption that $\mathrm{basis}(\mathbf{M}^*)\subseteq\mathcal{B}$. Furthermore, the numerator of Eq.(A.10) can be upper bounded by

$$|\mathbf{m}'^\top\mathbf{E}_{\mathcal{B}}\,\mathrm{vec}(\mathbf{M}^*)|\le\max_{\|\mathbf{y}\|=\|\mathbf{x}\|=1}|\mathbf{y}^\top\mathbf{E}_{\mathcal{B}}\mathbf{x}|=\rho(\mathbf{E}_{\mathcal{B}})\le\rho(\mathbf{E},r). \quad \text{(A.11)}$$

Also, we estimate the denominator of Eq.(A.10) by

$$\begin{aligned}&\lambda'-\mathbf{m}'^\top\mathbf{A}_{\mathcal{B}}\mathbf{m}'\\ &\ge\lambda_1(\mathbf{A}^*)-\rho(\mathbf{E},r)-\mathbf{m}'^\top\mathbf{A}^*_{\mathcal{B}}\mathbf{m}'-\mathbf{m}'^\top\mathbf{E}_{\mathcal{B}}\mathbf{m}'\\ &\ge\lambda_1(\mathbf{A}^*)-\rho(\mathbf{E},r)-|\lambda_j(\mathbf{A}^*)|_{j\ge2}-\rho(\mathbf{E},r)\\ &\ge\delta\lambda-2\rho(\mathbf{E},r). \quad \text{(A.12)}\end{aligned}$$

Putting (A.10) - (A.12) together, we obtain $|t|\le L|s|$ where $L=\rho(\mathbf{E},r)/(\delta\lambda-2\rho(\mathbf{E},r))$. It implies that $1=s^2+t^2\le s^2(1+L^2)$, therefore $s^2\ge 1/(1+L^2)$. By multiplying $\mathrm{vec}(\mathbf{M}^*)$ to (A.9) we can obtain $s=\langle\mathbf{M}^*,\mathbf{M}(\mathcal{B})\rangle$. We assume that $s>0$ without loss of generality, as otherwise we can replace $\mathrm{vec}(\mathbf{M})^*$ with $-\mathrm{vec}(\mathbf{M})^*$. Then we apply the above results to the norm

$$\begin{aligned}&\|\mathbf{M}^*-\mathbf{M}(\mathcal{B})\|_F^2=2-2\langle\mathbf{M}^*,\mathbf{M}(\mathcal{B})\rangle\\ &=2-2s\le2\frac{\sqrt{1+L^2}-1}{\sqrt{1+L^2}}\le\frac{2L^2}{1+L^2}=\delta(r). \quad \text{(A.13)}\end{aligned}$$

Thus, the expected bound is proved. ∎

The two following lemmas describe the error introduced by the matrix spectrum truncation, which admits a tighter bound compared to the sparse vector truncation case.

Lemma A.4. Let $\mathbf{M}^*$ denote the optimum matrix subject to (I.3) with $\mathrm{rank}(\mathbf{M}^*)\le r^*$, and $\mathcal{T}_r(\cdot):\mathbb{R}^{d_1\times d_2}\to\mathbb{R}^{d_1\times d_2}$ is the singular values truncation operator, which remains the largest r singular values and truncates the other ones to zero. Let $r_0\ge r^*$, then for any $\mathbf{M}\in\mathbb{R}^{d_1\times d_2}$ we have

$$\|\mathcal{T}_r(\mathbf{M})-\mathbf{M}^*\|_F^2\le\Big(1+\frac{2\sqrt{r^*}}{\sqrt{r_0-r^*}}\Big)\cdot\|\mathbf{M}-\mathbf{M}^*\|_F^2. \quad \text{(A.14)}$$

Proof: Suppose that the singular value decomposition of $\mathbf{M}$ and $\mathbf{M}^*$ are in the form of $\mathbf{M}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ and $\mathbf{M}^*=\mathbf{U}^*\mathbf{\Sigma}^*(\mathbf{V}^*)^T$ respectively, where $\mathbf{\Sigma}$ and $\mathbf{\Sigma}^*$ are near diagonal matrices in the following form:

$$\begin{bmatrix}\sigma_1 & & \\ & \sigma_2 & \\ & & \ddots\end{bmatrix}, \quad \text{(A.15)}$$

where $\sigma_1\ge\sigma_2\ge\sigma_3\ge\cdots\ge0$ are the rearranged singular values of $\mathbf{M}$ or $\mathbf{M}^*$. Following the technique in [51], as $\mathbf{U}(\mathbf{U}^*)$ and $\mathbf{V}(\mathbf{V}^*)$ are unitary in singular value decomposition scheme, we have

$$\begin{aligned}&\|\mathcal{T}_r(\mathbf{M})-\mathbf{M}^*\|_F^2-\|\mathbf{M}-\mathbf{M}^*\|_F^2 \quad \text{(A.16)}\\ &=\|\mathcal{T}_r(\mathbf{M})\|_F^2-\|\mathbf{M}\|_F^2+2\langle\mathbf{M}-\mathcal{T}_r(\mathbf{M}),\mathbf{M}^*\rangle\\ &=\|\mathcal{T}_r(\mathbf{\Sigma})\|_F^2-\|\mathbf{\Sigma}\|_F^2+2\langle\mathbf{M}-\mathcal{T}_r(\mathbf{M}),\mathbf{M}^*\rangle. \quad \text{(A.17)}\end{aligned}$$

Plugging in Von Neumann's trace inequality [52] $\langle \mathbf{A}, \mathbf{B}\rangle \leq \Sigma_{i=1}^{\min\{\mathrm{rank}(\mathbf{A}),\mathrm{rank}(\mathbf{B})\}}\sigma_i(\mathbf{A})\cdot\sigma_i(\mathbf{B})$ for matrices $\mathbf{A},\mathbf{B}\in\mathbb{R}^{d_1\times d_2}$, we can reduce (A.16) as follows.

$$\begin{aligned}
&\|\mathcal{T}_r(\mathbf{M})-\mathbf{M}^*\|_F^2-\|\mathbf{M}-\mathbf{M}^*\|_F^2\\
=&\|\mathcal{T}_r(\mathbf{\Sigma})\|_F^2-\|\mathbf{\Sigma}\|_F^2+2\langle\mathbf{M}-\mathcal{T}_r(\mathbf{M}),\mathbf{M}^*\rangle \qquad (A.18)\\
\leq&\|\mathcal{T}_r(\mathbf{\Sigma})\|_F^2-\|\mathbf{\Sigma}\|_F^2+2\sum_{i=1}^{r^*}\sigma_i(\mathbf{M}-\mathcal{T}_r(\mathbf{M}))\cdot\sigma_i(\mathbf{M}^*)\\
=&|\mathcal{T}_r(\mathbf{\Sigma})\|_F^2+2\sum_{i=1}^{r^*}(\sigma_{i+r}(\mathbf{M})-\sigma_{i+r}(\mathcal{T}_r(\mathbf{M})))\cdot\sigma_i(\mathbf{M}^*)\\
=&\|\mathcal{T}_r(\mathbf{\Sigma})-\mathbf{\Sigma}^*\|_F^2-\|\mathbf{\Sigma}-\mathbf{\Sigma}^*\|_F^2. \qquad (A.19)
\end{aligned}$$

Then Lemma 3.3 in [51] implies the result. ∎

Next, the main lemma demonstrates to what extent capacity of our overall STPower iteration method can decrease the estimation error at each iteration. Note that in the following we measure the error of an iterate $\mathbf{M}$ using $\min\{\|\mathbf{M}-\mathbf{M}^*\|_F,\|\mathbf{M}+\mathbf{M}^*\|_F\}$. For simplicity, we still write $\|\mathbf{M}-\mathbf{M}^*\|_F$ for the error.

Lemma A.5. Suppose that $r_0\geq r^*$. Let $r=2r_0+r^*$. If $|\langle\mathbf{M}^{(t)},\mathbf{M}^*\rangle|>\delta(r)+\omega$ for $t\in\mathbb{N}$ and some $\omega\in(0,1)$, then there exists $0<\beta<1$ satisfying

$$\|\widehat{\mathbf{M}}^{(t+1)}-\mathbf{M}^*\|_F\leq\beta\|\mathbf{M}^{(t)}-\mathbf{M}^*\|_F+2\delta(r). \qquad (A.20)$$

Proof: We denote the column index set of basis vectors of $\mathbf{M}^{(t)}$ as $\mathcal{B}_t$. Let $\mathcal{B}=\mathcal{B}_t\cup\mathcal{B}_{t+1}\cup\mathrm{basis}(\mathbf{M}^*)$. We redefine

$$\mathrm{vec}(\mathbf{M}^{(t+0.5)})=\mathrm{vec}(\widetilde{\mathbf{M}}^{(t+0.5)})=\frac{\mathbf{A}_{\mathcal{B}}\,\mathrm{vec}(\mathbf{M}^{(t)})}{\|\mathbf{A}_{\mathcal{B}}\,\mathrm{vec}(\mathbf{M}^{(t)})\|}. \qquad (A.21)$$

Since replacing $\mathrm{vec}(\mathbf{M}^{(t+0.5)})$ with $\mathrm{vec}(\widetilde{\mathbf{M}}^{(t+0.5)})$ has no impact on the result iteration sequence $\{\mathbf{M}^{(t)}\}$ (the remaining part of $\mathrm{vec}(\mathbf{M}^{(t+0.5)})$ is determined by the transformation of column basis), the redefinition is applicable to the following proof for simplicity.

Without loss of generality, we assume that $\langle\mathbf{M}^{(t+0.5)},\mathbf{M}(\mathcal{B})\rangle\geq 0$ and $\langle\mathbf{M}^{(t)},\mathbf{M}^*\rangle\geq 0$ as the sign can be appropriately changed. From Lemma A.1, we have

$$\begin{aligned}
&\|\mathbf{M}^{(t+0.5)}-\mathbf{M}(\mathcal{B})\|_F^2\\
\leq&[1-\frac{1-\gamma(r)^2}{2}(1+m_{\mathcal{B}}^{(t)})m_{\mathcal{B}}^{(t)}]\|\mathbf{M}^{(t)}-\mathbf{M}(\mathcal{B})\|_F^2\\
\leq&[1-0.5(1-\gamma(r)^2)\omega(1+\omega)]\|\mathbf{M}^{(t)}-\mathbf{M}(\mathcal{B})\|_F^2 \qquad (A.22)
\end{aligned}$$

where $m_{\mathcal{B}}^{(t)}=\langle\mathbf{M}^{(t)},\mathbf{M}(\mathcal{B})\rangle$. The second inequality comes from Lemma A.1 and the assumption made before, i.e., $\langle\mathbf{M}^{(t)},\mathbf{M}(\mathcal{B})\rangle\geq\langle\mathbf{M}^{(t)},\mathbf{M}^*\rangle-\delta(r)\geq\omega$. Then, we have

$$\begin{aligned}
&\|\mathbf{M}^{(t+0.5)}-\mathbf{M}^*\|_F \qquad (A.23)\\
\leq&\|\mathbf{M}^{(t+0.5)}-\mathbf{M}(\mathcal{B})\|_F+\|\mathbf{M}(\mathcal{B})-\mathbf{M}^*\|_F\\
\leq&\sqrt{1-0.5(1-\gamma(r)^2)\omega(1+\omega)}\|\mathbf{M}^{(t)}-\mathbf{M}(\mathcal{B})\|_F+\delta(r)\\
\leq&\sqrt{\frac{2-(1-\gamma(r)^2)\omega(1+\omega)}{2}}\big(\|\mathbf{M}^{(t)}-\mathbf{M}^*\|_F+\|\mathbf{M}^*-\mathbf{M}(\mathcal{B})\|_F\big)\\
&+\delta(r)\\
\leq&\sqrt{\frac{2-(1-\gamma(r)^2)\omega(1+\omega)}{2}}\|\mathbf{M}^{(t)}-\mathbf{M}^*\|_F+2\delta(r).
\end{aligned}$$

According to Lemma A.4 and assumption $r_0\geq r^*$, we can finally obtain

$$\begin{aligned}
&\|\widehat{\mathbf{M}}^{(t+1)}-\mathbf{M}^*\|_F \qquad (A.25)\\
\leq&\sqrt{1+\frac{2\sqrt{r^*}}{\sqrt{r_0-r^*}}}\cdot\sqrt{\frac{2-(1-\gamma(r)^2)\omega(1+\omega)}{2}}\\
&\cdot\|\mathbf{M}^{(t+0.5)}-\mathbf{M}^*\|_F\\
\leq&\sqrt{\left(1+\frac{2\sqrt{r^*}}{\sqrt{r_0-r^*}}\right)\cdot\left(1-\frac{(1-\gamma(r)^2)\omega(1+\omega)}{2}\right)}\\
&\cdot\|\mathbf{M}^{(t)}-\overline{\mathbf{M}}\|_F+2\delta(r) \qquad (A.26)\\
=&\beta\|\mathbf{M}^{(t)}-\mathbf{M}^*\|_F+2\delta(r), \qquad (A.27)
\end{aligned}$$

where

$$\beta=\sqrt{\left(1+\frac{2\sqrt{r^*}}{\sqrt{r_0-r^*}}\right)\cdot\left(1-\frac{(1-\gamma(r)^2)\omega(1+\omega)}{2}\right)}$$

with $\gamma(r)<1$ as the ratio of the second largest to largest eigenvalue of the matrix and appropriate $\omega\in(0,1)$. Hence, we can get the desired contraction result. ∎

After obtaining the contraction relation between the successive iterates with an additional statistical error, we can immediately reach our first main result of long-term relation for the rank recovery error as below.

Proof: According to the quantitative relation between $\delta(r)$, and β, we divide the proof in following two cases:

(1) $1-\frac{2\delta^2(r)}{(1-\beta)^2}<\delta(r)$: It follows naturally that

$$\begin{aligned}
&\|\mathbf{M}^{(0)}-\mathbf{M}^*\|_F=\sqrt{2-2|\langle\mathbf{M}^{(0)},\mathbf{M}^*\rangle|} \qquad (A.28)\\
&\leq\sqrt{2-2\delta(r)}<\sqrt{2-2\left(1-\frac{2\delta^2(r)}{(1-\beta)^2}\right)}=\frac{2\delta(r)}{1-\beta}.
\end{aligned}$$

(2) $1-\frac{2\delta^2(r)}{(1-\beta)^2}\geq\delta(r)$: Under this constraint, first we prove that for all $t\geq 0$, $\mathbf{M}^{(t)}$ satisfies the condition $|\langle\mathbf{M}^{(t)},\mathbf{M}^*\rangle|\geq\delta(r)$ for Lemma A.5 to hold, then we will utilize Lemma A.5 iteratively to obtain the result. For the former purpose, the result for $t=0$ has been checked by the initial condition in Theorem III.2. Now assume that for $t\geq 1$, $|\langle\mathbf{M}^{(t-1)},\mathbf{M}^*\rangle|\geq\delta(r)$. The analysis will condition on two further cases:

(a) $\|\mathbf{M}^{(t-1)}-\mathbf{M}^*\|_F\geq\frac{2\delta(r)}{1-\beta}$: In this case, from definition $\|\widehat{\mathbf{M}}^{(t)}\|_F\leq 1$ and $|\langle\mathbf{M}^{(t)},\mathbf{M}^*\rangle|=\frac{|\langle\widehat{\mathbf{M}}^{(t)},\mathbf{M}^*\rangle|}{\|\widehat{\mathbf{M}}^{(t)}\|_F}\geq|\langle\widehat{\mathbf{M}}^{(t)},\mathbf{M}^*\rangle|$, combining Lemma A.5 we have

$$\begin{aligned}
&\sqrt{2-2|\langle\mathbf{M}^{(t)},\mathbf{M}^*\rangle|}\leq\sqrt{2-2|\langle\widehat{\mathbf{M}}^{(t)},\mathbf{M}^*\rangle|} \qquad (A.29)\\
&=\|\widehat{\mathbf{M}}^{(t)}-\mathbf{M}^*\|_F\leq\beta\|\mathbf{M}^{(t-1)}-\mathbf{M}^*\|_F+2\delta(r)\\
&\leq\|\mathbf{M}^{(t-1)}-\mathbf{M}^*\|_F=\sqrt{2-2|\langle\mathbf{M}^{(t-1)},\mathbf{M}^*\rangle|},
\end{aligned}$$

which suffices to show $|\langle\mathbf{M}^{(t)},\mathbf{M}^*\rangle|\geq|\langle\mathbf{M}^{(t-1)},\mathbf{M}^*\rangle|\geq\delta(r)$.

(b) $\|\mathbf{M}^{(t-1)} - \mathbf{M}^*\|_F < \frac{2\delta(r)}{1-\beta}$: Similarly in (a), we can show that

$$\sqrt{2 - 2|\langle \mathbf{M}^{(t)}, \mathbf{M}^* \rangle|} \leq \beta \|\mathbf{M}^{(t-1)} - \mathbf{M}^*\|_F + 2\delta(r) < \frac{2\delta(r)}{1-\beta}. \tag{A.30}$$

It follows that $|\langle \mathbf{M}^{(t)}, \mathbf{M}^* \rangle| > 1 - \frac{2\delta^2(r)}{(1-\beta)^2} \geq \delta(r)$.

From all the above, $|\langle \mathbf{M}^{(t)}, \mathbf{M}^* \rangle| \geq \delta(r)$ is proved under the assumption $|\langle \mathbf{M}^{(t-1)}, \mathbf{M}^* \rangle| \geq \delta(r)$, which by induction demonstrates $|\langle \mathbf{M}^{(t)}, \mathbf{M}^* \rangle| \geq \delta(r)$ for all $t \geq 0$. Next, by repeatedly applying Lemma A.5 to the first item in (A.29), we have

$$\begin{aligned}
&\|\mathbf{M}^{(t)} - \mathbf{M}^*\|_F \qquad (A.31)\\
\leq &\|\widehat{\mathbf{M}}^{(t)} - \mathbf{M}^*\|_F \\
\leq &\beta \|\mathbf{M}^{(t-1)} - \mathbf{M}^*\|_F + 2\delta(r) \\
\leq &\beta \left(\beta \|\mathbf{M}^{(t-2)} - \mathbf{M}^*\|_F + 2\delta(r) \right) + 2\delta(r) \\
\leq &\beta^t \|\mathbf{M}^{(0)} - \mathbf{M}^*\|_F + \sum_{i=0}^{t-1} \beta^i 2\delta(r) \\
\leq &\beta^t \|\mathbf{M}^{(0)} - \mathbf{M}^*\|_F + \sum_{i=0}^{\infty} \beta^i 2\delta(r) \\
\leq &\beta^t \|\mathbf{M}^{(0)} - \mathbf{M}^*\|_F + \frac{2\delta(r)}{1-\beta}, \quad \forall t \geq 0 \qquad (A.32)
\end{aligned}$$

Together with (A.28) we have obtain our final result. ∎

Next, we will proceed to prove the concentration result of matrix $\mathbf{A}$ below. From [7], we are able to formulate $\mathbf{A}$ and $\mathbf{A}^*$ as $\mathbf{A} = \frac{1}{n}\sum_{i=1}^n [y_i(\text{vec}(\mathbf{X}_i)\,\text{vec}(\mathbf{X}_i)_i^\top) - \mathbf{I}]$ and $\mathbf{A}^* = \mathbb{E}[y(\text{vec}(\mathbf{X})\,\text{vec}(\mathbf{X})^\top - \mathbf{I})]$ respectively. There will be some sets of interest of matrices, denoted by $\mathcal{S}_{\mathbf{M}^{d_1\times d_2}} = \{\mathbf{M} \in \mathbb{R}^{d_1\times d_2} \mid \|\mathbf{M}\|_F = 1\}$ and $\mathcal{B}_{\mathbf{M}(r)} = \{\mathbf{M} \in \mathbb{R}^{d_1\times d_2} \mid \text{rank}(\mathbf{M}) \leq r\}$. First, a new cover number of the space of interest of matrices is proved.

B. Proof of Lemma III.5

Proof: First, we introduce $\mathcal{N}_\epsilon^*$ as the *maximal ϵ-separated* set of $\mathcal{S}_{\mathbf{M}}^{d_1\times d_2} \cap \mathcal{B}_{\mathbf{M}}(r)$, in which any two elements $\mathbf{M}_x, \mathbf{M}_y$ satisfying $\|\mathbf{M}_x - \mathbf{M}_y\|_F \geq \epsilon$, in other words, they are always at least ϵ distance away. The maximal property implies that there exists no ϵ-separated subset $\mathcal{N}_\epsilon'$ of $\mathcal{S}_{\mathbf{M}}^{d_1\times d_2} \cap \mathcal{B}_{\mathbf{M}}(r)$ subject to $\mathcal{N}_\epsilon^* \subsetneq \mathcal{N}_\epsilon'$. Indeed, the definition of maximal ϵ-separated set guarantees that $\mathcal{N}_\epsilon^*$ is an ϵ-net of $\mathcal{S}_{\mathbf{M}}^{d_1\times d_2} \cap \mathcal{B}_{\mathbf{M}}(r)$. Otherwise there would exist $\mathbf{M}_0 \in \mathcal{S}_{\mathbf{M}}^{d_1\times d_2} \cap \mathcal{B}_{\mathbf{M}}(r)$ such that no point in $\mathcal{N}_\epsilon^*$ is within ϵ-far from $\mathbf{M}_0$. Therefore, $\mathcal{N}_\epsilon^* \cup \{\mathbf{M}_0\}$ would become the larger ϵ-separated subset containing $\mathcal{N}_\epsilon^*$, which contradicts the maximality of $\mathcal{N}_\epsilon^*$. Thus we need to bound $|\mathcal{N}_\epsilon^*| = N(\epsilon, r, d_1, d_2)$.

Now we proceed to cover the neighborhood of each element $\mathbf{M}_i \in \mathcal{N}_\epsilon^*$ within balls $\mathcal{B}_{\mathbf{M}_i'} = \{\mathbf{M}_i \| \|\mathbf{M}_i' - \mathbf{M}_i\|_F \leq \epsilon/2\}$. Due to the ϵ-separated property of $\mathcal{N}_\epsilon^*$, for any two elements $\mathbf{M}_1, \mathbf{M}_2 \in \mathcal{N}_\epsilon^*$ and any $\mathbf{M}' \in \mathcal{B}_{\mathbf{M}_1'}$, we have $\|\mathbf{M}' - \mathbf{M}_1\|_F \leq \epsilon$ and the triangle inequality $\|\mathbf{M}' - \mathbf{M}_2\|_F \geq \|\mathbf{M}_1 - \mathbf{M}_2\|_F - \|\mathbf{M}' - \mathbf{M}_1\|_F \geq \epsilon - \epsilon/2 \geq \epsilon/2$. That is to say, any two differently centered balls are disjoint. On the other hand, we need to check the volume of the area containing all the balls. For $\mathbf{M} = (m)_{ij} = (\mathbf{m}_1, \mathbf{m}_2, ..., \mathbf{m}_{d_2})$, where $\mathbf{m}_i (1 \leq i \leq d_2)$ is a column vector of $\mathbf{M}$, without loss of generality, we assume that the last $(d_2 - r)$ columns are linear combinations of the first r columns, and $\mathbf{m}_k = \sum_{l=1}^r c_{lk}\mathbf{m}_l$. Then for any matrix $\mathbf{M} \in \mathcal{S}_{\mathbf{M}}^{d_1\times d_2} \cap \mathcal{B}_{\mathbf{M}}(r)$ with elements $(m)_{ij}$ as coordinates we have

$$\sum_{i=1}^{d_1}\sum_{j=1}^{r} m_{ij}^2 + \sum_{j=r+1}^{d_2}\sum_{i=1}^{d_1}\left(\sum_{l=1}^{r} c_{lj}m_{il}\right)^2 = 1 \tag{A.33}$$

We can see actually there are rd_1 independent coordinates $(m_{11}, m_{12}, ..., m_{d_1 r})$ in quadratic form (A.33). For the symmetry among $\{m_{1k}, m_{2k}, ..., m_{d_1 k}\}_{k=1}^r$, when reformulating (A.33) into the standard form, we obtain common coefficients for variables $\{m_{1k}', m_{2k}', ..., m_{d_1 k}'\}$ with respect to a given k:

$$\sum_{i=1}^{r} \frac{\sum_{j=1}^{d_1} {m_{ji}'}^2}{a_i^2} = 1, \tag{A.34}$$

which corresponds to a super ellipsoid in $\mathbb{R}^{rd_1}$ where $\{\{a_1\}_{i=1}^{d_1}, \{a_2\}_{i=1}^{d_1}, ..., \{a_r\}_{i=1}^{d_1}\}$ are the lengths of the principal semi-axes. Similarly, we can derive that the ball $\mathcal{B}_{\mathbf{M}_i'}$ reduces to an super ellipsoid $\mathbf{E}_{\mathbf{M}_i'}$ in $\mathbb{R}^{rd_1}$ with principal semi-axis lengths $\{\{\frac{\epsilon}{2}a_1\}_{i=1}^{d_1}, \{\frac{\epsilon}{2}a_2\}_{i=1}^{d_1}, ..., \{\frac{\epsilon}{2}a_r\}_{i=1}^{d_1}\}$. Then it is shown that all the balls we defined before are contained in the super ellipsoid $\bar{\mathbf{E}}$ with $\{\{(\frac{\epsilon}{2}+1)a_1\}_{i=1}^{d_1}, \{(\frac{\epsilon}{2}+1)a_2\}_{i=1}^{d_1}, ..., \{(\frac{\epsilon}{2}+1)a_r\}_{i=1}^{d_1}\}$ principal semi-axis lengths. Let $\mathcal{B}_0$ be the Euclidean ball of radii 1, since

$$\begin{aligned}
|\bar{\mathbf{E}}| \geq |\bigcup_{i=1}^{N(\epsilon,r,d_1,d_2)} \mathbf{E}_{\mathbf{M}_i'}| &= \sum_{i=1}^{N(\epsilon,r,d_1,d_2)} |\mathbf{E}_{\mathbf{M}_i'}| \\
&= N(\epsilon, r, d_1, d_2) \cdot |\mathbf{E}_{\mathbf{M}_i'}|, \qquad (A.35)
\end{aligned}$$

we have

$$N(\epsilon, r, d_1, d_2) \leq \frac{|\bar{\mathbf{E}}|}{|\mathbf{E}_{\mathbf{M}_i'}|} = \frac{\prod_{i=1}^r \left[(\frac{\epsilon}{2}+1)a_i\right]^{d_1}}{\prod_{i=1}^r \left(\frac{\epsilon}{2}a_i\right)^{d_1}} = \left(\frac{2}{\epsilon}+1\right)^{rd_1} \tag{A.36}$$

Similarly, by selecting r basis rows in $\mathbf{M}$ it is straightforward that

$$N(\epsilon, r, d_1, d_2) \leq \left(\frac{2}{\epsilon}+1\right)^{rd_2}. \tag{A.37}$$

As it has been postulated that $d_1 \leq d_2$ without any loss of generality, we have proved (III.7). We continue with (A.37) by taking $\mathbf{M}_1 \in \mathcal{N}(\epsilon, r, d_1, d_2)$, $\mathbf{M}_2 \in \mathcal{S}_{\mathbf{M}}^{d_1\times d_2} \cap \mathcal{B}_{\mathbf{M}}(r)$, and $\|\mathbf{M}_1 - \mathbf{M}_2\|_F \leq \epsilon$

$$\begin{aligned}
&|\,\text{vec}(\mathbf{M}_1)^\top \mathbf{S}\,\text{vec}(\mathbf{M}_1) - \text{vec}(\mathbf{M}_2)^\top \mathbf{S}\,\text{vec}(\mathbf{M}_2)| \\
=&|\,\text{vec}(\mathbf{M}_1)^\top \mathbf{S}[\text{vec}(\mathbf{M}_1) - \text{vec}(\mathbf{M}_2)] \\
&+ \text{vec}(\mathbf{M}_2)^\top \mathbf{S}[\text{vec}(\mathbf{M}_1) - \text{vec}(\mathbf{M}_2)]| \\
\leq &\|\mathbf{S}\|\|\mathbf{M}_1\|_F\|\mathbf{M}_1 - \mathbf{M}_2\|_F + \|\mathbf{S}\|\|\mathbf{M}_2\|_F\|\mathbf{M}_1 - \mathbf{M}_2\|_F \\
\leq &2\epsilon\|\mathbf{S}\|. \qquad (A.38)
\end{aligned}$$

It follows that

$$|\operatorname{vec}(\mathbf{M}_1)^\top \mathbf{S} \operatorname{vec}(\mathbf{M}_1)| \geq |\operatorname{vec}(\mathbf{M}_2)^\top \mathbf{S} \operatorname{vec}(\mathbf{M}_2)| - 2\epsilon\|\mathbf{S}\|. \tag{A.39}$$

Taking the maximum on both sides of (A.39) and noting that

$$\|\mathbf{S}\| \geq \max_{\mathbf{M}_2 \in \mathcal{S}_{\mathbf{M}}^{d_1\times d_2} \cap \mathcal{B}_{\mathbf{M}}(r)} |\operatorname{vec}(\mathbf{M}_2)^\top \mathbf{S} \operatorname{vec}(\mathbf{M}_2)|,$$

we can reach the result

$$\begin{aligned}&\max_{\mathbf{M}_1 \in \mathcal{N}(\epsilon,r,d_1,d_2)} |\operatorname{vec}(\mathbf{M}_1)^\top \mathbf{S} \operatorname{vec}(\mathbf{M}_1)| \\ &\geq (1-2\epsilon) \max_{\mathbf{M}_2 \in \mathcal{S}_{\mathbf{M}}^{d_1\times d_2} \cap \mathcal{B}_{\mathbf{M}}(r)} |\operatorname{vec}(\mathbf{M}_2)^\top \mathbf{S} \operatorname{vec}(\mathbf{M}_2)|.\end{aligned}$$

■

C. Proof of Theorem III.6

Lemma A.6. For any fixed $\mathbf{M} \in \mathcal{S}_{\mathbf{M}}^{d_1\times d_2} \cap \mathcal{B}_{\mathbf{M}}(r)$, under Assumption III.4 we have

$$\mathbb{P}\left\{|\operatorname{vec}(\mathbf{M})^\top \mathbf{A} \operatorname{vec}(\mathbf{M}) - \operatorname{vec}(\mathbf{M})^\top \mathbf{A}^* \operatorname{vec}(\mathbf{M})| \geq \epsilon\right\} \leq e^{-\frac{n\epsilon^2}{10K}},$$

where K is the constant stated in Assumption III.4.

Proof: Let $\{y_i, \mathbf{X}_i\}_{i=1}^n$ be n independent samples drawn from model $y = f(\langle \mathbf{M}^*, \mathbf{X}\rangle)$. According to Assumption III.4, we have bounded $yT(\mathbf{X})$. On the other hand, we have

$$\operatorname{vec}(\mathbf{M})^\top \mathbf{A} \operatorname{vec}(\mathbf{M}) = \operatorname{vec}(\mathbf{M})^\top \frac{\sum_{i=1}^n [y_i T_i(\mathbf{X})]}{n} \operatorname{vec}(\mathbf{M})$$

and

$$\operatorname{vec}(\mathbf{M})^\top \mathbf{A}^* \operatorname{vec}(\mathbf{M}) = \operatorname{vec}(\mathbf{M})^\top \mathbf{E}[yT(\mathbf{X})] \operatorname{vec}(\mathbf{M}) \tag{A.40}$$

Combining Assumption III.4 and appendix of [44], according to the Bernstein-type inequality for independent random variables [50], for any $\epsilon > 0$ we can obtain

$$\begin{aligned}&\mathbb{P}\{|\operatorname{vec}(\mathbf{M})^\top (\mathbf{A} - \mathbf{A}^*) \operatorname{vec}(\mathbf{M})| \geq \epsilon\} \qquad (A.41)\\ =&\mathbb{P}\left\{\left|\frac{\sum_{i=1}^n [y_i T_i(\mathbf{X})]}{n} - \mathbf{E}(yT(\mathbf{X}))\right| \geq \epsilon\right\} \leq \exp\left(-\frac{n\epsilon^2}{10K}\right).\end{aligned}$$

Hence, we complete the proof. ■

Using the two lemmas introduced above, we are ready to complete the proof of the third main result below.

Proof of Theorem 3.5: As the rank restricted norm $\rho(\mathbf{A} - \mathbf{A}^*, r) = \sup_{\mathbf{M}\in\mathcal{S}_{\mathbf{M}}^{d_1\times d_2}\cap\mathcal{B}_{\mathbf{M}}(r)} |\operatorname{vec}(\mathbf{M})^\top (\mathbf{A} - \mathbf{A}^*) \operatorname{vec}(\mathbf{M})|$ is derived from the union bound of $|\operatorname{vec}(\mathbf{M})^\top (\mathbf{A} - \mathbf{A}^*) \operatorname{vec}(\mathbf{M})|$, we first fix the subset $\mathcal{P} \subset \{1, ..., d_2\}$, we define

$$\mathcal{B}_{\mathcal{P}}^{d_2} = \{\mathbf{M} \in \mathcal{S}_{\mathbf{M}}^{d_1\times d_2} | \text{the basis columns are indexed by } \mathcal{P}\}.$$

Similarly we can also define

$$\mathcal{B}_{\mathcal{P}}^{d_1} = \{\mathbf{M} \in \mathcal{S}_{\mathbf{M}}^{d_1\times d_2} | \text{the basis rows are indexed by } \mathcal{P}\}.$$

For any $\epsilon > 0$ and d_2 (choosing column vectors as basis), we define events:

$$\Omega_{\mathcal{P}}^{d_2} := \left\{\sup_{\mathbf{M}\in\mathcal{S}_{\mathbf{M}}^{d_1\times d_2}\cap\mathcal{B}_{\mathcal{P}}^{d_2}} |\operatorname{vec}(\mathbf{M})^\top (\mathbf{A} - \mathbf{A}^*) \operatorname{vec}(\mathbf{M})| \geq 2\epsilon\right\},$$

$$\Omega_{\mathbf{M}} := \left\{|\operatorname{vec}(\mathbf{M})^\top (\mathbf{A} - \mathbf{A}^*) \operatorname{vec}(\mathbf{M})| \geq \epsilon\right\}.$$

Let $\mathcal{N}_{\mathcal{P}}^{d_2}$ be the $\frac{1}{4}$-net of $\mathcal{S}_{\mathbf{M}}^{d_1\times d_2} \cap \mathcal{B}_{\mathcal{P}}^{d_2}$, from Lemma III.5, we can see that in case of $\Omega_{\mathcal{P}}^{d_2}$, $\sup_{\mathbf{M}\in\mathcal{N}_{\mathcal{P}}^{d_2}} |\operatorname{vec}(\mathbf{M})^\top (\mathbf{A} - \mathbf{A}^*) \operatorname{vec}(\mathbf{M})| \geq \frac{1}{2} \sup_{\mathbf{M}\in\mathcal{S}_{\mathbf{M}}^{d_1\times d_2}\cap\mathcal{B}_{\mathcal{P}}^{d_2}} |\operatorname{vec}(\mathbf{M})^\top (\mathbf{A} - \mathbf{A}^*) \operatorname{vec}(\mathbf{M})| \geq \frac{1}{2} \cdot 2\epsilon = \epsilon$, also considering the fact that $\mathcal{N}_{\mathcal{P}}^{d_2}$ is compact and the supremum can be reached by elements in $\mathcal{N}_{\mathcal{P}}^{d_2}$, hence we can obtain

$$\begin{aligned}\Omega_{\mathcal{P}}^{d_2} &\subset \left\{\sup_{\mathbf{M}\in\mathcal{N}_{\mathcal{P}}^{d_2}} |\operatorname{vec}(\mathbf{M})^\top (\mathbf{A} - \mathbf{A}^*) \operatorname{vec}(\mathbf{M})| \geq \epsilon\right\} \\ &\subset \left\{\exists \mathbf{M}_0 \in \mathcal{N}_{\mathcal{P}}^{d_2}, \text{s.t. } |\operatorname{vec}(\mathbf{M}_0)^\top (\mathbf{A} - \mathbf{A}^*) \operatorname{vec}(\mathbf{M}_0)| \geq \epsilon\right\} \\ &\subset \bigcup_{\mathbf{M}\in\mathcal{N}_{\mathcal{P}}^{d_2}} \left\{|\operatorname{vec}(\mathbf{M})^\top (\mathbf{A} - \mathbf{A}^*) \operatorname{vec}(\mathbf{M})| \geq \epsilon\right\}.\end{aligned} \tag{A.42}$$

Letting $|\mathcal{N}_{\mathcal{P}}^{d_2}| = r$ and using Lemma III.5 and (A.6), we have

$$\mathbb{P}\{\Omega_{\mathcal{P}}^{d_2}\} \leq \sum_{\mathbf{M}\in\mathcal{N}_{\mathcal{P}}^{d_2}} \mathbb{P}\{\Omega_{\mathbf{M}}\} \tag{A.43}$$

$$\leq |\mathcal{N}_{\mathcal{P}}^{d_2}| \cdot \exp\left(-\frac{n\epsilon^2}{10K}\right) \leq 9^{rd_1} \cdot \exp\left(-\frac{n\epsilon^2}{10K}\right). \tag{A.44}$$

Similarly, we can prove

$$\mathbb{P}\{\Omega_{\mathcal{P}}^{d_1}\} \leq 9^{rd_2} \cdot \exp\left(-\frac{n\epsilon^2}{10K}\right). \tag{A.45}$$

Next we generalize the result to an arbitrary subset $\mathcal{P} \subset \{1, .., d_2\}$ or $\mathcal{P} \subset \{1, .., d_2\}$ with cardinality r. We define $\mathcal{P}_1 = \{1, ..., d_1\}$, $\mathcal{P}_2 = \{1, ..., d_2\}$, and

$$\Omega'_{\mathcal{P}} := \left\{\sup_{\mathbf{M}\in\mathcal{S}_{\mathbf{M}}^{d_1\times d_2}\cap\mathcal{B}_{\mathbf{M}}(r)} |\operatorname{vec}(\mathbf{M})^\top (\mathbf{A} - \mathbf{A}^*) \operatorname{vec}(\mathbf{M})| \geq \epsilon\right\}.$$

We can obtain

$$\mathbb{P}\{\Omega'_{\mathcal{P}}\} \leq \sum_{\mathcal{K}\in\{\mathcal{P}_1,\mathcal{P}_2\},\mathcal{P}\subset\mathcal{K}} \mathbb{P}\{\Omega_{\mathcal{P}}\} \tag{A.46}$$

$$\begin{aligned}&\leq 2\binom{\max(d_1,d_2)}{r} \max(\mathbb{P}\{\Omega_{\mathcal{P}}^{d_1}\}, \mathbb{P}\{\Omega_{\mathcal{P}}^{d_2}\}) \\ &\leq 9^{r\max(d_1,d_2)}\binom{\max(d_1,d_2)}{r} \cdot 2\exp\left(-\frac{n\epsilon^2}{10K}\right).\end{aligned} \tag{A.47}$$

Setting $\epsilon = \sqrt{\frac{Kr\max(d_1,d_2)\log\max(d_1,d_2)}{n}}$, from the sufficient largeness assumption of n implying $\frac{\epsilon^2}{K^2} \leq \frac{\epsilon}{K}$ and $\binom{\max(d_1,d_2)}{r} \sim \left(\frac{\max(d_1,d_2)}{r}\right)^r$, we obtain

$$\rho(\mathbf{A} - \mathbf{A}^*, r) = \mathcal{O}_p\left(\sqrt{\frac{r\max(d_1,d_2)\log\max(d_1,d_2)}{n}}\right), \tag{A.48}$$

which completes the proof. ■

REFERENCES

[1] T. Zhao, Z. Wang, and H. Liu, "Nonconvex low rank matrix factorization via inexact first order oracle," in *Proc. of Advances in Neural Information Processing Systems*, 2015.

[2] D. Goldberg, D. A. Nichols, B. M. Oki, and D. B. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, vol. 35, no. 12, pp. 61–70, 1992.

[3] R. M. Bell, Y. Koren, and C. Volinsky, "The bellkor 2008 solution to the netflix prize," *Statistics Research Department at AT&T Research*, vol. 1, 2008.

[4] J. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[5] R. M. Bell and Y. Koren, "Scalable collaborative filtering with jointly derived neighborhood interpolation weights," in *Proc. of the 7th IEEE International Conference on Data Mining (ICDM 2007)*, pp. 43–52, Oct. 2007.

[6] Y. Chen and Y. Chi, "Harnessing structures in big data via guaranteed low-rank matrix estimation," *arXiv preprint arXiv:1802.08397*, 2018.

[7] M. Neykov, Z. Wang, and H. Liu, "Agnostic estimation for misspecified phase retrieval models," in *Proc. of Advances in Neural Information Processing Systems 29*, pp. 4089–4097, 2016.

[8] L. Ambrosio and G. Dal Maso, "A general chain rule for distributional derivatives," *Proceedings of the American Mathematical Society*, vol. 108, no. 3, pp. 691–702, 1990.

[9] M. Hristache, A. Juditsky, and V. Spokoiny, "Direct estimation of the index coefficient in a single-index model," *Annals of Statistics*, pp. 595–623, 2001.

[10] P. Yu, J. Du, and Z. Zhang, "Single-index partially functional linear regression model," *Statistical Papers*, pp. 1–17, 2018.

[11] G. Feng, B. Peng, L. Su, and T. T. Yang, "Semi-parametric single-index panel data models with interactive fixed effects: Theory and practice," *Journal of Econometrics*, 2019.

[12] N. Vaswani, S. Nayer, and Y. C. Eldar, "Low-rank phase retrieval," *IEEE Trans. Signal Processing*, vol. 65, no. 15, pp. 4059–4074, 2017.

[13] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[14] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proc. of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, pp. 665–674, 2013.

[15] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.

[16] S. Tu, R. Boczar, M. Simchowitz, M. Soltanolkotabi, and B. Recht, "Low-rank solutions of linear matrix equations via procrustes flow," in *Proc. of the 33nd International Conference on Machine Learning*, pp. 964–973, 2016.

[17] S. M. Kakade, V. Kanade, O. Shamir, and A. Kalai, "Efficient learning of generalized linear and single index models with isotonic regression," in *Proc. of Advances in Neural Information Processing Systems*, pp. 927–935, 2011.

[18] M. Soltani and C. Hegde, "Fast low-rank matrix estimation without the condition number," *arXiv preprint arXiv:1712.03281*, 2017.

[19] G. Lecué, S. Mendelson, et al., "Minimax rate of convergence and the performance of empirical risk minimization in phase recovery," *Electronic Journal of Probability*, vol. 20, 2015.

[20] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[21] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[22] T. T. Cai, X. Li, Z. Ma, et al., "Optimal rates of convergence for noisy sparse phase retrieval via thresholded wirtinger flow," *The Annals of Statistics*, vol. 44, no. 5, pp. 2221–2251, 2016.

[23] Y. Chen and E. Candes, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," in *Proc. of Advances in Neural Information Processing Systems*, pp. 739–747, 2015.

[24] S. Becker, V. Cevher, and A. Kyrillidis, "Randomized low-memory singular value projection," *arXiv preprint arXiv:1303.0167*, 2013.

[25] T. Bendory, R. Beinert, and Y. C. Eldar, "Fourier phase retrieval: Uniqueness and algorithms," in *Compressed Sensing and its Applications*, pp. 55–91. Springer, 2017.

[26] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *Applied and Computational Harmonic Analysis*, vol. 39, no. 2, pp. 277–299, 2015.

[27] Y. Chen, X. Yi, and C. Caramanis, "A convex formulation for mixed regression with two components: Minimax optimal rates," in *Annual Conference on Learning Theory, 2014*, pp. 560–604, 2014.

[28] X. Li and V. Voroninski, "Sparse signal recovery from quadratic measurements via convex programming," *SIAM Journal on Mathematical Analysis*, vol. 45, no. 5, pp. 3019–3033, 2013.

[29] H. Ohlsson, A. Y. Yang, R. Dong, and S. S. Sastry, "Compressive phase retrieval from squared output measurements via semidefinite programming," *arXiv preprint arXiv:1111.6323*, 2011.

[30] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Proc. of Advances in Neural Information Processing Systems*, pp. 2796–2804, 2013.

[31] H. Zhang, Y. Zhou, Y. Liang, and Y. Chi, "Reshaped wirtinger flow and incremental algorithms for solving quadratic systems of equations," *arXiv preprint*, 2006.

[32] H. Zhang and Y. Liang, "Reshaped wirtinger flow for solving quadratic system of equations," in *Proc. of Advances in Neural Information Processing Systems*, pp. 2622–2630, 2016.

[33] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 773–794, 2018.

[34] X.-T. Yuan and T. Zhang, "Truncated power method for sparse eigenvalue problems," *Journal of Machine Learning Research*, vol. 14, no. Apr, pp. 899–925, 2013.

[35] C. Thrampoulidis, E. Abbasi, and B. Hassibi, "Lasso with non-linear measurements is equivalent to one with linear measurements," in *Proc. of Advances in Neural Information Processing Systems*, pp. 3420–3428, 2015.

[36] Y. Plan and R. Vershynin, "The generalized lasso with non-linear observations," *IEEE Transactions on Information Theory*, vol. 62, no. 3, pp. 1528–1537, 2016.

[37] R. Ganti, N. Rao, R. M. Willett, and R. Nowak, "Learning single index models in high dimensions," *arXiv preprint arXiv:1506.08910*, 2015.

[38] F. Han and H. Wang, "Provable smoothing approach in high dimensional generalized regression model," *arXiv preprint arXiv:1509.07158*, 2015.

[39] S. Nayer, P. Narayanamurthy, and N. Vaswani, "Phaseless PCA: Low-rank matrix recovery from column-wise phaseless measurements," in *Proc. International Conference on Machine Learning*, pp. 4762–4770, 2019.

[40] Z. Hao and A. AghaKouchak, "Multivariate standardized drought index: a parametric multi-index model," *Advances in Water Resources*, vol. 57, pp. 12–18, 2013.

[41] T. E. Booth, "Power iteration method for the several largest eigenvalues and eigenfunctions," *Nuclear Science and Engineering*, vol. 154, no. 1, pp. 48–62, 2006.

[42] M. Janzamin, H. Sedghi, and A. Anandkumar, "Score function features for discriminative learning: Matrix and tensor framework," *arXiv preprint arXiv:1412.2863*, 2014.

[43] C. Stein, P. Diaconis, S. Holmes, G. Reinert, et al., "Use of exchangeable pairs in the analysis of simulations," pp. 1–25, 2004.

[44] Z. Yang, K. Balasubramanian, Z. Wang, and H. Liu, "Learning non-Gaussian multi-index model via second-order stein's method," in *Proc. of Advances in Neural Information Processing Systems*, 2017.

[45] M. H. Gutknecht, "A brief introduction to krylov space methods for solving linear systems," in *Frontiers of Computational Science*, pp. 53–62. Springer, 2007.

[46] Y. Plan, R. Vershynin, and E. Yudovina, "High-dimensional estimation with geometric constraints," *Information and Inference: A Journal of the IMA*, vol. 6, no. 1, pp. 1–40, 2017.

[47] G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3, 2012.

[48] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *CoRR*, vol. abs/1011.3027, 2010.

[49] S. Negahban, M. J. Wainwright, et al., "Estimation of (near) low-rank matrices with noise and high-dimensional scaling," *The Annals of Statistics*, vol. 39, no. 2, pp. 1069–1097, 2011.

[50] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *arXiv preprint arXiv:1011.3027*, 2010.

[51] X. Li, R. Arora, H. Liu, J. Haupt, and T. Zhao, "Nonconvex sparse learning via stochastic optimization with progressive variance reduction," *arXiv preprint arXiv:1605.02711*, 2016.

[52] L. Mirsky, "A trace inequality of john von neumann," *Monatshefte für Mathematik*, vol. 79, no. 4, pp. 303–306, 1975.

PROOF OF REMARK III.3

Proof: When $n = \Omega_p\left(\left(\frac{1+c}{\delta\lambda+(1+c)\lambda}\right)^2 r\max(d_1,d_2)\log\max(d_1,d_2)\right)$ for some $0 < c < 1$, by the definition of $\rho(\mathbf{E}, r)$ and Theorem III.6, we have

$$\rho(\mathbf{E}, r) < \frac{\delta\lambda + (1+c)\lambda}{1+c}, \tag{A.49}$$

where we assume the constants in notations Ω_p and $\mathcal{O}_p$ to be 1. By the definition of $\gamma(r)$, we obtain

$$\gamma(r) < c. \tag{A.50}$$

Furthermore, if

$$\sqrt{\left(1 + \frac{2\sqrt{r^*}}{\sqrt{r_0 - r^*}}\right) \cdot \left(1 - \frac{(1-c^2)\omega(1+\omega)}{2}\right)} < 1 \tag{A.51}$$

holds, we can guarantee $\beta < 1$. (A.51) can be equivalently written as

$$g(\omega) = \omega(\omega+1) > \frac{4}{\left(2 + \sqrt{\frac{r_0}{r^*} - 1}\right)(1-c^2)}. \tag{A.52}$$

Note that $g(\omega)$ as a function of ω is increasing on $(0, 1)$. Hence, to make the appropriate $\omega \in (0, 1)$ exist, we require

$$\left(2 + \sqrt{\frac{r_0}{r^*} - 1}\right)(1-c^2) > 2. \tag{A.53}$$

After simplification we obtain

$$r_0 > \left(1 + \left(\frac{2}{1-c^2} - 2\right)^2\right) \cdot r^*, \tag{A.54}$$

which concludes the proof. ■

ADDITIONAL QUANTITATIVE RESULTS

In this section, we provide all the reconstruction errors in Figure 7 to demonstrate the fair performance of our algorithm.

TABLE III: Additional comparison of reconstruction errors with other related works.

Reconstruction Error	STPower	FGD	SVD	SVDs	MAPLE
digit, f_2	0.514	1.431	0.897	1.352	1.368
digit, f_3	0.372	0.814	1.374	1.403	0.898
airplane, f_2	0.582	1.273	1.273	1.416	1.389
airplane, f_3	0.417	1.273	1.368	1.425	1.274
car, f_2	0.588	1.376	1.457	1.457	1.457
car, f_3	0.363	1.457	1.491	1.491	1.457
horse, f_2	0.380	1.434	1.462	1.205	1.462
horse, f_3	0.372	1.462	1.453	1.429	1.462

We measure the quantitative performance of the algorithms using the reconstruction error, i.e., $\|\bar{\mathbf{M}} - \mathbf{M}^*\|_F$, where the Frobenius norm of each matrix, i.e., $\bar{\mathbf{M}}$ and $\mathbf{M}^*$ has been normalized to 1.