Leveraged Matrix Completion With Noise

Xinjian Huang, Weiwei Liu, Senior Member, IEEE, Bo Du, Senior Member, IEEE, and Dacheng Tao, Fellow, IEEE

Abstract—Completing low-rank matrices from subsampled measurements has received much attention in the past decade. Existing works indicate that $O(nr \log^2(n))$ datums are required to theoretically secure the completion of an $n \times n$ noisy matrix of rank $r$ with high probability, under some quite restrictive assumptions: 1) the underlying matrix must be incoherent and 2) observations follow the uniform distribution. The restrictiveness is partially due to ignoring the roles of the leverage score and the oracle information of each element. In this article, we employ the leverage scores to characterize the importance of each element and significantly relax assumptions to: 1) not any other structure assumptions are imposed on the underlying low-rank matrix and 2) elements being observed are appropriately dependent on their importance via the leverage score. Under these assumptions, instead of uniform sampling, we devise an ununiform/biased sampling procedure that can reveal the “importance” of each observed element. Our proofs are supported by a novel approach that phrases sufficient optimality conditions based on the Golfing scheme, which would be of independent interest to the wider areas. Theoretical findings show that we can provably recover an unknown $n \times n$ matrix of rank $r$ from just about $O(nr \log^2(n))$ entries, even when the observed entries are corrupted with a small amount of noisy information. The empirical results align precisely with our theories.

Index Terms—Leverage score, low rank, matrix completion, noise.

I. INTRODUCTION

MATRIX completion [1], [2], [3], [4], [5], [6], [7], [8], [9] is a fundamental task aimed at recovering a low-rank matrix from a small subset of its elements. This problem has attracted considerable attention from researchers due to its wide range of applications across various fields, including image processing [10], [11], recommendation systems [12], multitask learning [13], [14], dimensionality reduction [15], clustering and localization in sensor networks [16], drug-target interactions prediction [17], and traffic speed estimation [18], to name a few. However, in real-world application, noise is an inevitable factor during the data acquisition process, as it reflects the uncertainty of the environment and/or the measurement processes. For example, in the context of the Netflix problem, users’ ratings are uncertain [19]. Similarly, in the positioning problem, local instances are imperfect [20]. Additionally, in the acquisition of the functional MRI, its signal may be contaminated by subject motion artifacts [21].

To formulate the above-mentioned problem precisely, imagine that we consider the scenario where our objective is to recover an unknown low-rank matrix $\hat{L} \in \mathbb{R}^{n_1 \times n_2}$ from a collection of partially observed and corrupted entries $X$ as follows:

$$X_{ij} = \hat{L}_{ij} + \hat{S}_{ij}, \quad (i,j) \in \mathcal{O}$$

where $\hat{S} \in \mathbb{R}^{n_1 \times n_2}$ is a matrix representing random noises, and we only observe entries over an index subset $\mathcal{O} \subseteq [n_1] \times [n_2]$ with $[n_1] := \{1, \ldots, n_1\}$ and $[n_2] := \{1, \ldots, n_2\}$. The aim is to reliably recover $\hat{L}$ given the incomplete and even grossly corrupted data.

Candès and Plan [22] first showed that this goal can be achieved by means of a principled convex program:

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \quad \text{s.t.} \quad \mathcal{P}_\mathcal{O}(X) = \mathcal{P}_\mathcal{O}(L) + S$$

where $\|L\|_*$ denotes the nuclear norm (the sum of the singular values) of $L$, $\|S\|_1 = \sum_{i,j} |S_{ij}|$ denotes the entrywise $l_1$ norm and $\mathcal{P}_\mathcal{O}(-)$ is the orthogonal projection operator which keeps the elements in $\mathcal{O}$ invariant and 0 otherwise; and $X \in \mathbb{R}^{n_1 \times n_2}$ is the matrix containing the known entries (with values known only on $\mathcal{O}$). Specifically, they suggest that at least $O(nr \log^2(n))$ observations are required to theoretically secure completing an $n \times n$ noisy matrix of rank $r$ with high probability under the assumptions that the observed elements follow the uniform distribution and that the low-rank matrix to be recovered should satisfy incoherence property [23]. The incoherence property, while useful in various applications, such as non-negative matrix factorization [24] and face recognition [25], necessitates that the row or column spaces of the considered data should be diffuse [22], [26].

Unfortunately, these restrictive assumptions do not hold in the majority of real-world applications. For example, in the sensor network localization problem [27], information observed by the key nodes is evidently more informative than that observed by the edge nodes [28]. Thus, it is necessary...
to consider the “importance” of elements being observed during the sampling processing, which may result in nonuniform sample processing. In this context, leverage score can offer a promising solution. It was first introduced to detect outliers in regression diagnostics [29], and has proven successful in analyzing the large-scale data and implementing randomized matrix due to its capacity to measure the correlation of the dominant subspace with the canonical basis [30]. For an element in the matrix, its “importance” thus can be characterized by the sum of the leverage scores of its corresponding row and column.

In this article, we address two interconnected questions simultaneously.

1) Can we devise a specific sampling strategy suitably dependent on the “importance” of each observed entry?

2) Does this sampling strategy works effectively even under noisy cases?

Indeed, we show that achieving reasonably accurate matrix completion from noisy sampled entries is feasible, given that the sampling process incorporates a biased distribution based on the leverage scores of the target low-rank matrix. Specifically, we consider a noisy low-rank matrix completion problem, in which each sampled element follows a specific distribution defined by its leverage scores, and the incoherence property of the low-rank matrix is nonessential in our approach. Under these reasonable conditions, we devise a more compact sample complexity $O(nr \log^2(n))$ with respect to noisy low-rank matrix completion. This sampling upper bound is derived mainly on the basis of Golffing scheme (an elegant technique to construct dual certificates [31], [32], [33]) and several concentration inequalities involving two norms defined by the leverage score. Our theoretical findings are further validated through empirical results.

The main contributions of this article are summarized as follows.

1) Our theoretical results show that if the sampling probability follows a biased distribution determined by the row and column leverage scores of the underlying matrix, only $O(nr \log^2(n))$ observed entries are needed to exactly recover the underlying low-rank matrix with high probability, even in the presence of corruption in a subset of the observed entries (Theorem 1).

2) Our proof techniques introduce a new method for formulating sufficient optimality conditions (Lemma 1) based on the Golffing scheme by constructing a particular matrix-valued random process that converges to the dual certificate $Y$ (Section V-B).

3) Two norms, which involve leverage scores, are presented to better characterize the probability of each element being observed. Furthermore, several concentration properties [34] and upper bounds with respect to these norms are also rigorously derived (Section V-C).

The remainder of the article is organized as follows. In Section II, we briefly review some related work. Section III elaborates on some preliminaries about the problem we considered. Our main theoretical findings are presented in Section IV. Proof of our theoretical results are provided in Section V. Section VI reports our empirical results. Section VII concludes the work. The proofs of some theorems are available in the Appendix in the supplementary material.

II. RELATED WORK

Matrix completion has been extensively studied in both theory and algorithm due to its wide application in numerous scenarios. If $S$ in the convex optimization (2) equals to 0, then it involves the following realistic matrix completion problem

$$\min_{L} \|L\|_*$$

s.t. \( P_O(L) = P_O(X) \).

Next, we first give some notations utilized in this article, and then briefly review some literature on matrix completion with and without noises from both theory and algorithm perspectives.

Notations: \( X_{ij} \) denotes the \((i, j)\)th element of a matrix \( X \in \mathbb{R}^{m \times n} \), \( X_i \) and \( X_j \) are the \(i\)th row and \(j\)th column of \( X \), respectively. \( X^* \) denotes the transpose of \( X \). There are five norms associated with a matrix \( X \): 1) \( \|X\|_F \) denotes the Frobenius norm; 2) \( \|X\|_* \) denotes the nuclear norm; 3) \( \|X\|_1 \) denotes the spectral norm and \( \|X\|_1 \) represent the \( l_1 \); and 4) \( \|X\|_\infty \) represent the \( l_\infty \) norms of the long vector stacked by \( X \). The inner product between two matrices is \( \langle X, Y \rangle = \text{trace}(X^*Y) \). We denote by Range(\( P \)) the range of an operator \( P \). A linear operator \( A \) acts on the space of matrices and \( \|A\| \) denotes the operator norm given by \( \|A\| = \sup_{\|X\|_F = 1} \|A(X)\|_F \).

A. Theory

In general, restoration of a matrix from a small amount of observations is impossible. However, if the unknown matrix has low-rank structure, then accurate and even exact recovery is possible.

Candès and Recht [1] provide the first algorithm and theoretical guarantees for low-rank matrix completion, where they show that the nuclear norm minimization problem (3) works when the low-rank matrix is incoherent and the sampling process is uniform and independent of the matrix, obtaining that the underlying low-rank matrix can be exactly recovered with high probability from only $O(n^{1.25}r \log(n))$ entries. Subsequent works have refined provable completion results that $O(nr \log^2(n))$ entries are needed for incoherent matrices recovery under the uniform random sampling model [31], [32], [35]. The same complexity has also been obtained under a more suitable setting. Chen et al. [36] proved that any coherent matrix can be exactly recovered from $O(nr \log^2(n))$ entries if the random sampling processing follows a specific biased distribution. If the row space is coherent and the column space is still incoherent, Krishnamurthy and Singh [37] established that merely $O(n^{1.5} \log(n))$ entries can exactly recover the low-rank matrix with high probability under an adaptive sampling strategy. Further, they also prove that $O(nr^{1.5} \log(r))$ entries can recover the noisy low-rank matrix. For life-long or online matrix completion, Balcan and Zhang [38] established an optimal guarantee that exactly recovers an $\mu$-incoherent matrix by probability at least $1 - \delta$ with sample complexity $O(\mu n \log(r/\delta))$ in the
TABLE I
SHORT SUMMARY OF SEVERAL UPPER BOUNDS FOR LOW-RANK MATRIX COMPLETION UNDER DIFFERENT SETTINGS

| Ref. | Noisy | Sampling | Assumption | Upper Bound |
|------|-------|----------|------------|-------------|
| [1]  | no    | uniform  | incoherence| $O(n^{1.5}r\log(n))$ |
| [35] | no    | uniform  | incoherence| $O(nr\log^{2}(n))$ |
| [22] | yes   | uniform  | incoherence| $O((n^2 \log^{2}(n))$ |
| [38] | yes   | uniform  | incoherence| $O((n\mu r \log^{2}(\frac{1}{\mu}))$ |
| [39] | no    | uniform  | incoherence| $O((\mu r \log^{2}(n))$ |
| [36] | no    | non-uniform| coherence | $O((nr\log^{2}(n))$ |
| [37] | no    | adaptive | coherence  | $O((n^{1.5}r \log(r))$ |
| [38] | yes   | adaptive | coherence  | $O((nr^{1.5} \log(n))$ |
| ours | yes   | non-uniform| coherence  | $O((nr\log^{2}(n))$ |

context of sparse random noise. More recently, based on leave-one-out technique, Ding and Chen [39] showed that only $O((\mu r \log(\mu r) n \log(n))$ observations are suffice to recover an $n \times n$ $\mu$-incoherence matrix of rank $r$ by using nuclear norm minimization methods. Although Ding and Chen [39] and Balcan and Zhang [38] established a tighter sampling bound for noise-free and noisy matrix completion, respectively, these results are developed under the $\mu$-incoherence assumption, a stronger assumption than coherence considered in this article.

To make a clear comparison, Table I summarizes these theoretical results. Our theoretical findings establish that the required number of observations for successfully recovering a noisy low-rank matrix is $O(nr \log^{2}(n))$, which is in the same order as the state-of-the-art results when no corruptions exist and more reasonable assumptions are satisfied.

B. Algorithm

Existing matrix completion methods can be generally classified into three categories: 1) regularization-based methods; 2) matrix factorization-based methods; and 3) others.

1) Regularization-Based Method: The regularization-based method for matrix completion involves using singular values or their variations to construct different norms to surrogate rank function. The most famous surrogation for the rank function is the nuclear norm, which corresponds to the nuclear norm-based matrix completion problem (3). The singular value threshold (SVT) method [40] and the inexact augmented Lagrange multiplier (IALM) method [41], [42] are two representative methods designed to solve the optimization problem (3). Afterward, some other regularizers are also developed to surrogate the low-rank function, such as weighted nuclear norm [43], [44], truncated nuclear norm [45], [46], Shatten $p$-norm [47], [48], weighted Shatten $p$-norm [49], and Schatten capped $p$-norm [50]. Specific algorithms are rigorously devised concerning these different norms. Note that these norms can be used not only in matrix completion but also in other computer vision tasks, such as image denoising [51], [52], [53], image inpainting [54], etc.

The key challenge of the regularization-based method is the computationally expensive calculation of singular values, especially in large scale cases. To improve the computational efficiency, matrix approximation is often used as an efficient alternative, thus inducing the following section.

2) Matrix Factorization-Based Method: This kind of method is to factorize or approximate the original matrix by the product of two or more small-sized matrices. Compared with the regularization-based methods, this type of method is usually computationally cheap and less memory-consuming.

The maximum margin factorization (MMF) method [55] is a representative factorization-based method. However, empirical studies show that the MMF method only can obtain suboptimal solutions. The low-rank matrix fitting (LMaFit) method [56] is devised by constructing a nonlinear successive over-relaxation algorithm. LMaFit only requires solving a series of linear least-squares problems, which enables this method to handle the large-scale problem well. Recently, Shang et al. [57], [58] proposed a bilinear factorization method to solve the low-rank matrix recovery problem. The method approximates the original matrix by two small-sized matrix and some specific norms, that is nuclear norm, are imposed on each small-sized matrix to induce the low-rank property. Factor group-sparse regularization is also developed for completing matrices [48]. Additionally, it proves that the Schatten-$p$ norm is the sum of two group-sparse norms, which greatly enhances computational efficiency [48]. Based on the factorization framework, Wang et al. [59] proposed a novel robust matrix completion scheme via using the truncated-quadratic loss function, and half-quadratic theory is adopted for its optimization.

3) Others: Besides the aforementioned methods, some other methods are also extensively developed. Based on a sum of multiple orthonormal side information and nuclear-norm regularization, Ledent et al. [60] proposed an interpretable approach to matrix completion with a provable convergence. To handle the data matrices with nonlinear structures, Fan and Chow [61] proposed nonlinear matrix completion (NLMC), which extends the conventional matrix completion method to nonlinear structures. For high-rank matrix completion problem, a novel online method is proposed by using kernel trick, where it maps the data into a high-dimensional polynomial feature space [62]. The proposed online method enjoys much lower space and time complexity since the data admit a low-dimensional subspace in this feature space. Based on the correntropy criterion, He et al. [63] proposed a half-quadratic alternating steepest descent (HQ-ASD) algorithm for a robust matrix completion problem (2). To further utilize the smooth Riemannian manifold of a matrix with a fixed-rank, Riemannian optimization [64] was introduced to accelerate the optimization process [65]. Based on linear latent variable models, deep matrix factorization (DMF) [66] is proposed for nonlinear matrix completion. Recently, learning-based methods have also been introduced to the task of completion [67]. For distributed matrix completion problem, Abubaker et al. [68] proposed a framework for scaling stratified SGD through significantly reducing the communication overhead. Zhang et al. [69] employed the alternating direction method of multiplier (ADMM) with two dual variables to optimize the generalized nonconvex nonsmooth low-rank matrix recovery problems.
III. PRELIMINARY

Suppose matrix $X \in \mathbb{R}^{n_1 \times n_2}$ is the sum of an underlying low-rank matrix $\bar{L} \in \mathbb{R}^{n_1 \times n_2}$ and a sparse “noises” matrix $\tilde{S} \in \mathbb{R}^{n_1 \times n_2}$. We consider the following problem: suppose we only observe a subset $O \subseteq [n_1] \times [n_2]$ of the entries of $X$; the remaining entries are unobserved. Our goal is to exactly and provably recover $\bar{L}$ from partially observed entries with noise. Formally, we focus on the matrix completion problem

$$\left(\hat{L}, \hat{S}\right) = \arg \min_{L,S} \|L\|_* + \lambda\|S\|_1$$

s.t. $\mathcal{P}_O(X) = \mathcal{P}_O(L) + S$ \hspace{1cm} (4)

where $S$ is supported on the index matrix $\Omega \subset O$, $\mathcal{P}_O(X)$ is the matrix obtained by setting the entries of $X$ that are outside the observed set $O$ to zero, and $\lambda$ is a parameter that trades off between these two elements of the objective function. The value of $\lambda$ is chosen for a theoretical guarantee of exact recovery in Theorem 1. The nuclear norm is used as a convex surrogate for the rank of a matrix and the $l_1$ norm is used as a convex surrogate for its sparsity [70].

Notice that the observed data is $\mathcal{P}_O(L) + S$, where $O \subseteq [n_1] \times [n_2]$ and $S$ is supported on $\Omega \subset O$. We assume that the index set $O$ of the observation data is obtained by nonuniform sampling with probability $p_{ij}$.

Random uniform corruption of these observations with probability $q$ yields the index set $\Omega$ of the “noise” matrix $S$. Specifically, the index matrix $O, \Omega$ and sparse “noise” matrix $S$ satisfy the following model.

**Model 1:**
1) $L$ is supported by $O \subseteq [n_1] \times [n_2]$; $O$ is determined by Bernoulli sampling with nonuniformly probability with $p_{ij}$, denoted as $O \sim \text{Ber}(p_{ij})$. That is, to say, $p_{ij}$ represents the probability that the $(i,j)$th entry of $L$ to be observed or sampled.
2) Assume that $\Omega$ is uniformly sampled from $O$ with probability $q$ and that “noise” matrix $S$ is supported by $\Omega$. In other words, given $(i,j) \in O$, we have $\mathbb{P}(i,j) \in \Omega | (i,j) \in O) = q$. This implies that $\Omega$ is determined by Bernoulli sampling with nonuniformly probability $q_{ij}$, denoted as $\Omega \sim \text{Ber}(q_{ij})$.
3) Define $\Gamma := O/\Omega$. We then have $\Gamma \sim \text{Ber}(p_{ij}(1-q))$.
4) Define $\text{sgn}(S) = \mathcal{P}_\Omega(K)$, where $K \in \mathbb{R}^{n_1 \times n_2}$ and its entries are either 1 or -1.

We assume $L$ is of rank $r$ and its reduced singular value decomposition (SVD) is denoted as $L = U\Sigma V^*$, where $U \in \mathbb{R}^{n_1 \times r}$, $\Sigma \in \mathbb{R}^{r \times r}$ and $V \in \mathbb{R}^{n_2 \times r}$. We next provide the definition of leverage scores, which is used to determine the nonuniformly sampling process.

**Definition 1 (Leverage Scores):** For a real-valued matrix $L \in \mathbb{R}^{n_1 \times n_2}$ with rank $r$, its SVD is $U\Sigma V^*$. Then its row leverage score $\mu_i(L)$ for any row $i$ and column leverage score $\nu_j(L)$ for any column $j$ are defined as

$$\mu_i = \frac{n_1}{r}\|UU^*e_i\|_2^2, \quad \text{for } i = 1, 2, \ldots, n_1$$

$$\nu_j = \frac{n_2}{r}\|VV^*e_j\|_2^2, \quad \text{for } j = 1, 2, \ldots, n_2$$

1Note that the value of $p_{ij}$ is determined by the leverage scores of each datum. Details can be seen in Section IV to follow.

where $e_i$ is the $i$th canonical basis vector in Euclidean space (the vector with all entries equal to 0 but the $i$th equal to 1) with appropriate dimension.

Note that the leverage scores of the matrix $X$ are nonnegative, and are functions of the column and row spaces of $L$. The standard coherence parameter $\mu$ of $L$ used in the previous literature [1], [23] corresponds to a global upper bound on the leverage scores, that is $\mu = \max_{ij}\{\mu_i(L), \nu_j(L)\}$. Clearly, standard coherence parameters characterize the quality of a matrix from a holistic perspective, while leverage scores from a local perspective. Therefore, the leverage scores can be considered as localized versions of the standard coherence parameter.

Two norms $\mu(\infty)$-norm and $\mu(\infty, 2)$-norm with respect to leverage scores are needed in the following concentration properties establishment. The $\mu(\infty, 2)$-norm of a matrix $Z \in \mathbb{R}^{m \times n}$ is defined as

$$\|Z\|_{\mu(\infty, 2)} := \max_{a,b} \left\{ \sqrt{m} \mu_{a}^* \|Z_{a,:}\|_2, \sqrt{n} \mu_{b}^* \|Z_{:,b}\|_2 \right\}$$

which is the maximum of the weighted column and row norms of $Z$. $\mu(\infty)$-norm of $Z$ is defined as

$$\|Z\|_{\mu(\infty)} := \max_{a,b} \|Z_{ab}\|_{\mu(\infty)}$$

which is the weighted elementwise magnitude of $Z$. $\|Z\|_{\mu(\infty, 2)}$ and $\|Z\|_{\mu(\infty)}$ are exactly norms. Detailed proofs can be seen in Appendix A in the supplementary material.

IV. LEVERAGED MATRIX COMPLETION WITH NOISE

A. Main Results

To facilitate the derivation of the upper bound for noisy matrix completion by using leverage scores, Model 1 is transformed into the following model in an equivalent manner.

**Model 2:**
1) Fix an $n \times n$ matrix $K$, whose entries are either 1 or -1.
2) Define two independent random subset of $[n_1] \times [n_2]$: a) $\Gamma' \sim \text{Ber}(p_{ij}(1-2q))$ and b) $\Omega' \sim \text{Ber}(2p_{ij}q/[1-p_{ij} + 2q_{ij}])$. Let $O := \Gamma' \cup \Omega'$, it is easy to verify that $O \sim \text{Ber}(p_{ij})$.
3) Define an random matrix $W \in \mathbb{R}^{n_1 \times n_2}$ with independent entries $W_{ij}$ satisfying $\mathbb{P}(W_{ij} = 1) = \mathbb{P}(W_{ij} = -1) = (1/2)$.
4) Define $\Omega'' \subset \Omega'$, where $\Omega'' := \{(i,j) | W_{ij} = K_{ij}, (i,j) \in \Omega'\}$. Then define $\Omega := \Omega'' \cup \Gamma' = (1/2)\Omega' \cup \Gamma'$ and $\Omega := O/\Omega$.
5) Let $\text{sgn}(S) := \mathcal{P}_\Omega(K)$.

Clearly, in both Model 1 and Model 2, if we fix $(O, \Omega)$, the whole setting is deterministic. Therefore, the probability of $(\hat{L}, \hat{S}) = (L, S)$ is determined by the joint distribution of $(O, \Omega)$. Besides, it is easy to verify that the joint distribution of $(O, \Omega)$ in the two models is identical.

We are now ready to state our main results presented as follows. For simplicity, results provided here are on the basis of square matrix with size $n \times n$; similar results can be extended to a general rectangle case in the same fashion.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Theorem 1: Under the Model 2, if each element \((i,j)\) is independently observed with probability \(p_{ij}\) and satisfies
\[
p_{ij} \geq \max \left\{ c_p \left( \frac{(\mu_i + \nu_j) r \log^2(n)}{n}, \frac{1}{n^2} \right) \right\}
\]
\[q \leq c_q \text{ and } \lambda = \left[ \frac{1}{(24 \sqrt{n \log n})} \right], \text{then } (\hat{L}, \hat{S}) \text{ is the unique optimal solution to problem (2) with probability at least } 1 - Cn^{-5} \text{ for a positive constant } C, \text{ provided that the positive constants } c_p \text{ is sufficiently large and } c_q \text{ is sufficiently small.}

Proof Sketch: Due to space constraints, we present only the outline of the theorem’s proof here. The full proof is outlined in Section V and the Appendix in the supplementary material. The high-level roadmap of the proof follows.

Remark 1: In Theorem 1, \(c_p\) and \(c_q\) are universal positive constants, whose values can be well-designed during the proofs of the corresponding inequalities or lemmas. Details can be see at Section V and the Appendixes in the supplementary material.

Remark 2: The power of our results is that one can recover a low-rank matrix with rank \(r\) from nearly minimal number of samples in the order of \(O(nr \log^2 n)\) even when a constant proportion of these samples has been corrupted. From experimental results, we know that this corruption proportion empirically approximates 0.6 by using leverage sampling, while 0.18 by using uniform sampling (see Section VI for detail). Moreover, this theorem implies that elements with higher-leverage scores should be sampled with higher probability. Informally, elements with higher-leverage scores have more “important information” of the matrix, thereby tolerating larger noise density.

Remark 3: Theorem 1 states us that only \(O(nr \log^2(n))\) datums can theoretically guarantee the exact recovery of a noisy low-rank matrix. This upper bound on sampling is consistent with to the findings in [36]. However, our results are based on the assumption of mild noise contamination in the sampled data, while the results in [36] are based on the assumption of no noise contamination. Consequently, our results can be viewed as an extension of the work in [36], with broader applicability.

B. Leverage Sampling in Practice

So far, we have established that one can exactly recover an arbitrary \(n \times n\) rank-\(r\) matrix from just about \(O(nr \log^2(n))\) observations if sampled in accordance with the leverage scores. However, partial observations (sometimes even corrupted) result in the lack of a prior knowledge about the leverage scores, preventing the implementation of leveraged sampling in practice.

To overcome the restriction on applications, we find that \(X\) and \(L\), intuitively, share the same row and column space and that perturbation or sparse noise \(\tilde{S}\) does not change the leverage scores if \(\text{range}(X) = \text{range}(L)\). These findings provide a possible way to calculate leverage scores via the observed matrix \(X\). Formally, the following theorem can theoretically ensure that leverage scores of \(L\) can be approximated via observed matrix \(X\).

Theorem 2: Let \(X\), \(L\), and \(S\) are defined in (2), with \(\text{Range}(L) = \text{Range}(X)\) and \(\delta = \|S\|_{\|L^\dagger\|} \leq (1/2)\), where \(L^\dagger\) denotes the Moore–Penrose inverse of \(L\) Then
\[
\left\| \tilde{\mu}_i - \mu_i \right\| \leq \left( 2 \sqrt{1 - \mu_i} \right) \delta, \quad \text{for } i = 1, \ldots, m \quad (5)
\]
and
\[
\left\| \tilde{\nu}_j - \nu_j \right\| \leq \left( 2 \sqrt{1 - \nu_j} \right) \delta, \quad \text{for } j = 1, \ldots, n \quad (6)
\]
where \(\tilde{\mu}_i\) and \(\tilde{\nu}_j\) denote the \(i\)th row and \(j\)th column leverage score of \(X\); and \(\mu_i\) and \(\nu_j\) denote the \(i\)th row and \(j\)th column leverage score of \(L\).

Proof Sketch: Based on principal angle theory [71], it can be observed that the contribution of noise \(\tilde{S}\) to the range of \(L\) is limited. Then by the oracle information of the leverage scores and [71, Th. 2.4], our results hold under appropriate assumptions. Detailed proofs can be seen in Appendix B in the supplementary material.

Remark 4: From Theorem 2, we get that the relative leverage score difference between \(L\) and \(X\) is bounded by a very small constant under some suitable conditions. Thus, we can directly use the observation data to calculate the leverage score by existing methods [36], [72]. For the number of samples, we can rank the leverage scores in a descending order and then select the top \(N\) samples, as long as \(N\) satisfies the sampling upper bound provided in Remark 2.

Based on Theorem 2 and the existing methods [36], [72], we propose Algorithm 1 for leveraged matrix completion with noise: first estimating the leverage scores of the noisy matrix from a small number of uniform samples, then using these estimated leverage scores to select the remaining samples. Specifically, given a total budget of \(N\) samples, draws a subset \(O\) uniformly without replacement such that \(|O| = \theta N\), where \(\theta \in [0, 1]\) denotes the fraction of the budget to estimate the leverage scores of the underlying matrix. Then take a rank-\(r\) approximation to \(P_O(X)\), obtaining the estimated leverage scores \(\hat{\mu}_i = \mu_i(U\hat{S}^\dagger V^T)\) \(\nu_j = \nu_j(U\hat{S}^\dagger V^T)\). Last, generate the remaining \((1 - \theta)N\) samples by sampling without replacement with distribution \(p_{ij} \propto (\hat{\mu}_i + \hat{\nu}_j)r \log^2(n)/n\), obtaining the new set of samples \(\hat{O}\). Take a union of \(O\) and \(\hat{O}\) to construct the observation data \(P_{O \cup \hat{O}}(X)\) as constraints in robust matrix completion problem (2).

V. PROOF OF THEOREM 1

Our analysis of nonuniform error bound is based on leverage scores, where \(\mu(\infty)\)-norm and \(\mu(\infty, 2)\)-norm are utilized.
Algorithm 1: Leverage Sampling for Matrix Completion With Noise

**Input:** Noisy data $X$, rank $r$, sampling budget $N$, $\theta$

1. Draw a subset $\mathcal{O}$ by sampling uniformly without replacement such that $|\mathcal{O}|=\theta N$.
2. Compute rank-$r$ approximation to $\mathcal{P}_O(X)$, $\bar{U}\bar{\Sigma}\bar{V}^T$.
3. Calculate the estimated leverage scores $\bar{\mu}_i = \mu_i(\bar{U}\bar{\Sigma}\bar{V}^T)$ and $\bar{v}_j = v_j(\bar{U}\bar{\Sigma}\bar{V}^T)$.
4. Generate the remaining $(1-\theta)N$ samples by sampling without replacement with distribution $p_{ij} \propto (\bar{\mu}_i + \bar{v}_j)r\log^2(n)/n$.
5. Replace $\mathcal{P}_O$ with $\mathcal{P}_{O,\hat{S}}(X)$ in problem (2) to obtain

$$
\min_{L, S} \|L\|_* + \lambda \|S\|_1 \\
\text{s.t.} \quad \mathcal{P}_{O,\hat{S}}(X) = \mathcal{P}_{O,\hat{S}}(L) + S.
$$

6. Solve problem (7) to obtain $\hat{L}$.

**Output:** Recovered problem (7) to obtain $\hat{L}$.

B. Construction of the Dual Certificate

Since $O := \Gamma \cup \Omega'$, we know that $\mathcal{P}_{O,\Gamma'}(W) = \mathcal{P}_{\Omega'}(W)$. From Model 2, we know that the distribution of $(\Gamma', \mathcal{P}_{\Omega'}(W))$ and $(\Gamma', 2\mathcal{P}_{\Omega'}(W) - \mathcal{P}_O(W))$ are same.

Suppose there exist $Y_1$ and $Y_2$ satisfying

$$
\|\mathcal{P}_{\Gamma'}(Y_1 + Y_2) + \mathcal{P}_{\Gamma'}[\lambda(2\mathcal{P}_{\Omega'}(W) - 2UV^*)]\|_F \\
\leq \|\mathcal{P}_{\Gamma'}(Y_1) + \mathcal{P}_{\Gamma'}[\lambda(2\mathcal{P}_{\Omega'}(W) - \mathcal{P}_O(W) - UV^*)]\|_F \\
+ \|\mathcal{P}_{\Gamma'}(Y_2) + \mathcal{P}_{\Gamma'}(\lambda\mathcal{P}_O(W) - UV^*)\|_F
$$

and

$$
\|\mathcal{P}_{\Gamma'}(Y_1 + Y_2) + \mathcal{P}_{\Gamma'}[\lambda(2\mathcal{P}_{\Omega'}(W))]\|_F \\
\leq \|\mathcal{P}_{\Gamma'}(Y_1) + \mathcal{P}_{\Gamma'}[\lambda(2\mathcal{P}_{\Omega'}(W) - \mathcal{P}_O(W))]\|_F \\
\leq \|\mathcal{P}_{\Gamma'}(Y_2) + \mathcal{P}_{\Gamma'}(\lambda\mathcal{P}_O(W))\|_F
$$

Then we know that $Y = [(Y_1 + Y_2)/2]$ satisfies the condition (8). To prove Theorem 1, we need to prove that there exists $Y$ satisfying

$$
\|\mathcal{P}_{\Gamma'}(Y) + \mathcal{P}_{\Gamma'}[\lambda\mathcal{P}_O(W) - UV^*)\|_F \leq \frac{\lambda}{n^3} \\
\|\mathcal{P}_{\Gamma'}(Y) + \mathcal{P}_{\Gamma'}[\lambda\mathcal{P}_O(W)]\|_F \leq \frac{1}{4} \\
\|\mathcal{P}_{\Gamma'}(Y)\|_\infty \leq \frac{\lambda}{4} \\
\mathcal{P}_{\Gamma'}(Y) = 0
$$

with high probability under the assumptions of Lemma 1.

Notice that $1' \sim \text{Ber}(p_2t(1 - 2q))$. Suppose that $\rho$ satisfies $1 - p_2(1 - 2q) = (1 - (p_2(1 - 2q))/6)(1 - \rho)^{-2}$, where $t = \lfloor 5\log n + 1 \rfloor$. We know that $\rho \leq (C\rho)/(|\log n|)$. Define $\rho_1 = \rho_2 = (p_2(1 - 2q))/6$, $\rho_3 = \cdots = \rho_t = \rho$. Let $\Gamma' = \Gamma_1 \cup \Gamma_2 \cup \ldots \cup \Gamma_t$, where $\Gamma_k \sim \text{Ber}(\rho_k)$ independently.

Construct

$$
\begin{align*}
X_0 &= \mathcal{P}_{\Gamma'}(UV^* - \lambda\mathcal{P}_{\Omega'}(W)) \\
X_k &= \left(\mathcal{P}_{\Gamma'} - \frac{1}{\rho_k}\mathcal{P}_{\Gamma_k}\mathcal{P}_{\Gamma'}\right)X_{k-1} \\
Y &= \sum_{k=1}^t \frac{1}{\rho_k}\mathcal{P}_{\Gamma_k}X_{k-1}
\end{align*}
$$

By this construction, we see that

$$
X_k = \mathcal{P}_{\Gamma'}(X_k), \quad k = 0, 1, \ldots, t
$$

which implies that $X_k$ is in the range of $\mathcal{P}_{\Gamma'}(X_k)$.

C. Validity of the Dual Certificate

We next show that $Y$ satisfies all the constraints (9)–(12) simultaneously under our assumptions. The inequality (12) is immediately held by the construction of $Y$. Before validating the constraints (9)–(11), we present some lemmas.

Lemma 2 (Matrix Bernstein Inequality [74]): Let $X_1, X_2, \ldots, X_n \in \mathbb{R}^{n_1 \times n_2}$ be independent zero mean random matrices. Suppose

$$
\max\left\{\left\|\mathbb{E}\sum_{k=1}^n X_kX_k^*\right\|_F, \left\|\mathbb{E}\sum_{k=1}^n X_k^T X_k\right\|_F\right\} \leq \sigma^2
$$

A. Sufficient Condition for Optimality

We first derive the first order subgradient sufficient conditions for Problem (2) as below.

**Lemma 1:** If $p_{ij} \geq \max\{c_\rho(\mu_i + v_j)r\log^2(n)/n, (1/n^5)\}$, $\lambda = 1/(2\sqrt{n}\log n)$ and there exists a dual variable $Y \in \mathbb{R}^{n \times n}$ satisfying

$$
\begin{align*}
\|\mathcal{P}_{\Gamma'}(Y + \lambda\mathcal{P}_{\Omega'}(W) - UV^*)\|_F &\leq \frac{\lambda}{n^3} \\
\|\mathcal{P}_{\Gamma'}(Y + \lambda\mathcal{P}_{\Omega'}(W))\|_F &\leq \frac{1}{4} \\
\|\mathcal{P}_{\Gamma'}(Y)\|_\infty &\leq \frac{\lambda}{4} \\
\mathcal{P}_{\Gamma'}(Y) &\neq 0
\end{align*}
$$

then the solution $(\hat{L}, \hat{S})$ is the unique optimal solution to the original optimization problem (2).

**Proof:** The proof details of this lemma can be found in Appendix C in the supplementary material.
and $\|X_k\| \leq B$ almost surely for all $k$. Then we have
\[ \mathbb{P}\left\{ \left\| \sum_{k=1}^{n} X_k \right\| \geq t \right\} \leq (n_1 + n_2) \exp\left( \frac{-t^2/2}{Bt/3 + \sigma^2} \right). \]
As a consequence, for any $c > 0$, we have
\[ \left\| \sum_{k=1}^{n} X_k \right\| \leq 2\sqrt{\alpha^2 \log(n_1 + n_2) + cB \log(n_1 + n_2)} \]
with probability at least $1 - (n_1 + n_2)^{1-c}$.

**Lemma 3:** If $p_{ij} \geq c_p((\mu_i + \nu_j)r\log^2(n))/n$ for all $(i, j)$ and $q \leq c_q$, then with high probability
\[ \left\| \frac{1}{(1 - 2q)p_{ij}} \mathcal{P} \mathcal{P} \mathcal{P} \mathcal{P} - \mathcal{P} \right\| \leq \frac{1}{2} \]
provided that $c_p$ is sufficiently large and $c_q$ is sufficiently small.

From Lemma 3, we have
\[ \left\| X_k \right\|_F \leq \frac{1}{2} \left\| X_{k-1} \right\|_F , \quad k = 1, 2, \ldots, t \]
with high probability, provided $c_p$ is sufficiently large and $c_q$ is sufficiently small. Therefore, it is easy to obtain
\[ \left\| X_k \right\|_F \leq \frac{1}{2^k} \left\| X_0 \right\|_F , \quad k = 1, 2, \ldots, t. \]

**Lemma 4:** Suppose $Z$ is a fixed $n \times n$ matrix and $Z \in \text{Range}(\mathcal{P})$. If $p_{ij} \geq c_p((\mu_i + \nu_j)r\log^2(n))/n$ for all $(i, j)$ and $q \leq c_q$, then with high probability
\[ \left\| \frac{1}{(1 - 2q)p_{ij}} \mathcal{P} \mathcal{P} \mathcal{P} \mathcal{P} - \mathcal{P} \right\| \leq \frac{1}{2} \left( \|Z\|_{\mu(\infty)} + \|Z\|_{\mu(\infty, 2)} \right) \]
provided that $c_p$ is sufficiently large and $c_q$ is sufficiently small.

**Proof:** The proof details of this lemma are provided in Appendix D in the supplementary material.

**Lemma 5:** Suppose $Z$ is a fixed $n \times n$ matrix and $Z \in \text{Range}(\mathcal{P})$. If $p_{ij} \geq c_p((\mu_i + \nu_j)r\log^2(n))/n$ for all $(i, j)$ and $q \leq c_q$, then with high probability
\[ \left\| \frac{1}{(1 - 2q)p_{ij}} \mathcal{P} \mathcal{P} \mathcal{P} \mathcal{P} - \mathcal{P} \right\|_{\mu(\infty)} \leq \frac{1}{2} \|Z\|_{\mu(\infty)} \]
provided that $c_p$ is sufficiently large and $c_q$ is sufficiently small.

**Proof:** The proof details of this lemma are provided in Appendix E in the supplementary material.

**Corollary 1:** Suppose $Z$ is a fixed $n \times n$ matrix and $Z \in \text{Range}(\mathcal{P})$. If there exists a $\alpha > 0$ such that $p_{ij} \geq c_0((\mu_i + \nu_j)r)/[\alpha n^2]$ $\geq c((\mu_i + \nu_j)r\log^2(n))/n$ for all $(i, j)$, provided $c_0$ and $c$ is sufficiently large and $q$ is sufficiently small. Then with high probability
\[ \left\| \frac{1}{(1 - 2q)p_{ij}} \mathcal{P} \mathcal{P} \mathcal{P} \mathcal{P} - \mathcal{P} \right\|_{\mu(\infty)} \leq \frac{\alpha}{2} \|Z\|_{\mu(\infty)}. \]

Note that this Corollary can be seen as a generalization of [33, Lemma 3.1].

Armed with the above lemmas, dual certifications of (9)-(11) can be extensively conducted, which is provided in Appendix F in the supplementary material.

**VI. EXPERIMENTS**

In this section, we provide numerical experiments for solving extensive synthetic and real-world problems to demonstrate the effectiveness of our proposed leveraged sampling strategy and the correctness of the theoretical findings.

**A. Synthetic Experiment**

We first conduct an experiment by considering a simulated task on artificially generated data, whose goal is to restore a clean matrix from partially observed entries with noises. We declare that a trial is successful if $\|L - L\|_F/\|L\|_F \leq 0.05$. Robust bilinear factorization (RBF) [57], a classical decomposition-based method for robust matrix completion problems, is utilized to solve problem (7).

The low-rank matrix $L$ is constructed by $L = X_1X_2^T$, where the entries of $X_1, X_2 \in \mathbb{R}^{n \times r}$ are independently sampled from Gaussian distribution $N(0, [1/n^2])$. In order to verify that sampling by leverage scores can bear more corruptions, we study the following two types of models.

1) **Uniform Sampling + Uniform Corruption (UU):** Each entry is sampled with equal probability $p_{ij} = p$, for all $(i, j) \in [n] \times [n]$; each observed entries are corrupted by $S$, where $P(S_{ij} = 1) = P(S_{ij} = -1) = (1/2)$ and $E(S_{ij}) = q$, for all $(i, j) \in O$.

2) **Leveraged Sampling + Uniform Corruption (LU):** Each entry is sampled with probability
\[ p_{ij} = p \frac{n^2 \sqrt{\mu_i + \nu_j}}{\sum_{i,j} \sqrt{\mu_i + \nu_j}} \]
for all $(i, j) \in [n] \times [n]$; $S$ is the same as UU.

In LU model, the probability is adaptive to the leverage scores.

We first demonstrate that exact recovery of a low-rank matrix with noise not only depends on the percentage of the entries but also how entries are observed. For UU and LU models, we set $n = 1000$, $r = 10$. Fig. 1 shows the successful frequency versus the observed percentage $p$, where we fix $q = 0.1$. For each value of $p$, we perform 50 trials.
of independent observations and error corruptions and count the number of successes. We observe that the LU model outperforms the UU model. With same ratio of successes, LU needs less observed entries; with same observed entries, LU can obtain higher ratio of success. This is because LU is based on leverage scores, which can be used to characterize the importance of each element in a matrix.

We next study the influence of the corruption percentage on the ratio of successes. We also set $n = 1000$, $r = 10$. Fig. 2 shows the successful frequency versus the corruption percentage $q$, where we fix $p = 0.4$. For each value of $q$, the ratio of successes are obtained in the same way as above. We observe that the LU model performs significantly higher robustness to noise. Specifically, in this setting, the capacity of noise immunity increases about 2.33 times, from 0.18 to 0.6. This is mainly because sampling by leverage scores can reveal “dominating” elements of a low-rank matrix.

B. Collaborative Filtering

In this section, we propose to verify the effectiveness of our proposed leverage sampling method on real-world applications—collaborative filtering; this is a technique for some recommender systems, aiming to recommend movies to its users. Formally, it predicts the unknown preference of a user on a set of unrated items according to other similar users or items.

MovieLens\(^2\) and Jester\(^3\) are two widely used datasets for recommender systems [75], [76]. For MovieLens, we select MovieLens 100K (ML-100K) and MovieLens 1M (ML-1M) in our experiments. ML-100K contains 100,000 ratings for 1682 movies by 943 users and the ratings range from 1 to 5. ML-1M contains 1 000 209 ratings for about 3900 movies by 6040 users and the ratings range from 1 to 5. Jester dataset is a joke rating dataset. It consists of three rating matrices, namely, Jester-1, Jester-2, and Jester-3. Ratings of these datasets are uniformly distributed between 1 and 5.

| Datasets | #Users | #Items | #Rated Items | Range |
|----------|--------|--------|--------------|-------|
| ML-100K  | 943    | 1,682  | 100,000      | [1,5] |
| ML-1M    | 6,040  | 3,706  | 1,000,209    | [1,5] |
| Jester-1 | 24,983 | 100    | 100,000      | [-10,10] |
| Jester-2 | 23,500 | 100    | 100,000      | [-10,10] |
| Jester-3 | 24,938 | 100    | 60,000       | [-10,10] |

Table II

Information of the Real Datasets Utilized in Our Experiments

| $\phi_1$ | $\phi_2$ | $\theta$ | RBF-uni | RBF-lev | HQASD-uni | HQASD-lev |
|----------|----------|----------|---------|---------|-----------|-----------|
| 0.9      | 0.6      | 0.8294   | 0.8109  | 0.8100  | 0.8073    |
| 0.7      | 0.8229   | 0.8033   | 0.8020  | 0.7981  |
| 0.8      | 0.8297   | 0.8174   | 0.8044  | 0.8006  |
| 0.9      | 0.8324   | 0.8263   | 0.8221  | 0.8147  |

Table III

Average RMSE of Ten Trials by Different Sampling Strategies on ML-100K Dataset Under Different Parameters Setting
also see that parameter $\theta = 0.7$ leads to the best performance, which is consistent with the empirical findings in [36].

Furthermore, empirical results also demonstrate the effectiveness and the superiority of our proposed algorithm, implying that just a small fraction of observations, even corrupted, to rate a few selected movies according to the estimated leverage scores obtained by previous samples have the potential to greatly improve the quality of the recovered preference matrix.

### VII. CONCLUSION

The incoherence condition presents a challenge in many real-world scenarios. To address this challenge, we propose a biased sampling processing method based on the row and column leverage scores of the underlying matrix. We demonstrate that an unknown $n \times n$ matrix of rank $r$ can be exactly recovered from approximately $O(nr\log^2(n))$ entries, even in cases where some entries are corrupted. Numerical experiments support our theoretical results and demonstrate the effectiveness of the biased sampling processing.

We propose a leverage score-based biased sampling strategy for matrix completion with noise. Our analysis of the sampling upper bound is rigorous. However, the sampling lower bound remains an open problem worthy of exploration. In addition, developing other methods for estimating leverages and tuning the sampling procedure would be interesting. The extension of the results and techniques presented in this article for matrix completion has potential implications for broader fields and is therefore of independent interest.

### REFERENCES

[1] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” Found. Comput. Math., vol. 9, no. 6, pp. 717–727, 2009.

[2] F. Nie, Z. Li, Z. Hu, R. Wang, and X. Li, “Robust matrix completion with column outliers,” IEEE Trans. Cybern., vol. 52, no. 11, pp. 12042–12055, Nov. 2022.

[3] X. P. Li, Z.-L. Shi, Q. Liu, and H. C. So, “Fast robust matrix completion via entry-wise $\ell_0$-norm minimization,” IEEE Trans. Cybern., early access, Dec. 6, 2022. doi: 10.1109/TCYB.2022.3234070.

[4] S. Zhang and M. Wang, “Correction of corrupted columns through fast robust hankel matrix completion,” IEEE Trans. Signal Process., vol. 67, no. 10, pp. 2580–2594, May 2019.

[5] Y. He and G. K. Atia, “Coarse to fine two-stage approach to robust tensor completion of visual data,” IEEE Trans. Cybern., early access, Sep. 5, 2022. doi: 10.1109/TCYB.2022.3108932.

[6] S. Foulcart, D. Nederl, R. Pathak, Y. Plan, and M. Wootters, “Weighted matrix completion from non-random, non-uniform sampling patterns,” IEEE Trans. Inf. Theory, vol. 67, no. 2, pp. 1264–1290, Feb. 2021.

[7] X. Li, H. Zhang, and R. Zhang, “Matrix completion via non-convex relaxation and adaptive correlation learning,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 2, pp. 1981–1991, Feb. 2023.

[8] X.-P. Li, Z.-Y. Wang, Z.-L. Shi, H. C. So, and N. D. Sidiropoulos, “Robust tensor completion via capped Frobenius norm,” IEEE Trans. Neural Netw. Learn. Syst., early access, Jan. 23, 2023. doi: 10.1109/TNNSL.2023.3236415.

[9] M. C. Tsakiris, “Low-rank matrix completion theory via plücker coordinates,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 8, pp. 10084–10099, Aug. 2023.

[10] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo, “On the applications of robust PCA in image and video processing,” Proc. IEEE, vol. 10, no. 8, pp. 1427–1457, Aug. 2018.

[11] Y. Lu, Z. Lai, X. Li, W. K. Wong, C. Yuan, and D. Zhang, “Low-rank 2-D neighborhood preserving projection for enhanced robust image representation,” IEEE Trans. Cybern., vol. 49, no. 5, pp. 1859–1872, May 2019.

[12] Z. Kang, C. Peng, and Q. Cheng, “Top-N recommender system via matrix completion,” in Proc. AAAI Conf. Artif. Intell., 2016, pp. 179–185.

[13] Y. Mao, W. Liu, and X. Lin, “Adaptive adversarial multi-task representation learning,” in Proc. Int. Conf. Mach. Learn., 2020, pp. 6724–6733.

[14] Y. Mao, S. Yun, W. Liu, and B. Du, “Tchebycheff procedure for multi-task text classification,” in Proc. Annu. Meeting Assoc. Comput. Linguist., 2020, pp. 4217–4226.

[15] Y. Lu, Z. Lai, Y. Xu, X. Li, D. Zhang, and C. Yuan, “Low-rank preserving projections,” IEEE Trans. Cybern., vol. 46, no. 8, pp. 1900–1913, Aug. 2016.

[16] L. Wang, X. Zhang, and Q. Gu, “A unified computational and statistical framework for nonconvex low-rank matrix estimation,” in Proc. Int. Conf. Artif. Intell. Statist., 2017, pp. 981–990.

[17] J. Li, J. Wang, H. Lv, Z. Zhang, and Z. Wang, “IMCHGAN: Inductive matrix completion with heterogeneous graph attention networks for drug-target interactions prediction,” IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 19, no. 2, pp. 655–665, Mar./Apr. 2021.

[18] X. Wang, Y. Wu, D. Zhuang, and L. Sun, “Low-rank hankel tensor completion for traffic speed estimation,” IEEE Trans. Intell. Transp. Syst., vol. 24, no. 5, pp. 3223–3229, 2023.

[19] Y. Liao, W. Du, P. Geurts, and G. Lodue, “DMFSGD: A decentralized matrix factorization algorithm for network distance prediction,” IEEE/ACM Trans. Netw., vol. 21, no. 5, pp. 1511–1524, Oct. 2013.
Y. Shi, J. Zhang, and K. B. Letaief, “Low-rank matrix completion,” IEEE Trans. Inf. Theory, vol. 61, no. 5, pp. 2099–2923, May 2015.

Y. Lu, C. Yuan, W. Zhu, and X. Li, “Structurally incoherent low-rank nonnegative matrix factorization for image classification,” IEEE Trans. Image Process., vol. 27, no. 11, pp. 5248–5260, Nov. 2018.

C.-P. Wei, C.-F. Chen, and Y.-C. F. Wang, “Robust face recognition with structurally incoherent low-rank matrix decomposition,” IEEE Trans. Image Process., vol. 24, no. 3, pp. 2304–3070, Aug. 2014.

Y. Chen, H. Xu, C. Caramanis, and S. Sanghavi, “Robust matrix completion and corrupted columns,” in Proc. Int. Conf. Mach. Learn., 2011, pp. 873–880.

S. Oh, A. Montanari, and A. Karbasi, “Sensor network localization from local connectivity: Performance analysis for the MDS-MAP algorithm,” in Proc. IEEE Inf. Theory Workshop Inf. Theory, 2010, pp. 1–5.

Y. Shi, J. Zhang, and K. B. Letaief, “Low-rank matrix completion for topological interference management by Riemannian pursuit,” IEEE Trans. Wireless Commun., vol. 15, no. 7, pp. 4703–4717, Jul. 2016.

D. C. Hoaglin and R. E. Welsch, “The hat matrix in regression and ANOVA,” Amer. Statist., vol. 32, no. 1, pp. 17–22, 1978.

M. W. Mahoney, “Randomized algorithms for matrices and data,” Found. Trends® @ Mach. Learn., vol. 3, no. 2, pp. 123–224, 2010.

D. Gross, “Recovering low-rank matrices from few coefficients in any basis,” IEEE Trans. Inf. Theory, vol. 57, no. 3, pp. 1548–1566, Mar. 2011.

B. Recht, “A simpler approach to matrix completion,” J. Mach. Learn. Res., vol. 12, no. 104, pp. 3413–3430, 2011.

E. J. Candés, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis: Recovery of a low-rank matrix from a corrupted observation,” Found. Trends® @ Mach. Learn., vol. 6, no. 3, pp. 1–37, 2011.

J. A. Tropp, “An introduction to matrix concentration inequalities,” Foundations Trends® @ Mach. Learn., vol. 8, nos. 1–2, pp. 1–230, 2015.

E. J. Candés and T. Tao, “The power of convex relaxation: Near-optimal matrix completion,” IEEE Trans. Inf. Theory, vol. 56, no. 5, pp. 2053–2080, May 2010.

Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward, “Completing any low-rank matrix, provably,” J. Mach. Learn. Res., vol. 16, no. 4, pp. 2999–3034, 2015.

A. Krishnamurthy and A. Singh, “Low-rank matrix and tensor completion via adaptive sampling,” in Proc. Adv. Neural Inf. Process. Syst., vol. 26, 2013, pp. 836–844.

M.-F. Balcan and H. Zhang, “Noise-tolerant life-long matrix completion via adaptive sampling,” in Proc. Adv. Neural Inf. Process. Syst., vol. 29, 2016, pp. 2955–2963.

L. Ding and Y. Chen, “Leave-one-out approach for matrix completion: Primal and dual analysis,” IEEE Trans. Inf. Theory, vol. 66, no. 11, pp. 7274–7301, Nov. 2020.

J.-F. Cai, E. J. Candés, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” SIAM J. Optim., vol. 20, no. 4, pp. 1956–1982, 2010.

Z. Lin, M. Chen, L. Wu, and Y. Ma, “The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices,” 2013. [Online]. Available: https://arxiv.org/abs/1009.5055.

Z. Lin, R. Liu, and Z. Su, “Linearized alternating direction method with adaptive penalty for low rank representation,” in Proc. Adv. Neural Inf. Process. Syst., vol. 24, 2011, pp. 612–620.

S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, “Weighted nuclear norm minimization and its applications to low level vision,” Int. J. Comput. Vis., vol. 121, no. 2, pp. 183–208, 2017.

S. Gu, L. Zhang, W. Zuo, and X. Feng, “Weighted nuclear norm minimization with application to image denoising,” in Proc. Conf. Comput. Vis. Pattern Recognit., 2014, pp. 2862–2869.

Y. Hu, D. Zhang, X. Ye, X. L. Li, and X. He, “Fast and accurate matrix completion via truncated nuclear norm regularization,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 9, pp. 2117–2120, Sep. 2013.

T.-H. Oh, Y.-W. Tai, J.-C. Bazin, H. Kim, and I. S. Kweon, “Partial sum minimization of singular values in robust PCA: Algorithm and applications,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 4, pp. 754–768, Apr. 2016.

F. Nie, H. Huang, and C. Ding, “Low-rank matrix recovery via efficient Schatten p-norm minimization,” in Proc. 26th AAAI Conf. Artif. Intell., 2012, pp. 655–661.
Xinjian Huang received the Ph.D. degree from the School of Computer Science, Wuhan University, Wuhan, China, in 2021. He is currently a Lecturer with the School of Cyber Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. His research interests include machine learning, computer vision, and the security of artificial intelligence.

Weiwei Liu (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of Technology Sydney, Sydney, NSW, Australia, in 2017, under the supervision of Prof. Ivor W. Tsang. He is currently a Full Professor with the School of Computer Science, Wuhan University, Wuhan, China. His research results have been published in prestigious journals and leading conferences, such as Journal of Machine Learning Research, IEEE Transactions on Pattern Analysis and Machine Intelligence, NeurIPS, ICML, and COLT. His current research interest is statistical machine learning.

Bo Du (Senior Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2010. He is currently a Professor with the School of Computer Science and the Institute of Artificial Intelligence, Wuhan University, where he is also the Director of the National Engineering Research Center for Multimedia Software. He has more than 80 research articles published in the IEEE Transactions on Image Processing, IEEE Transactions on Cybernetics, IEEE Transactions on Pattern Analysis and Machine Intelligence, and IEEE Transactions on Geoscience and Remote Sensing. Fourteen of them are ESI hot articles or highly cited articles. His major research interests include machine learning, computer vision, and image processing.

Dr. Du serves as an Associate Editor for Neural Networks, Pattern Recognition, and Neurocomputing. He also serves as a reviewer for 20 Science Citation Index magazines, including IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Cybernetics, IEEE Transactions on Geoscience and Remote Sensing, IEEE Transactions on Image Processing, and IEEE Geoscience and Remote Sensing Letters.

Dacheng Tao (Fellow, IEEE) received the Ph.D. degree from the University of London, London, U.K., in 2007. He is an Australian Laureate Fellow, a Professor of Computer Science, and a Peter Nicol Russell Chair with the School of Computer Science; an Advisor and Chief Scientist of the Digital Sciences Initiative; and the Founding Director of the Sydney AI Centre, Faculty of Engineering, The University of Sydney, Darlington, NSW, Australia. His research is detailed in one monograph and over 200 publications in prestigious journals and proceedings at prominent conferences, such as IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Image Processing, IEEE Transactions on Neural Networks and Learning Systems, International Journal of Computer Vision, Journal of Machine Learning Research, NeurIPS, ICML, CVPR, ICCV, ECCV, AAAI, IJCAI, ICDM, and ACM SIGKDD, with several best paper awards.

Dr. Tao received the 2015 and 2020 Australian Eureka Prize, the 2018 IEEE ICDM Research Contributions Award, and the 2021 IEEE Computer Society McCluskey Technical Achievement Award. He is a Fellow of the Australian Academy of Science, the Royal Society of NSW, TWAS, AAAS, and ACM.