Fast exact recovery of noisy matrix from few entries: the infinity norm approach

BaoLinh Tran and Van Vu
Department of Mathematics
Yale University
New Haven, CT 06511
l.tran@yale.edu and van.vu@yale.edu

Abstract

The matrix recovery (completion) problem, a central problem in data science, involves recovering a matrix A from a relatively small random set of entries. While such a task is generally impossible, it has been shown that one can recover A exactly in polynomial time, with high probability, under three basic and necessary assumptions: (1) the rank of A is very small compared to its dimensions (low rank), (2) A has delocalized singular vectors (incoherence), and (3) the sample size is sufficiently large. Various algorithms address this task, including convex optimization by Candes, Recht, and Tao (2009), alternating projection by Hardt and Wooters (2014), and low-rank approximation with gradient descent by Keshavan, Montanari, and Oh (2009, 2010). In applications, Candes and Plan (2009) noted that it is more realistic to assume noisy observations. In such cases, the above approaches provide approximate recovery with small root mean square error, which is difficult to convert into exact recovery. Recently, results by Abbe et al. (2017) and Bhardwaj et al. (2023) on approximation in the infinity norm showed that one can recover A even in the noisy case, provided A has bounded precision. However, beyond the three basic assumptions, they either required that the condition number of A be small (Abbe and Fan, 2017) or that the gaps between consecutive singular values be large (Bhardwaj et al., 2023). These additional assumptions conflict, with one requiring singular values to be close together and the other suggesting they should be far apart. It is thus natural to conjecture that neither is necessary. In this paper, we demonstrate that this is indeed the case. We propose a simple algorithm for exact recovery of noisy data, relying solely on the three basic assumptions. The core step of the algorithm is a straightforward truncated singular value decomposition, which is highly efficient. To analyze the algorithm, we prove a new infinity norm version of the classical Davis-Kahan perturbation theorem, improving an earlier result in (Bhardwaj et al., 2023). Our proof employs a combinatorial contour integration argument and is entirely distinct from all previous approaches.

1 Introduction

1.1 Problem description

A large matrix $A \in \mathbb{R}^{m\times n}$ is hidden, except for a few revealed entries in a set $\Omega \subset [m] \times [n]$. We call Ω the set of *observations* or *samples*. The matrix A_Ω, defined by

$$(A_\Omega)_{ij} = A_{ij} \text{ for } (i,j) \in \Omega, \text{ and 0 otherwise,} \tag{1}$$

is called the *observed* or *sample* matrix. The task is to recover A, given A_Ω. This is the *matrix recovery* (or *matrix completion*) problem, a central problem in data science that has received significant

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

attention in recent years, motivated by a number of real-world applications. Examples include building recommendation systems, notably the **Netflix challenge** [1]; reconstructing a low-dimensional surface based on partial distance measurements from a sparse network of sensors [2, 3]; repairing missing pixels in images [4]; and system identification in control [5]. See the surveys by Li et al. [6] and Davenport and Romberg [7] for additional applications.

Researchers have proposed two models: (a) Ω is sampled uniformly among subsets with the same size, or (b) Ω has independently chosen entries, each with the same probability p, called the *sampling density*, which can be known or hidden. The models are interchangeable in mathematical analysis through a simple conditioning trick [4]. Most papers use (b), and we will do so in our paper. Very recently, researchers have also explored models where the sample entries are correlated [8–11]; these are beyond the scope of this paper.

In this paper, we focus on exact recovery (to find all entries exactly). For this problem to make sense, it is important to assume that A has a *fixed precision*, namely all entries of A are integer multiples of a positive constant ε. (Otherwise, it is impossible even to write down an entry exactly, let alone compute it.) In most practical contexts, ε does not depend on the size of the matrix. For instance, in the Netflix problem, all entries are half-integers, so $\varepsilon = 1/2$. If all entries have two decimal places, then $\varepsilon = 0.01$.

It has been pointed out by Candes and Plan [12] that in practice, data is often perturbed by noise, so we can only observe a partially hidden and noisy version of A. The main goal of this paper is to find an exact recovery for A from such a noisy sample.

1.2 Basic notation and assumptions

Throughout this paper, A is an $m \times n$ matrix, with $N := \max\{m, n\}$. Consider the SVD of $A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$, where $r := \operatorname{rank} A$, and the singular values are ordered: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. We write $A_s = \sum_{i=1}^{s} \sigma_i u_i v_i^T$ for the best rank-s approximation of A. An important parameter that appears in many papers in this area is the *condition number* of A: $\kappa = \kappa(A) := \sigma_1/\sigma_r$.

The *coherence parameter* is given by $\mu_0 = \mu_0(A) = \max\{\mu(U), \mu(V)\}$, where

$$\mu(U) := \max_{i \in [m]} \frac{m}{r} \|e_i^T U\|^2 = \frac{m\|U\|_{2,\infty}^2}{r}, \tag{2}$$

and analogously for $\mu(V)$. The 2-to-∞ norm of a matrix M is given by $\|M\|_{2,\infty} := \sup\{\|Mu\|_\infty : \|u\|_2 = 1\}$. It should be noted that $\|M\|_{2,\infty}$ is simply the largest L2 norm among the rows of M.

We use C to denote a positive constant, whose value depends on the context. When C depends on a set of parameters $a_1, a_2, \ldots, a_k$, we write $C(a_1, a_2, \ldots, a_k)$.

In most existing works on matrix recovery, researchers make the following three assumptions

- *Low-rank:* One assumes that $r := \operatorname{rank} A$ is much smaller than $\min\{m, n\}$. Many papers assume r is bounded ($r = O(1)$), while $m, n \to \infty$.
- *Incoherence:* One requires that the rows and columns of A are sufficiently "spread out", so the information does not concentrate in a small set of entries, which could be easily overlooked by random sampling. In technical terms, one needs μ_0 to be small.
- *Sufficient sampling size/density:* The sampling density (or the size of Ω) is sufficiently large. To ensure all rows and columns are sampled, one needs $|\Omega| \geq CN \log N$. We will typically work in the regime $|\Omega| = N \log^{O(1)} N$, which is optimal up to a logarithmic term.

These assumptions have been shown to be necessary; see [13, 4, 7], and have been used in most papers on exact recovery. We will refer to them as the **basic assumptions**.

In the noisy setting, we observe entries from $A + Z$, a noisy version of A, where Z represents the noise matrix. Most studies assume that Z has independent entries with mean 0, but not necessarily from the same distributions. In many papers directly related to our work, it was assumed that Z has bounded entries with probability 1. We will do the same here but note that we can replace this with weaker conditions, such as the entries of Z having light tails (via a standard truncation trick). We let $A_{\Omega,Z} := (A + Z)_\Omega$ denote the noisy observed matrix.

1.3 A brief summary of existing results and our contributions

There is a vast literature on matrix completion. The problem of exact recovery in the pure (noiseless) case, under the three basic assumptions, was first solved by Candès and Recht [13], Candès and Tao [4], Recht [14], using *nuclear norm minimization* (**NNM**), which seeks to find the matrix with the smallest nuclear norm that satisfies the observed samples via *semidefinite programming* (**SDP**). Their idea is based on convexifying the intuitive but NP-hard approach of minimizing the rank given the observations. However, the best-known solvers for the SDP run in time $O(|\Omega|^2 N^2)$, which is $O(N^4 \log^2 N)$ in the best case [6], and the calculation may be sensitive to noise [4]. Therefore, it is reasonable to seek faster and more noise-robust algorithms, potentially sacrificing some generality.

In Hardt and Wootters [15], Hardt [16], the authors offered another intuitive but NP-hard approach: minimize $\|X_\Omega - A_\Omega\|_F$ subject to $\operatorname{rank} A = r$, and proposed an approximate solution with *alternating projections*, which switches between optimizing the column and row spaces, given the other. In Keshavan et al. [17, 18], the authors proposed taking a low-rank approximation of A_Ω via SVD, then using it as an initial value for a GD-based algorithm to approximate A. In both approaches, it is important that the condition number κ be small. See [19, Section A] in the supplementary appendix for a more detailed discussion of the role of κ in these results. See also [20–23].

Recovery with noisy data, while more practical, as pointed out by Candes and Plan [12] in their influential survey, is clearly a harder problem. The authors of the methods discussed above all extended their studies to this case, achieving approximate solutions with guarantees in the Frobenius norm. However, it seems very difficult to turn these approximations into exact recovery.

Recently, there has been progress in achieving *exact recovery* in the noisy case by Abbe and Fan [24] and Bhardwaj et al. [25], with simple and fast algorithms. The basic idea is to show that a properly chosen low-rank approximation of the observed (noisy) matrix, under an appropriate assumption, approximates the ground truth matrix A in the infinity norm, with an error less than $\varepsilon/2$, where ε is the discretization unit of the entries of A. Once this is achieved, a simple rounding off recovers A exactly (see the last few paragraphs of Subsection 1.1). The heart of the matter is the analysis of the algorithms, which requires new mathematical ideas, as proving approximation in the infinity norm is significantly more challenging than a similar task under the Frobenius or spectral norm.

In what follows, we denote by $A_{\Omega,Z}$ the matrix obtained from the observed entries of $A + Z$. This is the input of the recovery algorithm. Beyond the three basic assumptions (see Subsection 1.2), these new works require an extra spectral assumption that either the condition number is small [24] or the gaps between consecutive singular values are large [25].

These assumptions are strong, and it is unclear how often they hold in practice. Figure 1.3 below is based on the Yale face database, a well-known and frequently used dataset; see Wainwright's book [26]. The data matrix in $\mathbb{R}^{165\times 77760}$ is created by flattening 165 greyscale 243×320 facial images into row vectors in $\mathbb{R}^{77760}$ and centering them by subtracting the average row. The singular values decay quickly, resulting in a rather large condition number; $\kappa \approx 10.45$ for the first 30 singular values. Consequently, the gaps between singular values are generally large, but there are still exceptions (for example, at index 15).

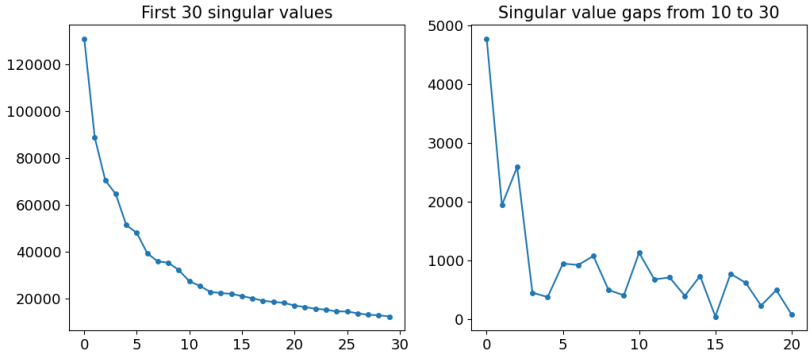

Figure 1: The Yale face database spectrum and spectral gaps

From a mathematical viewpoint, we observe a curious phenomenon: these two extra assumptions are seemingly contradictory. The first (small condition number) indicates that the singular values should be close to each other, while the second requires them to be apart. It is thus natural to conjecture that neither of these conditions is necessary.

In this paper, we prove this conjecture. We show that a properly chosen low-rank approximation of $A_{\Omega,Z}$ approximates A (with the required precision) in the infinity norm, using only the three basic assumptions. This not only provides a mathematically satisfying answer but also significantly expands the range of applications (regarding real datasets).

Our analysis of the algorithm is based on novel mathematical developments and is entirely distinct from previous approaches. It combines the contour integral method introduced in Tran and Vu [27] with a novel bound on random matrix powers, in a fairly nontrivial manner. This approach is robust, and we believe in its potential for future applications.

2 New results: a unifying exact recovery method

2.1 Formal setting and algorithm

To state the algorithm and our main result, let us restate the setting for clarity.

Setting 2.1 (Matrix completion with noise)**.** Consider the gound truth matrix A, the observed set Ω, and noise matrix Z. We assume: (1) $\|A\|_\infty \le K_A$ for a known parameter K_A; (2) we know an upper bound $r_{\max} \ge r$, without needing to know r; and (3) the noise Z has independent entries satisfying $\mathbf{E}[Z_{ij}] = 0$ and $\mathbf{E}\left[|Z_{ij}|^l\right] \le K_Z^l$ for all $l \in \mathbb{N}$ for a known parameter K_Z, without necessarily having the same distribution or variance. The parameters r, $r_{\max}$, K_A, K_Z can depend on m and n.

Algorithm 2.2 (Approximate-and-Round 2)**.** Input: the $m \times n$ sample matrix $A_{\Omega,Z}$ and the discretization unit ε of A's entries.

1. *Empirical rescaling:* Let $\hat{p} := (mn)^{-1}|\Omega|$ and $\hat{A} := \hat{p}^{-1} A_{\Omega,Z}$.
2. *Low-rank approximation:* Compute the truncated SVD $\hat{A}_{r_{\max}} = \sum_{i \le r_{\max}} \hat{\sigma}_i \hat{u}_i \hat{v}_i^T$.

   Take the largest index $s \le r_{\max} - 1$ such that $\hat{\sigma}_s - \hat{\sigma}_{s+1} \ge 20(K_A + K_Z)\sqrt{\frac{r_{\max}(m+n)}{\hat{p}}}$.

   If no such s exists, take $s = r_{\max}$. Let $\hat{A}_s := \sum_{i \le s} \hat{\sigma}_i \hat{u}_i \hat{v}_i^T$,
3. *Rounding off:* Round each entry of $\hat{A}_s$ to the nearest multiple of ε. Return $\hat{A}_s$.

Compared to the algorithm Approximate-and-Round used in [25], a minor difference is that we use an estimate $\hat{p}$ of p, which is highly accurate with high probability. Another innovation is a different threshold for the truncated SVD step that does not require knowledge of the parameter μ_0. From a complexity viewpoint, our algorithm is efficient, consisting of a truncated SVD on a matrix with $|\Omega|$ non-zero entries, and a rounding step, requiring only $O(|\Omega| r + mn) = O((pr+1)mn)$ FLOPs.

Our main theorem below provides sufficient conditions for exact recovery with this algorithm.

Theorem 2.3. *There is a universal constant $C > 0$ such that the following holds. Suppose $r_{\max} \le \log^2 N$. Under the model 2.1, assume that* the signal is sufficiently large*,*

$$\|A\| = \sigma_1 \ge 100 r K \sqrt{\frac{r_{\max} N}{p}},$$

for $K := K_A + K_Z$; and the sampling is sufficiently dense*,*

$$p \ge C\left(\frac{1}{m} + \frac{1}{n}\right) \max\left\{\log^4 N,\ \frac{r^3 K^2}{\varepsilon^2}\left(1 + \frac{\mu_0^2}{\log^2 N}\right)\right\} \log^6 N. \tag{3}$$

Then with probability $1 - O(N^{-1})$, the low-rank approximation step of Approximate-and-Round 2 recovers every entry of A within an absolute error $\varepsilon/3$. Consequently, if all entries are multiples integer of ε, the rounding-off step recovers A exactly.

As promised, our result unifies [24] and [25], removing both the condition number and the gap assumptions. Furthermore, our result also implies a RMSE recovery with the same error margin with no constraint on the condition number, an improvement over [17, 18, 15].

2.2 Analysis of the result

We begin with a small experiment to showcase the performance of the algorith. Each of the 13 datapoints corresponds to a pair (m, n) that is either $(2^l, 2^{l-1})$ or $(2^l, 2^l)$ for $l = 10, 11, \ldots, 16$.

The matrix $A \in \mathbb{R}^{m\times n}$ is randomly generated by $A_0 = XY^T$, where $X \in \mathbb{R}^{m\times r}$ and $Y \in \mathbb{R}^{n\times r}$, with iid $N(0, 1)$ entries, then $A := \frac{3}{4}A_0/\|A_0\|_\infty$. The noise matrix Z is generated with $Z_0 \in \mathbb{R}^{m\times n}$ having iid $N(0, 1/4)$ entries, then set $Z_{ij} = \operatorname{sgn}((Z_0)_{ij}) \max\{1/4, |(Z_0)_{ij}|\}$. This normalization makes $K_A + K_Z = 3/4 + 1/4 = 1$. For the sampling, we fix $p = 0.1$ in all datapoints.

In Figure 2, the plots show $N = m + n$ on the horizontal axis, on a log scale, and the RMSE and infinity norm errors on the vertical axes. Both decline rapidly with N. In the last datapoint, where $m = n = 2^{16} \approx 6000$, one can recover every entry of A to within an absolute error of 0.1.

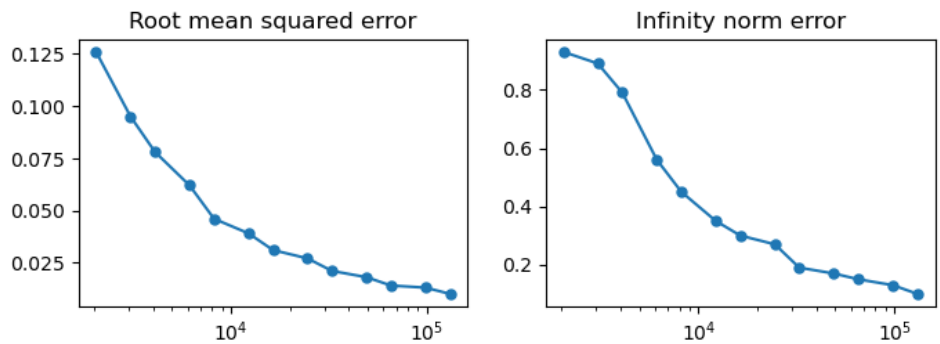

Figure 2: The RMSE and Infinity norm error of our method

Next, let us discuss the optimality of our result from a theoretical view.

Remark 2.4 (A high-level explanation for the thresholding rule)**.** Note that the observed matrix, rescaled by p^{-1}, is a perturbed version of the original matrix, since

$$\mathbf{E}\left[p^{-1}A_{\Omega,Z}\right] = p^{-1}(\mathbf{E}\left[A_\Omega\right] + \mathbf{E}\left[Z\right]) = A.$$

Therefore, one can write $p^{-1}A_{\Omega,Z} = \tilde{A} = A + E$, where E is a random matrix of mean 0 and independent entries. One can view E as a type of "noise" that includes both Z and the noise caused by the random sampling. Since it is a well-known fact in random matrix theory that $\|Z\| = O(K_Z\sqrt{N/p})$ (see [28, 29] for proofs), we simply want to cut off the singular values of $\tilde{A}$ at a level above $\|E\|$, as the well-documented BBP phenomenon [30–35] shows that the part of A below that is fully absorbed by E and becomes indistinguishable from noise. Under the assumptions we make, the signal extracted from the spectrum of $\tilde{A}$ above this threshold is a good approximation of A. We use the same argument in Remark 2.5 to show the necessity of the lower bound on σ_1.

In our theorem, we use only the three basic assumptions of Section 1.2, plus that $\|A\| = \sigma_1$ is large enough. This last assumption seems new, but it is not. It is often hidden in previous papers, and is guaranteed by the fact that A has low rank. Indeed

$$\sigma_1^2 = \|A\|^2 \geq \frac{1}{r}\|A\|_F^2.$$

Consider the representative case when $\|A\|_F^2 = \Theta(mn)$. Then $\sigma_1 \geq r^{-1/2}\|A\|_F \geq c\sqrt{\frac{mn}{r}}$. By the density condition (3) and the condition $r_{\max} \leq \log^2 N$, we have

$$p \geq \frac{r^3 r_{\max} K^2 N \log^4 N}{mn} \implies \sigma_1 \geq c\sqrt{\frac{mn}{r}} \gg (\log^2 N) rK\sqrt{\frac{r_{\max}N}{p}},$$

so the assumption on σ_1 in our theorem is automatically satisfied.

Remark 2.5 (Necessity of the signal assumption)**.** From the engineering perspective, it is intuitive that σ_1 should be large enough in order for any kind of recovery to be possible. In the opposite case when the intensity of the noise Z dominates the signal A, then the data is "too corrupted" and all interersting information are lost. Rigorously, the well-documented BBP phenomenon in random matrices [30–35] states that, if $\|Z\| \geq c\|A\| = c\sigma_1$, for a specific constant c, then $A + Z$ is indistinguishable from a fully random matrix, leaving no chance to recover A even if $A + Z$ is fully observed. Since it is also a well-known fact in random matrix theory that $\|Z\| = O(K_Z\sqrt{N/p})$ (see [28, 29] for proofs), the condition in Theorem 2.3 is simply $\sigma_1 \geq Cr\|Z\|$, which is optimal when $r = O(1)$.

Remark 2.6 (Optimality of the density bound)**.** The condition (3) looks complicated, but in the base case where $r = O(1)$, $K_A + K_Z = O(1)$, and uniformly random singular vectors, so that we have $\mu_0 = O(\log N)$, it reduces to

$$p \geq C \max\left\{\log^4 N,\ \varepsilon^{-2}\right\}\left(m^{-1} + n^{-1}\right)\log^6 N, \tag{4}$$

which is equivalent to $|\Omega| \geq CN\log^6 N \max\{\log^4 N, \varepsilon^{-2}\}$ in the uniform sampling model. The power of N is optimal to the theoretical limit (see Section 1.2). The power of $\log N$ can be further reduced but the details are tedious, and the improvement is not really important from the pratical view point. For recovery with precision ε, this bound grows with ε^{-2}, which is comparable to or better than all previous works (see [19, Section A]).

Remark 2.7 (Relaxing the bound on $r_{\max}$)**.** The condition $r_{\max} \leq \log^2 N$ in Theorem 2.3 can be avoided, at the cost having a more complicated sampling density bound, which connects p and r

$$p \geq C\left(\frac{1}{m} + \frac{1}{n}\right)\max\left\{\log^{10} N,\ \frac{r^4 r_{\max}\mu_0^2 K^2}{\varepsilon^2}, \frac{r^3 K^2}{\varepsilon^2}\left(1 + \frac{\mu_0^2}{\log^2 N}\right)\left(1 + \frac{r^3\log N}{N}\right)\log^6 N\right\}. \tag{5}$$

The proofs of Theorem 2.3 and Eq. (5) will be in the supplementary appendix, [19, Section B] . This shows that our result does not require any extra condition besides the mandatory large signal assumption.

In Section 3, we give a proof sketch for Theorem 2.3, asserting the correctness of our algorithm, Approximate-and-Round 2. The proof will boil down to obtaining a sharp bound in the infinity norm for the perturbation of the low-rank approximations, for which we introduce Theorem 7. By reframing the problem from a matrix perturbation perspective, we can view Theorem 7 as an infinity norm version of the classical Davis-Kahan-Wedin theorem.

We next explain the main ideas behind its proof, which combine the contour integral technique of Tran and Vu [27] (with necessary adjustments) and a novel *semi-isotropic* bound on powers of a random matrix. The (rather complex) detailed proofs will appear in the supplementary appendix, [19, Section C] .

3 Main ideas of the proof

3.1 The matrix perturbation perspective and the rise of the extra assumptions

Consider $\tilde{A}_s - A_s$. Let $\rho := \hat{p}/p$ and $\tilde{A} = p^{-1}A_{\Omega,Z} = \rho^{-1}\hat{A}$, we can write

$$\hat{A}_s - A = \rho^{-1}\tilde{A}_s - A = (\rho^{-1} - 1)A + \rho^{-1}(A_s - A) + \rho^{-1}(\tilde{A}_s - A_s).$$

We show that the three error terms on the right-hand side are small in the infinity norm. The first is easy, since ρ is close to 1 (by a Chernoff bound, see e.g. Hoeffding [36]). For the second, we have

$$\|A - A_s\|_\infty = \Big\|\sum_{i\geq s+1}\sigma_i u_i v_i^T\Big\|_\infty \leq \sigma_{s+1}\|U\|_{2,\infty}\|V\|_{2,\infty} \leq \frac{r\sigma_{s+1}\mu_0}{\sqrt{mn}},$$

which will be small by the way we choose s and the incoherence property. Most of the heavy lifting goes to bounding the third term, $\|\tilde{A}_s - A_s\|_\infty$.

Observe that $\mathbf{E}[A_\Omega] = pA$ and $\mathbf{E}[Z_\Omega] = 0$, so that $\mathbf{E}[\tilde{A}] = p^{-1}\mathbf{E}[A_\Omega + Z_\Omega] = A$. Therefore, $E := \tilde{A} - A$ is a random matrix with mean 0. This opens a way to use tools and ideas from random matrix theory in the analysis.

The approaches in [24, 25] both arrive at a bound on $\|\tilde{A}_s - A_s\|_\infty$, but at the cost of their respective extra assumptions. We give a brief overview of their methods.

Suppose $s = r = 1$ for simplicity, one can write

$$\begin{aligned}\tilde{A}_1 - A_1 &= \tilde{\sigma}_1 \tilde{u}_1 \tilde{v}_1^T - \sigma_1 u_1 v_1^T = (\tilde{\sigma}_1 - \sigma_1) u_1 v_1^T + \sigma_1 \big(\tilde{u}_1^T \tilde{v}_1^T - u_1 v_1^T\big) \\ &= (\Delta\sigma_1) u_1 v_1^T + \sigma_1 (\Delta u_1) v_1^T + \sigma_1 u_1 (\Delta v_1)^T + \sigma_1 (\Delta u_1)(\Delta v_1)^T,\end{aligned}$$

where we set $\Delta\sigma_1 := \tilde{\sigma}_1 - \sigma_1$ and analogously for other Δ-notations.

By Weyl's inequality, $|\Delta\sigma_1| \le \|E\|$, so the first term is bounded by $\|u_1\|_\infty \|v_1\|_\infty \|E\|$ in the infinity norm. The main challenge is to bound the middle two terms, which dominate the last one. By symmetry, let us focus on $\sigma_1(\Delta u_1) v_1^T$. We have

$$\|\sigma_1(\Delta u_1) v_1^T\|_\infty = \sigma_1 \|\Delta u_1\|_\infty \|v_1\|_\infty.$$

If σ_1 is large enough compared to $\|E\|$, [25] showed that the error Δu_1 is sufficiently "spread out", namely, $\|\Delta u_1\|_\infty \le C\|\Delta u_1\| \|u_1\|_\infty$. To bound $\|\Delta u_1\|$, the classic **Davis-Kahan-Wedin theorem** [37, 38] gives $\|\Delta u_1\| \le C\|E\|/\sigma_1$, so the final bound on this term looks like

$$\|\sigma_1(\Delta u_1) v_1^T\|_\infty \le C \|u_1\|_\infty \|v_1\|_\infty \|E\|.$$

Putting everything together, one obtains the desired (optimal) bound

$$\|\tilde{A}_1 - A_1\|_\infty \le C\|u_1\|_\infty \|v_1\|_\infty \|E\| \le \frac{C\mu_0}{\sqrt{mn}} \|E\|. \tag{6}$$

When extending this argument to rank $s > 1$, [24] and [25] used two different approaches. In [24], the authors wrote

$$\tilde{A}_s - A_s = U_s(\Delta\Sigma_s)V_s^T + (\Delta U_s)\Sigma_s V_s^T + U_s \Sigma_s (\Delta V_s)^T + (\Delta U_s)\Sigma_s(\Delta V_s)^T.$$

Again, one needs to bound $\|(\Delta U_s)\Sigma_s V_s^T\|_\infty$. The problem here is that $\|\Sigma_s\|$ is still σ_1 (instead of σ_s), which eventually leads to the appearnce of the condition number $\kappa = \sigma_1/\sigma_r$ in the final bound. Furthermore, the algorithm in [24] requires knowledge of the rank r of A. This is often not the case in practice. However, Keshavan et al. [17] showed that one can estimate r precisely with high probability, given that $\kappa = O(1)$. Thus, the assumption that the condition number is small is needed here for two different reasons: to compute the rank and to improve the quality of the bound; see also [19, Section A].

[25] circumvented this issue by considering

$$\tilde{A}_s - A_s = \sum_{i \le s} (\tilde{\sigma}_i \tilde{u}_i \tilde{v}_i^T - \sigma_i u_i v_i^T) = \sum_{i \le s} \Delta(\sigma_i u_i v_i^T),$$

then bounding each term $\Delta(\sigma_i u_i v_i^T)$ separately. Again, this boils down to bounding $\|\sigma_i(\Delta u_i) v_i^T\|_\infty$. For this, they use a (stronger) variant of Davis-Kahan theorem proved in O'Rourke et al. [39] (see [19, Section B]). As a trade-off, this approach requries the assumption that the singular values σ_i are well-separated for the (stronger) Davis-Kahan bound to hold.

Our new method, which is entirely different, allows us to avoid both assumptions. Let us first establish the quantitative statement:

Theorem 3.1. *Consider a deterministic matrix $A \in \mathbb{R}^{m\times n}$. and a random matrix $E \in \mathbb{R}^{m\times n}$ with independent entries satisfying $\mathbf{E}[E_{ij}] = 0$ and $\mathbf{E}\left[|E_{ij}|^l\right] \le p^{1-l}K^l$ for some $K > 0$ and $0 < p < 1$. Let $\tilde{A} = A + E$. Let $s \in [r]$ be an index satisfying*

$$\delta_s := \sigma_s - \sigma_{s+1} \ge 40rK\sqrt{N/p},$$

There are constant C, C_1 such that, if $p \ge C(m^{-1} + n^{-1})\log N$ (where $N = \max\{m, n\}$), then

$$\|\tilde{A}_s - A_s\|_\infty \le \frac{C_1(\mu_0 + \log N)\log^2 N}{\sqrt{mn}} r\sigma_s \left(\frac{K\sqrt{N}}{\sigma_s\sqrt{p}} + \frac{rK\sqrt{\log N}}{\delta_s\sqrt{p}} + \frac{r^2\mu_0 K \log N}{p\delta_s\sqrt{mn}} \right). \tag{7}$$

From Eq. (7), one can verify, with a routine computation, that the sampling density condition (3), with a sufficiently large constant C, implies $\|\tilde{A}_s - A_s\|_\infty \le \varepsilon/3$, which proves Theorem 2.3. For a detailed proof of Theorem 2.3, see [19, Section B].

3.2 Our new approach: contour integrals and bounds on random matrix powers

Let us now elaborate on our new approach used to prove Theorem 3.1.

We begin with the contour integral argument, adopted from the technique by P. Tran and Vu [27]. Instead of analyzing $\tilde{A}_s - A_s$ directly, we turn to their *symmetrized versions*. Denote

$$B_{\text{sym}} = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix}$$

for any matrix B, we have $\tilde{A}_{\text{sym}} = A_{\text{sym}} + E_{\text{sym}}$. Note that the symmetrization preserves the spectral norm and transforms the singular value decomposition into the eigendecomposition

$$A_{\text{sym}} = W\Lambda W^T = \sum_{|i|\in[r]} \sigma_i w_i w_i^T,$$

where for each $i \in [r]$,

$$w_i = \frac{1}{\sqrt{2}} \begin{bmatrix} u_i \\ v_i \end{bmatrix}, \qquad w_{-i} = \frac{1}{\sqrt{2}} \begin{bmatrix} u_i \\ -v_i \end{bmatrix}, \quad \sigma_{-i} = -\sigma_i.$$

The first key idea here is to compare $\tilde{A}_{\text{sym}}$ and A_{sym} via their *Stieltjes transforms*. Let z be a complex variable, we have the expansion (we pretend that convergence is not an issue for now)

$$z(zI - \tilde{A}_{\text{sym}})^{-1} - z(zI - A_{\text{sym}})^{-1} = \sum_{\gamma=1}^{\infty} z \left[(zI - A_{\text{sym}})^{-1} E_{\text{sym}}\right]^{\gamma} (zI - A_{\text{sym}})^{-1}.$$

If we integrate both sides over a contour Γ_s which encloses only the eigenvalues $\{\sigma_i : |i| \in [s]\}$, we obtain, by Cauchy's integration theorem (and some light calculation),

$$(\tilde{A}_s - A_s)_{\text{sym}} = \sum_{\gamma=1}^{\infty} \oint_{\Gamma_s} \frac{z\mathrm{d}z}{2\pi i} \left[(zI - A_{\text{sym}})^{-1} E_{\text{sym}}\right]^{\gamma} (zI - A_{\text{sym}})^{-1}.$$

Using the formula $(zI - A_{\text{sym}})^{-1} = \sum_{|i|\in[r]} (z(z-\sigma_i))^{-1} w_i w_i^T + z^{-1} I$, we can expand the right-hand side (via some calculations) to obtain

$$(\tilde{A}_s - A_s)_{\text{sym}} = \sum_{\gamma=1}^{\infty} \sum_{*,*,\ldots,*} \mathcal{C}(*,*,\ldots,*) E_{\text{sym}}^* w_* w_*^T E_{\text{sym}}^* w_* w_*^T \ldots w_* w_*^T E_{\text{sym}}^* w_* w_*^T E_{\text{sym}}^*,$$

where the asterisks stand for integer variables, i.e. w_* can be any w_i, E_{sym}^* can be any positive power of E_{sym}, and $\mathcal{C}$ is a scalar coefficient of the form

$$\mathcal{C}(*,*,\ldots,*) = \oint_{\Gamma_s} \frac{z\mathrm{d}z}{2\pi i} \frac{1}{z^*(z-\sigma_*)(z-\sigma_*)\ldots(z-\sigma_*)}.$$

There are conditions on these variables, but we will overlook them for now. At this point, one can bound $|\tilde{A}_s - A_s|$ by bounding the terms above and applying the triangle inequality. These bounds are fairly technical and require some delicate combinatorial and analytical consideration. This is essentially the approach introduced in [40].

Since we aim for the infinity norm, which is much harder to deal with than the spectral norm, we need to significantly modify the above argument.

Let us consider, for example, the 11-(upper left corner) entry of the matrix in question. We have

$$(\tilde{A}_s - A_s)_{11} = \sum_{\gamma=1}^{\infty} \sum_{*,\ldots,*} \mathcal{C}(*,*,\ldots,*) \big(e_{m+1}^T E_{\text{sym}}^* w_*\big) \big(w_*^T E_{\text{sym}}^* w_* w_*^T \ldots w_*^T E_{\text{sym}}^* w_*\big) \big(w_*^T E_{\text{sym}}^* e_1\big),$$

where $e_1, e_2, \ldots$ are standard basis vectors in $\mathbb{R}^N$. The upper left entry of a matrix M is $e_1^T M e_1$, but notice that we use e_{m+1} on the left due to the symmetrization.

To bound $\|\tilde{A}_s - A_s\|_\infty$, we need a strong bound on $e_j^T E_{\text{sym}}^a w_i$, for all indices j. To this end, we prove a new *semi-isotropic bound* for powers of E_{sym}, which is a key technical contribution of this paper.

These bounds take advantage of the fact that E has independent entries. The exact bounds are a bit technical, but for the simple case when $\mu_0 = O(1)$, they take the form

$$|e_j^T E_{\text{sym}}^a w_i| \leq \begin{cases} C \log N \sqrt{\frac{\mu_0 + \log N}{m}} \|E\|^a & \text{for } 1 \leq j \leq m, \\ C \log N \sqrt{\frac{\mu_0 + \log N}{n}} \|E\|^a & \text{for } m+1 \leq j \leq N. \end{cases}$$

By setting $a = 0$, one can also see that these bounds are optimal, up to $\log N$ factors. Plugging these bounds into the expansion, one can eventually obtain the bound of Theorem 3.1, by modifying the steps from [27] in a proper manner. This is, in itself, a challenging task, and we provide the full detailed proof in the supplementary appendix [19, Section C] .

Note that there have been similar isotropic-style bounds for powers of random matrices, notably by Mao et al. [41] and Fan et al. [42]. They both cover a wide range of cases, but were designed optimally for the use cases in their respective papers. In the supplementary document [19, Section B], we will briefly explain that they are not strong enough in our use case.

Besides the bound on the perturbation of low-rank approximations, we obtain similar bounds for the perturbation of singular vectors, in both the infinity and 2-to-∞ norms. For instance, we can show

$$\|\tilde{U}_s \tilde{U}_s^T - U_s U_s^T\|_\infty \leq \frac{C \log^2 N(\mu_0 + \log N)}{m} \left(\frac{\|E\|}{\sigma_s} + \frac{r\|U^T E V\|}{\delta_s} \right)$$

$$\|\tilde{U}_s \tilde{U}_s^T - U_s U_s^T\|_{2,\infty} \leq C \log N \sqrt{\frac{(\mu_0 + \log N)}{m}} \left(\frac{\|E\|}{\sigma_s} + \frac{r\|U^T E V\|}{\delta_s} \right)$$

under the assumptions of Theorem 3.1. We will use these bounds in a future paper.

3.3 Summary and roadmap for the rest of the paper

We started with the famous and influential matrix completion problem (Section 1.1, its three basic assumptions. Next, we discussed the the noisy setting, which is more realistic (Section 1.2, and focused on recent results concerning exact recovery [24, 25] (Section 1.3). These results require extra assumptions on the spectrum of the ground matrix. On one hand, these assumptions are rather strong and thus significantly limit the application of these results on real data sets. On the other hand, they, quite intriguingly, seem to contradict each other from the mathematical view point. This leads to the conjecture that neither is needed.

We next introduced our own result, which provides an efficient algorithm without requiring the above mentioned assumptions (Section 2), showing that the conjecture is correct. This algorithm obtains exact recovery under only three basic assumptions, and is the first such algorithm for noisy data. In the (easier) noiseless case, the only algorithm using only three basic assumptions is that of Candes et al. [13, 4, 14], which is based on convex optimization. Compared to this algorithm, our is simpler and faster, as it uses only on round of low rank approximation. (Low rank approximation is known to be an efficient operator, used very often in practice.) Thus, our result makes a contribution in the noiseless case as well.

Our main theorem is Theorem 2.3, which guarantees the correctness of our algorithm. This, in essence, is a matrix perturbation bound, and can be seen as an infinity norm variant of the classical Davis-Kahan theorem.

In Section 3, we sketch the proof of Theorem 2.3. We started with the sketch of the arguments of [24, 25], and showed why their extra assumptions are required. Next, we describe our new approach (Section 3.2), which combines the contour integration method intorduced in [27] with novel semi-isotropic bounds for random matrix powers. This approach avoids the use of the extra assumptions in earlier papers. Over all, this has led to a highly non-trivial, but robust and powerful, machinery to obtain matrix perburtation bounds in the infinity norm. We believe that this method will have many other applications.

The supplementary appendix [19], separate from the main body, will contain four sections.

In [19, Section A] , we dive deeper into the technical bounds of other matrix completion papers, including [24, 25] and the RMSE recovery papers, pointing out the strong assumptions they use (which mostly requiring the condition number of the ground matrix to be small), and demonstrate

that we do not need those assumptions. Another point of interest is the sample size needed to recover A with precision ε (or within a RMSE ε), which should grow with $1/\varepsilon$. We demonstrate that our growth factor, which is $1/\varepsilon^2$, is on par with the best of these results.

Consider the simple setting where $K_Z = O(1)$, $\mu_0 = O(\log N)$ and $r = O(\log^2 N)$. Table 1 summarizes the advantages we have over the main methods discussed in this paper.

	Method	Entry-wise recovery?	Time Complexity	Extra assumption to achieve optimal sampling bound
Convex Optimization	Candes and Plan	No	$O(\|\Omega\|^2(m+n)^2)$	Not optimal
Low-rank approx. with grad. descent cleaning up	Keshavan et al.	No	$O(\|\Omega\|r + mn + Lr(m+n))$, if one chooses L iterations for the clearning step	Condition number is small
Single-step low-rank approx. with singular value thresholding	Abbe and Fan	Yes	$O(\|\Omega\|r + mn)$	Condition number is small
	Bhardwaj et al.	Yes	$O(\|\Omega\|r + mn)$	Every singular value gap is large
	Our method	Yes	$O(\|\Omega\|r + mn)$	None

Table 1: Comparison of methods for noisy matrix completion

In [19, Section B] , we introduce our main technical machinery behind the proof of Theorem 3.1. We will fully shift the context to matrix perturbation, and introduce new notation to be used throughout the ensuing discussion and proofs. We will have one main theorem for the contour integral method [19, Theorem B.2] , another for the semi-isotropic bounds [19, Theorem B.4] , and a corollary of their combination [19, Theorem B.6] . We then use this theorem to prove Theorem 3.1, then prove Theorem 2.3 with Theorem 3.1.

In [19, Section C] , we provide the full proofs of the main technical theorems. To assist the readers, these proofs come with their own sketches. In particular, the proof sketch of [19, Theorem B.2] will be a more detailed version of the sketch in Section 3.2. All details will be presented, barring some cumbersome technical lemmas.

Finally, in [19, Section D] , we prove the technical lemmas. The details are heavy, but we hope that the revealed intution earlier will significantly help the readers follow the proofs.

Acknowledgments and Disclosure of Funding

The research is partially supported by Simon Foundation award SFI-MPS-SFM-00006506 and NSF grant AWD 0010308.

References

[1] Robert M. Bell and Yehuda Koren. Lessons from the netflix prize challenge. *SIGKDD Explor. Newsl.*, 9(2):75–79, 12 2007. ISSN 1931-0145. doi: 10.1145/1345448.1345465. URL `https://doi.org/10.1145/1345448.1345465`.

[2] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. ISSN 0209-9683. doi: 10.1007/BF01200757. URL `https://doi.org/10.1007/BF01200757`.

[3] Anthony Man-Cho So and Yinyu Ye. Theory of semidefinite programming for sensor network localization. *Math. Program.*, 109(2-3):367–384, 2007. ISSN 0025-5610,1436-4646. doi: 10.1007/s10107-006-0040-1. URL `https://doi.org/10.1007/s10107-006-0040-1`.

[4] Emmanuel J. Candès and Terence Tao. The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inform. Theory*, 56(5):2053–2080, 2010. ISSN 0018-9448,1557-9654. doi: 10.1109/TIT.2010.2044061. URL https://doi.org/10.1109/TIT.2010.2044061.

[5] M. Mesbahi and G. P. Papavassilopoulos. On the rank minimization problem over a positive semidefinite linear matrix inequality. *IEEE Trans. Automat. Control*, 42(2):239–243, 1997. ISSN 0018-9286,1558-2523. doi: 10.1109/9.554402. URL https://doi.org/10.1109/9.554402.

[6] Xiao Peng Li, Lei Huang, Hing Cheung So, and Bo Zhao. A survey on matrix completion: Perspective of signal processing, 2019. arXiv preprint https://arxiv.org/abs/1901.10885.

[7] Mark A. Davenport and Justin Romberg. An overview of low-rank matrix recovery from incomplete observations. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):608–622, 2016. doi: 10.1109/JSTSP.2016.2539100.

[8] Anish Agarwal, Munther Dahleh, Devavrat Shah, and Dennis Shen. Causal matrix completion. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 3821–3826. PMLR, 12–15 Jul 2023. URL https://proceedings.mlr.press/v195/agarwal23c.html.

[9] Anish Agarwal, Devavrat Shah, Dennis Shen, and Dogyoon Song. On robustness of principal component regression. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/923e325e16617477e457f6a468a2d6df-Paper.pdf.

[10] Anish Agarwal, Devavrat Shah, and Dennis Shen. On model identification and out-of-sample prediction of principal component regression: Applications to synthetic controls, 2023. URL https://arxiv.org/abs/2010.14449. arXiv preprint https://arxiv.org/abs/2010.14449.

[11] Anish Agarwal, Devavrat Shah, and Dennis Shen. Synthetic interventions, 2024. URL https://arxiv.org/abs/2006.07691. arXiv preprint https://arxiv.org/abs/2006.07691.

[12] Emmanuel J. Candès and Yaniv Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010. doi: 10.1109/JPROC.2009.2035722.

[13] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009. ISSN 1615-3375,1615-3383. doi: 10.1007/s10208-009-9045-5. URL https://doi.org/10.1007/s10208-009-9045-5.

[14] Benjamin Recht. A simpler approach to matrix completion. *J. Mach. Learn. Res.*, 12:3413–3430, 2011. ISSN 1532-4435,1533-7928.

[15] Moritz Hardt and Mary Wootters. Fast matrix completion without the condition number. In Maria Florina Balcan, Vitaly Feldman, and Csaba Szepesvári, editors, *Proceedings of The 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pages 638–678, Barcelona, Spain, 13–15 Jun 2014. PMLR. URL https://proceedings.mlr.press/v35/hardt14a.html.

[16] Moritz Hardt. Understanding alternating minimization for matrix completion. In *55th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2014*, pages 651–660. IEEE Computer Soc., Los Alamitos, CA, 2014. ISBN 978-1-4799-6517-5. doi: 10.1109/FOCS.2014.75. URL https://doi.org/10.1109/FOCS.2014.75.

[17] Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *IEEE Trans. Inform. Theory*, 56(6):2980–2998, 2010. ISSN 0018-9448,1557-9654. doi: 10.1109/TIT.2010.2046205. URL https://doi.org/10.1109/TIT.2010.2046205.

[18] Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from noisy entries. *J. Mach. Learn. Res.*, 11:2057–2078, 2010. ISSN 1532-4435,1533-7928.

[19] Anonymous authors. Supplementary appendix to the paper "fast exact recovery of noisy matrix from few entries: the infinity norm approach", 2025. URL https://github.com/thbl2016/Matrix-Completion-NeuRIPS2025/blob/main/Matrix_Completion_NeuRIPS2025_supplementary.pdf.

[20] Sourav Chatterjee. Matrix estimation by universal singular value thresholding. *Ann. Statist.*, 43(1):177–214, 2015. ISSN 0090-5364,2168-8966. doi: 10.1214/14-AOS1272. URL https://doi.org/10.1214/14-AOS1272.

[21] Sohom Bhattacharya and Sourav Chatterjee. Matrix completion with data-dependent missingness probabilities. *IEEE Trans. Inform. Theory*, 68(10):6762–6773, 2022. ISSN 0018-9448,1557-9654. doi: 10.1109/tit.2022.3170244. URL https://doi.org/10.1109/tit.2022.3170244.

[22] Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization (extended abstract). In *STOC'13—Proceedings of the 2013 ACM Symposium on Theory of Computing*, pages 665–674. ACM, New York, 2013. ISBN 978-1-4503-2029-0. doi: 10.1145/2488608.2488693. URL https://doi.org/10.1145/2488608.2488693.

[23] Srinadh Bhojanapalli and Prateek Jain. Universal matrix completion. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 1881–1889, Bejing, China, 22–24 Jun 2014. PMLR. URL https://proceedings.mlr.press/v32/bhojanapalli14.html.

[24] Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, and Yiqiao Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.*, 48(3):1452–1474, 2020. ISSN 0090-5364,2168-8966. doi: 10.1214/19-AOS1854. URL https://doi.org/10.1214/19-AOS1854.

[25] Abhinav Bhardwaj and Van Vu. Matrix perturbation: Davis-Kahan in the infinity norm. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 880–934. SIAM, Philadelphia, PA, 2024. ISBN 978-1-61197-791-2. doi: 10.1137/1.9781611977912.34. URL https://doi.org/10.1137/1.9781611977912.34.

[26] Martin J. Wainwright. *Principal component analysis in high dimensions*, page 236–258. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.

[27] Phuc Tran and Van Vu. New matrix perturbation bounds via combinatorial expansion i: Perturbation of eigenspaces, 2024. URL https://arxiv.org/abs/2409.20207. arXiv preprint https://arxiv.org/abs/2409.20207.

[28] Van Vu. Spectral norm of random matrices. *Combinatorica*, 27(6):721–736, 2007. ISSN 0209-9683,1439-6912. doi: 10.1007/s00493-007-2190-z. URL https://doi.org/10.1007/s00493-007-2190-z.

[29] Afonso S. Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.*, 44(4):2479–2506, 2016. ISSN 0091-1798,2168-894X. doi: 10.1214/15-AOP1025. URL https://doi.org/10.1214/15-AOP1025.

[30] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005. ISSN 0091-1798,2168-894X. doi: 10.1214/009117905000000233. URL https://doi.org/10.1214/009117905000000233.

[31] Delphine Féral and Sandrine Péché. The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.*, 272(1):185–228, 2007. ISSN 0010-3616,1432-0916. doi: 10.1007/s00220-007-0209-3. URL https://doi.org/10.1007/s00220-007-0209-3.

[32] Sandrine Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields*, 134(1):127–173, 2006. ISSN 0178-8051,1432-2064. doi: 10.1007/s00440-005-0466-z. URL https://doi.org/10.1007/s00440-005-0466-z.

[33] Mireille Capitaine, Catherine Donati-Martin, and Delphine Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. *Ann. Probab.*, 37(1):1–47, 2009. ISSN 0091-1798,2168-894X. doi: 10.1214/08-AOP394. URL https://doi.org/10.1214/08-AOP394.

[34] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.*, 227(1):494–521, 2011. ISSN 0001-8708,1090-2082. doi: 10.1016/j.aim.2011.02.007. URL https://doi.org/10.1016/j.aim.2011.02.007.

[35] Farzan Haddadi and Arash Amini. Eigenvectors of deformed Wigner random matrices. *IEEE Trans. Inform. Theory*, 67(2):1069–1079, 2021. ISSN 0018-9448,1557-9654. doi: 10.1109/tit.2020.3039173. URL https://doi.org/10.1109/tit.2020.3039173.

[36] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963. ISSN 0162-1459,1537-274X. URL http://links.jstor.org/sici?sici=0162-1459(196303)58:301<13:PIFSOB>2.0.CO;2-D&origin=MSN.

[37] Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7:1–46, 1970. ISSN 0036-1429. doi: 10.1137/0707001. URL https://doi.org/10.1137/0707001.

[38] Per-Åke Wedin. Perturbation bounds in connection with singular value decomposition. *Nordisk Tidskr. Informationsbehandling (BIT)*, 12:99–111, 1972. ISSN 0901-246X. doi: 10.1007/bf01932678. URL https://doi.org/10.1007/bf01932678.

[39] Sean O'Rourke, Van Vu, and Ke Wang. Random perturbation of low rank matrices: improving classical bounds. *Linear Algebra Appl.*, 540:26–59, 2018. ISSN 0024-3795,1873-1856. doi: 10.1016/j.laa.2017.11.014. URL https://doi.org/10.1016/j.laa.2017.11.014.

[40] BaoLinh Tran and Van Vu. The "power of few" phenomenon: the sparse case. *Random Structures Algorithms*, 66(1):Paper No. e21260, 16, 2025. ISSN 1042-9832,1098-2418. doi: 10.1002/rsa.21260. URL https://doi.org/10.1002/rsa.21260.

[41] Xueyu Mao, Purnamrita Sarkar, and Deepayan Chakrabarti. Estimating mixed memberships with sharp eigenvector deviations, 2019. URL https://arxiv.org/abs/1709.00407. arXiv preprint https://arxiv.org/abs/1709.00407.

[42] Jianqing Fan, Yingying Fan, Xiao Han, and Jinchi Lv. Asymptotic theory of eigenvectors for random matrices with diverging spikes, 2020. URL https://arxiv.org/abs/1902.06846. arXiv preprint https://arxiv.org/abs/1902.06846.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: In Section 1, we claim that (1) exact recovery in the finite precision sense is achieved via truncated SVD with an appropriate truncation point, combined with a rounding step, and that (2) this method works with only the three basic assumptions present in all previous works on exact matrix completion, plus one mild "strong signal" assumption, and that (3) the strong sinal assumption is necessary for noisy data. To justify the first two claims are resolved in the subsequent sections, we provide Algorithm 2.2 in Section 2 and assert its correctness with Theorem 2.3, which uses only the aforementioned assumptions. We prove Theorem 2.3 by shifting to the matrix perturbation perspective, using Theorem 3.1 to complete the proof (with details in Section B of the supplementary appendix). The third claim is justified in Remark 2.5 in Section 2.2.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: In Remark 2.6, we briefly mention that the power of the $\log N$ factor in our sample size assumption (to guarantee that the algorithm works) in Theorem 3.1 can be improved if one can refine the semi-isotropic bounds in Theorem B.4 in the supplementary appendix . In Section 1.3, we also make it clear that exact recovery for noiseless data, in the true sense, has been achieved by Candes, Recht and Tao with the nuclear norm minimization (NNM) method, with only the three basic assumptions. We emphasize that our method is both simpler and faster than NNM, that it is robust for noisy data, and that it requires fewer assumptions than previous fast methods for exact recovery.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.

- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The full proofs are in the supplementary appendix, with every theorem and lemma fully proven, or cited from another paper.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Regarding Figure 1.3, we have described the necessary steps to reproduce the graphs of the singular values and their gaps.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example

(a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
(b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: We feel that the description of the steps to reproduce Figure 1.3 has been clear enough, and the experiment itself is fairly short and non-essential to the results of the paper, so we do not feel the need to publish the code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [NA]

Justification: The paper does not include experiments that require training and testing of Machine Learning models.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: The paper does not include experiments that require training and testing of statistical hypotheses.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: The paper does not include experiments that require comparing performances of different algorithms, and thus does not need to specify the hardwares used. In fact, the experiments to produce Figure 1.3 can be run on any modern computer than can run Anaconda and Python 3.9.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have read and made sure that our paper follows the guidelines in the NeuRIPS Code of Ethics.

   Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: The paper concerns a purely technical problem and its solution, with real-life applications in technologies that socially and politically neutral.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: We have no data or models that are of high risks for misuse.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The Yale face dataset used in producing Figure 1.3 has been properly used and cited, as has the textbook in which it was mentioned.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: We have no released assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.