Proceedings of Machine Learning Research vol ???:1–60, 2025 38th Annual Conference on Learning Theory

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

author names withheld

Editor: Under Review for COLT 2025

Abstract

The matrix recovery (completion) problem, a central problem in data science and theoretical computer science, is to recover a matrix A from a relatively small random set of entries.

While such a task is impossible in general, it has been shown that one can recover A exactly in polynomial time, with high probability, from a random subset of entries, 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.

There are many different algorithms for the task, including convex optimization (Candès and Recht, 2009; Candès and Tao, 2010; Recht, 2011), alternating projection (Hardt, 2014; Hardt and Wootters, 2014), and low rank approximation with gradient descent (Keshavan et al., 2010a,b).

In applications, Candès and Plan (2010) pointed out that it is more realistic to assume that the target (data) matrix is noisy. In this case, these approaches provide an approximate recovery with small root mean square error. However, it is hard to transform such an approximate recovery to an exact one.

Recently, results by Abbe et al. (2020) and Bhardwaj and Vu (2024) concerning approximation in the infinity norm showed that we can achieve exact recovery even in the noisy case, given that the truth matrix A has bounded precision. However, this comes with a caveat. Beyond the three basic assumptions above, they required an extra assumption that either the condition number of A is small (Abbe et al., 2020) or the gap between consecutive singular values is large (Bhardwaj and Vu, 2024).

In this paper, we remove these extra spectral assumptions. As a result, we obtain the first algorithm for exact recovery in the noisy case, under only three basic assumptions. This algorithm is basically computing a low rank approximation, which is simple and fast.

The core of the analysis of our algorithm is an infinity norm version the clasical Davis-Kahan perturbation theorem, improving an earlier result by Bhardwaj and Vu (2024). Our proof relies on a combinatorial contour intergration argument, and is totally different from all previous approaches.

Contents

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 $\Omega \in [m] \times [n]$. We call Ω the set of *observations* or *samples*. The matrix A_Ω, defined by

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

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 which has been received lots of attention in recent years, motivated by a number of real-life applications. Examples include building recommendation systems, notably the **Netflix challenge** (Bell and Koren, 2007); reconstructing a low-dimensional geometric surface based on partial distance measurements from a sparse network of sensors (Linial et al., 1995; So and Ye, 2007); repairing missing pixels in images Candès and Tao (2010); system identification in control (Mesbahi and Papavassilopoulos, 1997). See the surveys by Li et al. (2019) and Davenport and Romberg (2016) for more applications.

It is standard to assume that the set Ω is *random*. Researchers have proposed two models: (a) Ω is sampled uniformly among subsets with the same size, or (b) that Ω has independently chosen entries, each with the same probability p, called the *sampling density*, which can be known or hidden. When doing mathematical analysis, it is often simple to replace the former model by the latter, using a simple trick: one samples the entries independently, and condition on the event that the sample size equals a given number.

The influential paper by Candès and Plan (2010) pointed out that data is often noisy. Thus, a more realistic model for the recovery problem is to consider $A' = A+Z$, for A being the low rank ground truth matrix and Z the noise. We observe a sparse matrix A'_Ω, where each entry of A' is observed with probability p and set to 0 otherwise. In other words, we have access to a small random set of noisy entries. Notice that in this case, the truth matrix A is still low rank, but the noisy matrix A', whose entries we observe, can have full rank.

In this paper, we focus on *exact recovery* in the *noisy setting*, where we want to recover all entries of A exactly even in the noisy model above. In what follows, we denote our input by $A_{\Omega,Z}$, emphasizing the presence of the noise.

1.2. Common settings and notation

To start, we introduce a set of notation:

- When discussing A, we denote $N := \max\{m,n\}$. Let the SVD of A be given by $A = U\Sigma V^T = \sum_{i=1}^r \sigma_i u_i v_i^T$, where $r := \operatorname{rank} A$, and the singular values σ_i are ordered: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. For each $s \in [r]$, let $A_s = \sum_{i=1}^s \sigma_i u_i v_i^T$ be the best rank-s approximation of A. Define B_s analogously for any matrix B. The *condition number* of A is $\kappa = \kappa(A) := \sigma_1/\sigma_r$.
- The *coherence parameter* of U is given by

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

where the 2-to-∞ norm of a matrix M is given by $\|M\|_{2,\infty} := \sup\{\|Mu\|_\infty : \|u\|_2 = 1\}$, which is the largest row norm of M. Define $\mu(V)$ similarly. When U and V are singular bases of A, we let $\mu_0(A) = \max\{\mu(U), \mu(V)\}$ and simply use μ_0 when A is clear from the context. The notion of coherence appears in many fields, and many works use the definition $\mu(U) = m\|U\|_\infty^2$, where the infinity norm is the absolute value of the largest entry. We stick to the definition above, which is consistent with the parameter μ_0 in many popular papers in this area (Candès and Recht, 2009; Candès and Tao, 2010; Recht, 2011; Keshavan et al., 2010a,b; Candès and Plan, 2010).

- We use C to denote a positive universal constant, whose value changes from place to place. The asymptotics notation are used under the assumption that $N \to \infty$. When the value of C depends on another set of parameters $a_1, a_2, \dots, a_k$, we simply use $C(a_1, a_2, \dots, a_k)$.

If we observe only A_Ω, filling out the missing entries is clearly impossible without extra assumptions. Most existing works on the topic made the following three assumptions:

- *Low-rank:* One assumes that $r := \operatorname{rank} A$ is much smaller than $\min\{m, n\}$. This assumption is crucial as it reduces the degree of freedom of A significantly, making the problem solvable, at least from an information theory stand point. Many papers assume r is bounded ($r = O(1)$), while $m, n \to \infty$.

- *Incoherence:* This assumption ensures 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 requires μ_0, and (sometimes) also μ_1 to be small.

- *Sufficient sampling size/density:* Due to a coupon collector effect, both random sampling models above need at least $N \log N$ observations to avoid empty rows or columns. Another lower bound is given by the degree of freedom: One needs to know $r(m + n - 1)$ parameters to compute A exactly. A more elaborate argument in Candès and Tao (2010) gives the lower bound $|\Omega| \geq CrN \log N$. This is equivalent to $p \geq Cr(m^{-1} + n^{-1}) \log N$ for the independent sampling model.

For more discussion about the necessity of these assumptions, we refer to Candès and Recht (2009); Candès and Tao (2010); Davenport and Romberg (2016). All results discussed in this paper make these assumptions, so we will refer to them as the *basic assumptions*.

In the noisy setting, most papers assume that Z has independent entries with mean 0, not necessarily of the same distributions. In the subcategory of works that our paper belongs to, which aims for exact recovery using a single SVD step with thresholding, it is also common to assume Z has bounded entries with probability 1.

1.3. Our goal: exact recovery in practice with finite precision

We say that a matrix A have precision ε_0, if its entries are integer multiples of a parameter $\varepsilon_0 > 0$. For instance, if all entries have two decimal places, then $\varepsilon_0 = .01$. In practice, all data stored in any computer system are binary numbers, and thus have precision at least 2^{-b}, where b is the number of bits the system uses to represent numbers.

In many pratical applications, such as recommendation systems, ε_0 is much larger. For instance, in the famous Netflix Challenge (Bell and Koren, 2007), the entries of A, which are movies ratings, are half integers from 1 to 5, so $\varepsilon_0 = 1/2$. In this paper, we propose and analyze an algorithm, which exploits this fact to recover A exactly, with only a mild assumption besides the basic ones. Our work is not the first to achieve exact recovery, but to the best of our knowledge, is the first to use only one SVD step without needing strong conditions on A such as having a small condition number or large singular value gaps.

1.4. A brief summary of existing methods

There is a huge literature on matrix completion. In this section, we summarize some of the main methods. Our focus is on the last category, *low-rank approximation with thresholding*. We encourage readers to read the surverys by Li et al. (2019) and Davenport and Romberg (2016) for more thorough reviews of the other approaches.

- *Nuclear norm minimization:* Proposed by Candès and Recht (2009) and subsequently perfected by Candès and Tao (2010) and Recht (2011), this method achieves *true exact recovery* (not a *close enough* approximation), requiring only the three basic assumptions. The idea is based on convexifying the intuitive but NP-hard approach of minimizing the rank given the observations. However, the best-known solvers for this method run in time $O(|\Omega|^2 N^2)$, which is $O(N^4 \log^2 N)$ in the best case (Li et al., 2019), making it less practical, and the calculation may be sensitive to noise (Candès and Tao, 2010). Therefore, faster algorithms that can perform well are still of interest.

- *Alternating projections:* Proposed by Hardt and Wootters (2014) and extended to the noisy case by Hardt (2014), this is method is based on another intuitive but NP-hard approach of fixing the rank, then minimizing the RMSE with the observations. The basic version of the algorithm switches between optimizing the column and row spaces, given the other, in alternating steps. Existing variants of alternating projections run well in practice, but to the best of our knowledge, require a small condition number and are iterative algorithms with an uncertain number of SVD steps.

- *Low-rank approximation with thresholding:* The general idea here is to view the sample matrix A_Ω as a rescaling of an unbiased random perturbation of A. This way, it is natural to first approximate A by taking a low rank approximation of $p^{-1}A_\Omega$ (where p is the sampling density) at an appropriate truncation point (*thresholding*), then add an extra cleaning step to make the recovery exact. The first step involves only one truncated SVD operation, and runs fast in practice. Our new algorithm in Section 1.6 belong to this category, along many other works (Keshavan et al., 2010a,b; Abbe et al., 2020; Bhardwaj and Vu, 2024) that we will discuss in the next part.

1.5. Low-rank approximation with thresholding: a dilemma

In this approach, one exploits the fact that A has finite precision (which is necessary for exact recovery to make sense in practice, see subsection 1.3). If each entry of A is an integer multiple of ε_0, then to achieve an exact recovery, it suffices to compute each entry with error less than $\varepsilon_0/2$, and then round it off. In other words, it is sufficient to obtain an approximation of A in the *inifnity norm.* It has been shown, under different extra assumptions, that low rank approximation fullfils this purpose.

The first infinity norm result was obtained by Abbe et al. (2020). They showed that the best rank-r approximation of $p^{-1}A_\Omega$ is close to A in the infinity norm (Abbe et al., 2020, Theorem 3.4). Technically, they proved that for some universal constant C, if

$$\sigma_r \geq C\kappa(\|A\|_\infty + \varsigma_Z)\sqrt{\frac{N\log N}{p}} \quad \text{and} \quad p \geq 6N^{-1}\log N,$$

where ς_Z is the standard deviation of each entry of Z, then

$$\|p^{-1}(A_\Omega)_r - A\|_\infty \leq C\mu_0^2\kappa^4(\|A\|_\infty + \varsigma_Z)\sqrt{\frac{\log N}{pN}}.$$

We remind the readers that $\kappa = \sigma_1/\sigma_r$ is the condition number. In other words, to achieve an error less than ε in the infinity norm, one needs

$$p \geq C^2\varepsilon^{-2}\mu_0^4\kappa^8(\|A\|_\infty + \varsigma_Z)^2 N^{-1}\log N, \tag{3}$$

If we turn this result into an algorithm (by adding the rounding-off cleaning step), then we face two issues: the algorithm needs to know the rank r, and the condition number κ has to be small ($\kappa = O(1)$), which is a rather a strong assumption.

Eliminating the condition number. Very recently, Bhardwaj and Vu (2024) proposed an SVD-based algorithm with a dynamically chosen truncation point, which has the advantage of working without knowing the rank r.

Algorithm 1 (Approximate-and-Round)

1. *Let $\tilde{A} := p^{-1}A_\Omega$ and compute the SVD: $\tilde{A} = \tilde{U}\tilde{\Sigma}\tilde{V}^T = \sum_{i=1}^{m\wedge n}\tilde{\sigma}_i\tilde{u}_i\tilde{v}_i^T$.*
2. *Let $\tilde{s}$ be the last index such that $\tilde{\sigma}_i \geq \frac{N}{8r\mu}$, where $\mu := N\max\{\|U\|_\infty^2, \|V\|_\infty^2\}$ is known.*
3. *Let $\hat{A} := \sum_{i=1}^{\tilde{s}}\tilde{\sigma}_i\tilde{u}_i\tilde{v}_i^T$.*
4. *Round off every entry of $\hat{A}$ to the nearest integer.*

Next, their bound on $|\Omega|$ does not include the condition number κ. In the case A has integer entries (the precision $\varepsilon_0 = 1$), they showed that with probability $1 - o(1)$, before the rounding step, one has $\|\hat{A} - A\|_\infty < 1/2$, guaranteeing an exact recovery of A after rounding, under the following assumptions: (1) $r = O(1)$ (low rank), (2) $\mu_0 = O(1)$ (incoherence), (3) $p \geq N^{-1}\log^{4.03} N$ (sampling density), (4) $\|A\|_\infty \leq K_A$ for a known constant K_A (bounded entries) and (5) $\min_{i\in[s]}(\sigma_i - \sigma_{i+1}) \geq Cp^{-1}\log N$.

In the case of a general precision $\varepsilon_0 = \varepsilon$ and a noisy matrix Z with entries bounded by K_Z, their analysis can be scaled to produce following sampling density condition:

$$p \geq C(r, \mu_0, K_A, K_Z)\varepsilon^{-5}N^{-1}\log^{4.03} N. \tag{4}$$

To the best of our knowledge, their work is the first that achieves exact recovery with a single SVD step without depending on the condition number. However, they need a new assumption on the gaps between consecutive singular values, and the scaling of ε in the sampling density condition is larger than Abbe et al. (2020).

The reader may have already noticed that this gap assumption, at least in spirit, goes into the *opposite direction* of the small condition number assumption. Indeed, if the gaps between the consecutive singular values are large, then its suggests that the singular values decay fast, and the condition number is also large. So, from the mathematical view point, the situation is quite intriguing. We have two valid theorems with *constrasting extra assumptions* (beyond the three basic assumptions). The most logical explanation here should be that neither assumption is in fact needed. This leads to the main question of this paper:

Question 2 *Can we use the fast low-rank approximation approach to obtain exact recovery in the noisy case with only the three basic assumptions (low rank, incoherence, density)?*

Another point of interest is the efficiency of the sampling density condition, which can be measured by the asymptotic growth of p in order to recover A within a precision ε. If possible, one should aim for this growth to be quadratic, comparable to Abbe et al. (2020).

In the next section, we will answer Question 2, introduce our algorithm and its analysis. Before beginning, we would like to mention the works of Keshavan et al. (2010a,b) and Chatterjee (2015), which also used low-rank approximation with thresholding. The different is they only aimed to achieve a RMSE (Frobenius norm) recovery, with the former also depending on the condition number.

1.6. New results: an affirmative answer to Question 2

1.6.1. An overview and the general setting

The goal of this paper is to give an affirmative answer to Question 2, in a sufficiently general setting. We will show that a variant of Approximate-and-Round will do the job. The technical core is a new mathematical method to extend and strengthen the classic Davis-Kahan-Wedin theorem (Wedin, 1972) for infinity norm estimates. Our technique takes heavy inspiration from the power series approach by Tran and Vu (2024), with non-trivial adjustments and additions. This is entirely different from all previous techniques, and is of independent interest.

This affirmative answer is a unifying result for fast recovery in the noisy setting. It provides the first and efficient exact recovery algorithm using only three basic assumptions. To state the algorithm and main result, let us restate the setting for clarification purposes.

Setting 3 (Matrix completion with noise) *Consider the truth matrix A, the observed set Ω, and noise matrix Z. We assume*

1. Known bound on entries: *We assume $\|A\|_\infty \le K_A$ for some known parameter K_A. This is the case for most real-life applications, as entries have physical meaning. For instance, in the Netflix Challenge $K_A = 5$.*

2. Known bound on rank: *We do not assume the knowledge of the rank r, but assume that we know some upper bound $r_{\max}$.*

3. Independent, bounded, centered 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 $l \in \mathbb{N}$ and $i \in [m]$, $j \in [n]$. We assume the knowledge of the upper bound K_Z. We do not require the entries to have the same distribution or even the same variance.*

We allow the parameters r, $r_{\max}$, K_A, K_Z to depend on m and n.

1.6.2. Our algorithm and theorem

We propose the following algorithm to recover A:

Algorithm 4 (Approximate-and-Round 2)

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

   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 $\hat{A}_s := \sum_{i \le s} \hat{\sigma}_i \hat{u}_i \hat{v}_i^T$,
4. Rounding off: *Round each entry of $\hat{A}_s$ to the nearest multiple of ε_0. Return $\hat{A}_s$.*

Compared to Approximate-and-Round, a minor difference is that we use an estimate $\hat{p}$ of p, which is very accurate with high probability. We next use a different cutoff point for the truncated SVD step that does not require knowing the parameter μ_0.

From a complexity viewpoint, our algorithm is very fast, being simply truncated SVD, taking only $O(|\Omega|r) = O(pmnr)$ FLOPs. Our main theorem below gives sufficient conditions for exact recovery with this algorithm.

Theorem 5 *There is a universal constant $C > 0$ such that the following holds. Suppose $r_{\max} \le \log^2 N$. Under the model 3, assume the following:*

- Large signal: *$\sigma_1 \ge 100 r K \sqrt{\frac{r_{\max} N}{p}}$, for $K := K_A + K_Z$.*
- Sampling density:

$$p \ge C\left(\frac{1}{m} + \frac{1}{n}\right) \max\left\{\log^4 N,\ \frac{r^3 K^2}{\varepsilon_0^2}\left(1 + \frac{\mu_0^2}{\log^2 N}\right)\right\} \log^6 N. \tag{5}$$

Then with probability $1 - O(N^{-1})$, the first three steps of Approximate-and-Round 2 recovers every entry of A within an absolute error $\varepsilon_0/3$. Consequently, if all entries are multiples integer of ε_0, the rounding-off step recovers A exactly.

If we resctrict to the noiseless case, our new result also provides some improvements. As discussed in Section 1.4, in the noiseless case, the only approach that achieves exact recovery without extra assumptions is the nuclear norm minimization (subsection 2.1), with trade-off being its slow running time. The other known methods are faster, at the cost of having extra assumptions, or having a weaker form of recovery such as RMSE.

Our algorithm gets the "best of the both worlds". First, it uses only the three basic assumptions. Next, it is also very fast and perhpas the simplest from the practical view point. Technically speaking, the algorithm is basically a truncated SVD, which is effective in both theory and practice. Compared to the spectral approach by Keshavan et al, it does not need the second phase (using gradient descent). Finally, the density bound is comparable to all previous works; see Remarks 7 and 8.

1.6.3. Analysis of the assumptions

Notice that we have removed the gap condition from Bhardwaj and Vu (2024), and thus obtained an exact recovery algorithm, using only the three basic assumptions: low rank, incoherence, density, and the practically common bounded entries assumption.

Well, almost! The reader, of course, has noticed that we have added (formally, at least) a new assumption that the leading singular value σ_1 of A has to be sufficiently large (large signal). In what follows, we argue that this assumption is *necessary*. Next, we show that it is not really new. Most of the times, while not stated explicitely, it is a corollary of the setting of the problem, and the bound is used in the analysis.

Remark 6 (Necessity of the signal assumption) *To start, one must notice, even at first glance, that σ_1 should be large enough. From a pure engineering viewpoint, if σ_1 is too small, then the intensity of the noise dominates the ground truth data (signal), making the observed data must be totally corrupted. Now comes the mathematical rigour. It is a well known fact in random matrix theory, called the BBP threshold phenomenon Baik et al. (2005), that if $\|Z\| \geq c\|A\| = c\sigma_1$, for a specific constant c, then the sum $A+Z$ behaves like a random matrix; see Baik et al. (2005); Féral and Péché (2007); Péché (2006); Capitaine et al. (2009); Benaych-Georges and Nadakuditi (2011); Haddadi and Amini (2021) for many results. For instance, the leading singular vectors of $A+Z$ look totally random and have nothing to do with the leading singular vectors of A. This shows that there is no chance that one can recover A from the (even fully observed) noisy matrix $A+Z$.*

In the results discussed earlier, this large signal assumption (if not explictly stated) is implicit in the setting of the model. To start, the the bound required for σ_1 is mild. In most cases, it is automatically satisfied by the simple fact that $\sigma_1^2 \geq r^{-1}\|A\|_F^2$. To see this, let us consider the base case when $n = \Theta(m)$, $\frac{1}{n}\|A\|_F = \Theta(1)$ (a normalization, for convenience), $r = O(1)$. In this case,

$$\sigma_1 \geq r^{-1}\|A\|_F = \tilde{\Omega}(n). \tag{6}$$

If we assume $K_A, K_Z = O(1)$, then our requirement on σ_1 in the theorem becomes

$$\sigma_1 = \Theta(\sqrt{n/p}),$$

which is guaranteed automatically (with room to spare) by (6), since we need $p = \Omega(\log n/n)$. As a matter of fact, the requirement $\sigma_1 = \Theta(\sqrt{n/p})$ is consistend with the BBP phenomenon *discussed above. Indeed, if we consider the rescaled input*

$$p^{-1}A_{\Omega,Z} = p^{-1}A_\Omega + p^{-1}Z,$$

then the $\sqrt{n/p}$ *is the order of magnitude of the spectral norm of noisy part* $\|\frac{1}{p}Z_\Omega\|$*, and the spectral norm of* $p^{-1}A_\Omega$ *is approximately* $\|A\|$*, with high probability. (It is well known that one can approximate the spectral norm by random sampling.)*

Remark 7 (Optimality of the density bound) *The condition (5) looks complicated, but in the base case where* A *has constant rank, constantly bounded entries, and uniformly random singular vectors, we have* $\mu_0 = O(\log N)$*,* $K_A, K_Z = O(1)$ *and* $r_{\max} = O(1)$*, it reduces to*

$$p \geq C \max\left\{\log^4 N,\ \varepsilon_0^{-2}\right\}\left(m^{-1} + n^{-1}\right)\log^6 N, \tag{7}$$

which is equivalent to $|\Omega| \geq CN\log^6 N \max\{\log^4 N, \varepsilon_0^{-2}\}$ *in the uniform sampling model. 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.*

Even when reduced to the noiseless case, this bound is comparable, up to a polylogarithmic factor, to all previous results, while having the key advantage that it does not contain a power of the condition number.

Remark 8 (Optimality of the dependency on $1/\varepsilon_0$**)** *The required* $|\Omega|$ *for exact recovery is* $O(N\log^{10} N)$ *until* $\varepsilon_0 < \log^{-2} N$*, then it grows quadratically with* $1/\varepsilon_0$*. All results we know so far achieve either the same or a worse factor of* $1/\varepsilon_0$*, suggesting that it may be optimal.*

Remark 9 (Relaxing the bound on $r_{\max}$**)** *The condition* $r_{\max} \leq \log^2 N$ *in Theorem 5 can be avoided, at the cost of a more complicated sampling density bound. The full form of our bound is*

$$p \geq C\left(\frac{1}{m} + \frac{1}{n}\right)\max\left\{\log^{10} N,\ \frac{r^4 r_{\max}\mu_0^2 K^2}{\varepsilon_0^2}, \frac{r^3K^2}{\varepsilon_0^2}\left(1 + \frac{\mu_0^2}{\log^2 N}\right)\left(1 + \frac{r^3\log N}{N}\right)\log^6 N\right\}. \tag{8}$$

The proof of the more general version of Theorem 5 with Eq. (8) replacing Eq. (5) will be in Appendix B. We do not know of a natural setting where one expects $r_{\max} > \log^2 N$*. Nevertheless this shows that our technique does not require any extra condition besides the large signal assumption.*

1.6.4. Roadmap for the rest of the paper

In Section 2, we will give a proof sketch for Theorem 5, asserting the correctness of our algorithm, Approximate-and-Round 2. We will reframe the problem from a matrix perturbation perspective, then introduce our main tools for the proof, Theorem 14 and its "random" version, Theorem 17, and briefly discuss them in the larger context of matrix perturbation theory.

Due to the technical cumbersome of these two theorems, we actually inserts an intermediate result, Theorem 10, which is a corollary of Theorem 17 specifically for the matrix completion setting, and use it to directly prove Theorem 5. Due to the format constraints, the full proof, as well as other technical proofs will be in the Appendices.

2. The main technical theorems

2.1. Proof sketch of Theorem 5

Theorem 5 aims to bound the difference $\hat{A}_s - A$ in the infinity norm. We achieve this using a series of intermediate comparisons, outlined below. Note that when we say an term is small, we mean that it will vanish as $\min\{m, n\} \to \infty$ under the conditions of Theorem 5.

1. Let $\rho := \hat{p}/p$. We bound $\|A - \rho^{-1}A\|_\infty$. As shown by a Chernoff bound, ρ is very close to 1, making this error small. It remains to bound

   $$\|\rho^{-1}A - \hat{A}_s\|_\infty = \rho^{-1}\|A - (p^{-1}A_{\Omega,Z})_s\|_\infty.$$

   Let $\tilde{A} := p^{-1}A_{\Omega,Z}$, this is equivalent to bounding $\|A - \tilde{A}_s\|_\infty$, since $\rho = \Theta(1)$.

2. We bound $\|A - A_s\|_\infty$. Some light calculations give $\|A - A_s\|_\infty \le \sigma_{s+1}\|U\|_{2,\infty}\|V\|_{2,\infty}$. Using the fact $\hat{\sigma}_i - \hat{\sigma}_{i+1}$ is small for all $i > s$, we can deduce that $\hat{\sigma}_{s+1}$ is small, making σ_{s+1} small too. Coupled with the incoherence property, this error will be small.

3. We bound $\|A_s - \tilde{A}_s\|_\infty$. Most of the heavy lifting is done here, so we will discuss it in detail. Observe that $\mathbf{E}\left[A_\Omega\right] = pA$ and $\mathbf{E}\left[Z_\Omega\right] = 0$, we have

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

   Let $E := \tilde{A} - A$. The above shows that E is a random matrix with mean 0. From a *matrix perturbation theory* view point, $\tilde{A}$ is an *unbiased* perturbation of A. Establishing a bound on $(A + E)_s - A_s$ in the infinity norm is one of the major goals of perturbation theory, and is the main technical contribution of our paper. We will use the following theorem. The essence of this theorem is that under certain conditions, a low rank approximation of the noisy matrix $A + Z$ approximates A very well entry-wise. This can be seen as an improved version of the main theorem of Bhardwaj and Vu (2024). Our approach to the proof also differs entirely from their paper.

Theorem 10 *Consider a fixed matrix $A \in \mathbb{R}^{m\times n}$. and a random matrix $E \in \mathbb{R}^{m\times n}$ with independent entries satisfying $\mathbf{E}\left[E_{ij}\right] = 0$ 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 and C' such that, if $p \ge C(m^{-1} + n^{-1})\log N$ where $N = m + n$, then

$$\|\tilde{A}_s - A_s\|_\infty \le C'\frac{(\log N + \mu_0)\log^2 N}{\sqrt{mn}} \cdot r\sigma_s\left(\frac{K}{\sigma_s}\sqrt{\frac{N}{p}} + \frac{rK\sqrt{\log N}}{\delta_s\sqrt{p}} + \frac{r^2\mu_0 K\log N}{p\delta_s\sqrt{mn}}\right). \quad (9)$$

Remark 11 *This theorem essentially gives a bound on the perturbation of the best rank-s approximation, given that the perturbation on A is random with sufficiently bounded moments and the singular values up to rank s are sufficiently separated from the rest. In matrix theorem, we know very few estimates using the inifnity norm, and we believe that this theorem is of independent interest.*

In practice, when applied to the matrix completion problem, this separation property is guaranteed thanks to the assumptions that the matrix A has low rank and σ_1 is large enough, guaranteeing that such index s exists. We do not need to make an extra assumption here.

Remark 12 *To make sense of the bound (9), we reveal its true form in terms of E under the "veil of randomness". For simplicity, consider only the case $m = \Theta(n)$, making both $\Theta(N)$. Suppose either of the following two popular scenarios: A is completely incoherent, namely $\mu_0 = O(1)$; or A has independent Gaussian $N(0,1)$ entries, making $\mu_0 = O(\log N)$. In this setting, the first factor on the right-hand side of (9) is simply $O\left(\frac{\log^3 N}{N}\right)$. As shown near the end (Section A.1), with probability $1 - O(N^{-1})$,*

$$\|E\| = O\left(K\sqrt{\frac{N}{p}}\right), \quad \|U^T E V\|_\infty = O\left(\frac{r\mu_0 K \log N}{p\sqrt{mn}} + \frac{K\sqrt{\log N}}{\sqrt{p}}\right).$$

Therefore, (9) is just the random version of the following:

$$\text{If } \delta_s \geq C_1 r\|E\|, \quad \text{then } \|\tilde{A}_s - A_s\|_\infty \leq \frac{C' \log^3 N}{N} r\sigma_s \left(\frac{\|E\|}{\sigma_s} + \frac{r\|U^T E V\|_\infty}{\delta_s}\right). \tag{10}$$

This form reveals the underlying structure of the bound: a Davis-Kahan-Wedin type bound of the change in the low-rank approximation, for the infinity norm. We will clarify what this means in the next subsection.

3. Davis-Kahan-Wedin theorem in the infinity norm

Let us reformalize the *matrix perturbation* setting. In this section, we do not need to know anything about the matrix completion context.

Setting 13 (Matrix perturbation) *Consider a fixed $m \times n$ matrix A with SVD*

$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T, \qquad \text{where } \sigma_1 \geq \sigma_2 \geq \ldots \sigma_r.$$

Consider a $m \times n$ matrix E, which can be deterministic or random, which we called the perturbation matrix. *Let $\tilde{A} = A + E$ be the* perturbed matrix *with the following SVD:*

$$\tilde{A} = \tilde{U}\tilde{\Sigma}\tilde{V}^T = \sum_{i=1}^{m\wedge n} \tilde{\sigma}_i \tilde{u}_i \tilde{v}_i^T \qquad \text{where } \tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \ldots \tilde{\sigma}_r.$$

Define the following terms related to A and $\tilde{A}$:

1. *For $k \in [r]$, $\delta_k := \sigma_k - \sigma_{k+1}$, using $\sigma_{r+1} = 0$, and let $\Delta_k := \delta_k \wedge \delta_{k-1}$.*

2. *For $S \subset [r]$, let $\sigma_S := \min\{\sigma_i : i \in S\}$ and $\Delta_S := \min\{|\sigma_i - \sigma_j| : i \in S, j \in S^c\}$.*

3. *For $S \subset [r]$, define the following matrices:*

$$V_S := \big[v_i\big]_{i \in S}, \quad U_S := \big[u_i\big]_{i \in S}, \quad A_S := \sum_{i \in S} \sigma_i u_i v_i^T.$$

When $S = [s]$ for some $s \in [r]$, we also use V_s, U_s, A_s in place of the three above.

Define analogous notations $\tilde{\delta}_k$, $\tilde{\Delta}_k$, $\tilde{\sigma}_S$, $\tilde{\Delta}_S$, $\tilde{V}_S$, $\tilde{U}_S$, and $\tilde{A}_S$ for $\tilde{A}$.

Some extra notation. To aid the presentation, we define the notation:

- $a_1 \wedge a_2 \wedge \ldots \wedge a_n := \min\{a_1, a_2, \ldots, a_n\}$ for every $a_1, a_2, \ldots, a_n \in \mathbb{R}$.
- $a_1 \vee a_2 \vee \ldots \vee a_n := \max\{a_1, a_2, \ldots, a_n\}$ for every $a_1, a_2, \ldots, a_n \in \mathbb{R}$.
- For each $a, b \in \mathbb{R}$, let $[a] := \{x \in \mathbb{Z} : 1 \le x \le a\}$, and $[[a, b]] := \{x \in \mathbb{Z} : a \le x \le b\}$.

3.1. The Davis-Kahan-Wedin theorem for the spectral norm

One of the most well-known results in perturbation theory is the **Davis-Kahan** $\sin\Theta$ **theorem**, proposed by Davis and Kahan (1970), which bounds the change in eigenspace projections by the ratio between the perturbation and the eigenvalue gap. The extension for singular subspaces, proven by Wedin (1972), states that:

$$\|\tilde{U}_s\tilde{U}_s^T - U_sU_s^T\| \vee \|\tilde{V}_s\tilde{V}_s^T - V_sV_s^T\| \le \frac{C\|E\|}{\delta_s} \text{ for a universal constant } C. \tag{11}$$

There are three challenges if one wants to apply this theorem to tackle the problem of bounding $\|\tilde{A} - A\|_\infty$.

1. The inequality above only concerns the change in the singular subspace projections, while the change in the low-rank approximation $\tilde{U}_s\tilde{\Sigma}_s\tilde{V}_s^T$ is needed.
2. The bound on the right-hand side requires the spectral gap-to-noise ratio at index s to be large to be useful, which is a strong assumption.
3. The left-hand side is the operator norm, while an infinity norm bound is needed for exact recovery after rounding.

A key observation is that, similarly to the Frobenius norm bound in Candès and Plan (2010), Eq. (11) works for all perturbation matrices E. Per the discussion in Tran and Vu (2024), the worst case (equality) only happens when there are special interactions between E and A. A series of papers by O'Rourke et al. (2018); Tran and Vu (2024) exploited the improbability of such interactions when E is random and A has low rank, and improved the bound significantly. For instance, O'Rourke et al. (2018) proved the following:

$$\|\tilde{V}_s\tilde{V}_s^T - V_sV_s^T\| \le C\sqrt{s}\left(\frac{\|E\|}{\sigma_s} + \frac{\sqrt{r}\|U^TEV\|_\infty}{\delta_s} + \frac{\|E\|^2}{\delta_s\sigma_s}\right),$$

with high probability, effectively turning the *noise-to-gap* on the right-hand side of Eq. (11) into the *noise-to-signal ratio*, which can be much smaller than the former in many cases.

Tran and Vu (2024) then improved the third term, at the cost of an extra factor of $\sqrt{r}$, which does not matter when A has constant rank. They showed that when

$$\frac{\|E\|}{\sigma_s} \vee \frac{2r\|U^T EV\|_\infty}{\delta_s} \vee \frac{\sqrt{2r}\|E\|}{\sqrt{\delta_s \sigma_s}} \leq \frac{1}{8},$$

then

$$\|\tilde{V}_s \tilde{V}_s^T - V_s V_s^T\| \leq Cr \left(\frac{\|E\|}{\sigma_s} + \frac{2r\|U^T EV\|_\infty}{\delta_s} + \frac{2ry}{\delta_s \sigma_s}. \right),$$

replacing the term $\|E\|^2$ in the previous result with the smaller $y := \frac{1}{2}\max_{i\neq j}(|u_i^T EE^T u_j| + |v_i^T E^T E v_j|)$. It is clear that this quantity is at most $\|E\|^2$. However, it can be significantly smaller in many cases, notably when E is *regular* (Tran and Vu, 2024),.

The notion of random regular matrices cover most models of random matrices used in practice, such as Wigner matrices or genearalized Wigner matrices.

Their method is easily adaptable to prove similar bounds for the deviation of other spectral entities, with respect to difference matrix norms. The main idea is to write the difference of the two entities as a contour integral, then use a combinatorial expansion to split this integral into many subsums, each of which can be treated using tools from complex analysis, linear algebra, and combinatorics. See Section 3 of Tran and Vu (2024) for a detailed discussion.

3.2. A new Davis-Kahan-Wedin theorem for the infinity norm

Our main result can be seen as the ìnfinity norm version of the results discussed above. Proving ìnfinity norm bounds is a recent trend in matrix theory. In particular, most recent progesses concerning universality of random matrices crucially rely on strong infinity norm bounds of the eigenvectors of the matrix (Erdös and Yau, 2017). Obtaining a strong infinity norm bound for a vector or a matrix is usually a highly non-trivial task, as demonstrated in the theory of random matrices, and we hope that the method introduced in this paper will be of independent interest.

Theorem 14 *Consider the objects in Setting 13. Define the following terms:*

$$\begin{aligned} \tau_1 &:= \max_{0\leq a\leq 10\log(m+n)} \frac{1}{\sqrt{r}} \max\left\{ \frac{\left\|(EE^T)^a U\right\|_{2,\infty}}{\|E\|^{2a}}, \frac{\left\|(EE^T)^a EV\right\|_{2,\infty}}{\|E\|^{2a+1}} \right\}, \\ \tau_2 &:= \max_{0\leq a\leq 10\log(m+n)} \frac{1}{\sqrt{r}} \max\left\{ \frac{\left\|(E^T E)^a V\right\|_{2,\infty}}{\|E\|^{2a}}, \frac{\left\|(E^T E)^a E^T U\right\|_{2,\infty}}{\|E\|^{2a+1}} \right\}. \end{aligned} \tag{12}$$

Consider an arbitrary subset $S \subset [r]$. Suppose that S satisfies

$$\frac{\|E\|}{\sigma_S} \vee \frac{2r\|U^T EV\|_\infty}{\Delta_S} \vee \frac{\sqrt{2r}\|E\|}{\sqrt{\Delta_S \sigma_S}} \leq \frac{1}{8}, \tag{13}$$

Then there is a universal constant C such that

$$\left\|\tilde{V}_S\tilde{V}_S^T - V_SV_S^T\right\|_\infty \leq C\tau_1^2 r\left(\frac{\|E\|}{\sigma_S} + \frac{2r\|U^TEV\|_\infty}{\Delta_S} + \frac{2ry}{\Delta_S\sigma_S}\right), \tag{14}$$

$$\left\|\tilde{V}_S\tilde{V}_S^T - V_SV_S^T\right\|_{2,\infty} \leq C\tau_1 r\left(\frac{\|E\|}{\sigma_S} + \frac{2r\|U^TEV\|_\infty}{\Delta_S} + \frac{2ry}{\Delta_S\sigma_S}\right), \tag{15}$$

where

$$y := \frac{1}{2}\max_{i\neq j}(|u_i^TEE^Tu_j| + |v_i^TE^TEv_j|).$$

When $S = [s]$ for some $s \in [r]$, we also have

$$\|\tilde{A}_s - A_s\|_\infty \leq C\tau_1\tau_2\sigma_s r\left(\frac{\|E\|}{\sigma_s} + \frac{2r\|U^TEV\|_\infty}{\delta_s} + \frac{2ry}{\delta_s\sigma_s}\right), \tag{16}$$

Analogous bounds for U and $\tilde{U}$ hold, with U and V swapped.

Remark 15 (Sharpness of the results I) *It is worth noting that the term*

$$r\left(\frac{\|E\|}{\sigma_s} + \frac{2r\|U^TEV\|_\infty}{\delta_s} + \frac{2ry}{\delta_s\sigma_s}\right)$$

is identical to the bound on $\left\|\tilde{V}_S\tilde{V}_S^T - V_SV_S^T\right\|_{\mathrm{op}}$ in Tran and Vu (2024), which is tight when $r = O(1)$. An routine corollary of their main result also implies that

$$r\sigma_s\left(\frac{\|E\|}{\sigma_s} + \frac{2r\|U^TEV\|_\infty}{\delta_s} + \frac{2ry}{\delta_s\sigma_s}\right)$$

is a tight upper bound for $\|\tilde{A}_s - A_s\|$ when $r = O(1)$.

The term y can be trivially upper-bounded by $\|E\|^2$. In fact, the slightly weaker bound with $\|E\|^2$ replacing y looks more natural and consistent with the condition (13). This bound was discovered by O'Rourke et al. (2018) and was the best-known until Tran and Vu (2024). In many cases, notably when E is a stochastic random matrix, *y can be much smaller than $\|E\|$ (see Tran and Vu (2024) for a detailed discussion).*

Remark 16 (Sharpness of the results II) *The terms τ_1 and τ_2 play the roles of the coherence parameters in the matrix completion setting. Practically, one replaces them with upper bounds when applying Theorem 14, as the theorem still works after such substitutions. A trivial choice is $\tau_1 = \tau_2 = 1/\sqrt{r}$, since we have*

$$\|(EE^T)^aU\|_{2,\infty} \leq \|E\|^{2a}, \quad \|(E^TE)^aEU\|_{2,\infty} \leq \|E\|^{2a+1},$$

and analogously for V. This is the best we can have in worst case analysis.

However, if E and A interact favorably, then we can get much better estimates. Let us first consider a bound from below. Setting $a = 0$, we get from Eq. (12) the lower bounds

$$\tau_1 \geq \frac{1}{\sqrt{r}}\|U\|_{2,\infty} = \sqrt{\frac{\mu(U)}{m}}, \quad \tau_2 \geq \frac{1}{\sqrt{r}}\|V\|_{2,\infty} = \sqrt{\frac{\mu(V)}{n}},$$

where $\mu(U)$ and $\mu(V)$ are the individual incoherence parameters from Eq. (2).

If these lower bounds are the truth, then by Remark 15, one gets, in philosophy, the following bounds from Theorem 14 (when $r = O(1)$):

$$\begin{aligned}
\left\|\tilde{V}_S\tilde{V}_S^T - V_SV_S^T\right\|_\infty &\le C\frac{\mu(V)}{n}\left\|\tilde{V}_S\tilde{V}_S^T - V_SV_S^T\right\|, \\
\left\|\tilde{V}_S\tilde{V}_S^T - V_SV_S^T\right\|_{2,\infty} &\le C\sqrt{\frac{\mu(V)}{n}}\left\|\tilde{V}_S\tilde{V}_S^T - V_SV_S^T\right\|, \\
\left\|\tilde{A}_s - A_s\right\|_\infty &\le C\sqrt{\frac{\mu(U)\mu(V)}{mn}}\left\|\tilde{A}_s - A_s\right\|.
\end{aligned} \tag{17}$$

These are the best possible bounds one can hope to produce with Theorem 14. But how good are they? *To answer this question, let us consider a simple case where $r = O(1)$, $\mu(V) = O(1)$, and $m = \Theta(n)$. Assume the best possible case for the parameters τ_2, which is that $\tau_2 = \sqrt{\mu(V)/n} = O(n^{-1/2})$. In this case, Eq.* (14) *asserts that*

$$\left\|\tilde{V}_S\tilde{V}_S^T - V_SV_S^T\right\|_\infty = O\left(\frac{1}{n}\left\|\tilde{V}_S\tilde{V}_S^T - V_SV_S^T\right\|\right).$$

On the other hand, we have

$$\left\|\tilde{V}_S\tilde{V}_S^T - V_SV_S^T\right\|_\infty = \Omega\left(\frac{1}{n}\left\|\tilde{V}_S\tilde{V}_S^T - V_SV_S^T\right\|_F\right) = \Omega\left(\frac{1}{n}\left\|\tilde{V}_S\tilde{V}_S^T - V_SV_S^T\right\|.\right)$$

Therefore, our bound says that in the best case scenario, the largest entry of the matrix is of the same magnitude as the average one, making Eq. (14) *sharp. The sharpness (in the best case) of Eq.* (15) *and Eq.* (16) *can be argued similarly.*

Now, the question is: *How close can we come to the best-case scenarios for τ_1 and τ_2?* It turns out that when E is random and the entries of U (and respectively V) are sufficiently spread out, then, given the above lowr bounds, we can achieve very close upper bounds for τ_1 (and respectively τ_2), which are only a polylogartihmic term appart (see the discussion on the random setting below)

The general incoherence assumption. Let us point out that in Theorem 14, the incoherence assumption (hidden in the definition of τ_1, τ_2) is quite different from the incoherence assumption discussed in the introduction, concerning the matrix completion problem. The original incoherence assumption starts to appear when we consider E to be a random matrix. Assume, for simplicity, that E is n by n with iid entries $N(0,1)$, then it is easy to show that the matrix $EE^T/\|E\|^2$ is (with high probability) is close to the indentity matrix I_n. Thus, the quantity $\frac{\|(EE^T)U\|_{2,\infty}}{\|E\|^2}$ is small if the rows of U are short.

The random setting. Now we derive a corrollary which addresses the case when E is random with independent entries. Recall that $N = m+n$. In this case, the term $\|U^TEV\|_\infty$ is the maximum among sums of independent random variables, which is $O\big(\varsigma(\sqrt{\log N} + M\|U\|_\infty\|V\|_\infty \log N)\big)$ by the Bernstein bound Hoeffding (1963); Chernoff (1952).The term

y has been analyzed in Tran and Vu (2024) and mentioned above. We use the trivial upper bound $y \leq \|E\|^2 = O(\varsigma^2 N)$, which is enough to prove the next theorem.

Regarding τ_1 and τ_2, our analysis later will give the estimates

$$\tau_1 = O\left(\log N \sqrt{\frac{\mu(U)}{m}} + \frac{\log^{3/2} N}{\sqrt{N}} + \frac{M \log^3 N}{\sqrt{N}} \cdot \sqrt{\frac{\mu(V)}{n}}\right),$$

and symmetrically for τ_2 by swapping U and V. In the simple case where $\mu(U), \mu(V) = O(1)$, $m = \Theta(n)$, this estimate reduces to

$$\tau_1 = O\left(\log N \sqrt{\frac{\mu(U)}{m}}\right)$$

whenever $M = O(N^{1/2} \log^{-2} N)$, which is off by a factor $\log N$ from the ideal case.

Combining the observations above, we get the following "random" version of Theorem 14:

Theorem 17 *Consider the objects in Setting 13. Let $\varepsilon \in (0,1)$ be arbitrary Suppose E is a random $m \times n$ matrix with independent entries satisfying:*

$$\mathbf{E}\left[E_{ij}\right] = 0, \quad \mathbf{E}\left[|E_{ij}|^2\right] \leq \varsigma^2, \quad \mathbf{E}\left[|E_{ij}|^l\right] \leq M^{l-2}\varsigma^l \quad \text{for all } \; p \in \mathbb{N}_{\geq 2}, \tag{18}$$

where M and ς are parameters. Let $N = m + n$. Define

$$\tau_1 := \frac{\|U\|_{2,\infty} \log N}{\sqrt{r}} + \frac{M\|V\|_{2,\infty} \log^3 N}{\sqrt{rN}} + \frac{\log^{3/2} N}{\sqrt{N}}, \tag{19}$$

and define τ_2 symmetrically by swapping U and V. For an arbitrary subset $S \subset [r]$, suppose

$$\frac{\varsigma\sqrt{N}}{\sigma_S} \;\vee\; \frac{r\varsigma(\sqrt{\log N} + M\|U\|_\infty \|V\|_\infty \log N)}{\Delta_S} \;\vee\; \frac{\varsigma\sqrt{rN}}{\sqrt{\Delta_S \sigma_S}} \leq \frac{1}{16}. \tag{20}$$

Let

$$R_S := \frac{\varsigma\sqrt{N}}{\sigma_S} + \frac{r\varsigma(\sqrt{\log N} + M\|U\|_\infty \|V\|_\infty \log N)}{\Delta_S} + \frac{2r\varsigma^2 N}{\Delta_S \sigma_S}.$$

There are universal constants c and C such that: If $M \leq cN^{1/2}\log^{-5} N$, then with probability at least $1 - O(N^{-1})$,

$$\left\|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\right\|_\infty \leq C\tau_2^2 r R_S. \tag{21}$$

$$\left\|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\right\|_{2,\infty} \leq C\tau_2 r R_S. \tag{22}$$

Analogous bounds for U and $\tilde{U}$ hold, with τ_2 replacing τ_1.

When $S = [s]$ for some $s \in [r]$, we slightly abuse the notation to let

$$R_s := R_{[s]} = \frac{\varsigma\sqrt{N}}{\sigma_s} + \frac{r\varsigma(\sqrt{\log N} + M\|U\|_\infty \|V\|_\infty \log N)}{\delta_s} + \frac{2r\varsigma^2 N}{\delta_s \sigma_s}.$$

Then with probability $1 - O(N^{-1})$,

$$\|\tilde{A}_s - A_s\|_\infty \leq C\tau_1\tau_2 r\sigma_s R_s. \tag{23}$$

Furthermore, for each $\varepsilon > 0$*, if the term* $\frac{2r\varsigma^2 N}{\Delta_S \sigma_S}$ *in* R_S *is replaced with*

$$\frac{r}{\Delta_S\sigma_S} \inf\left\{t : \mathbf{P}\left(\max_{i\neq j}(|v_i E^T E v_j| + |u_i E E^T u_j|) \leq 2t\right) \geq 1-\varepsilon\right\},$$

then all three bounds above hold with probability at least $1 - \varepsilon - O(N^{-1})$.

Going back to the matrix completion setting, we can use this theorem to prove Theorem 10. The proof is a simple application of the theorem for the assumptions of Theorem 10, and will be in Appendix 2.

In the Appendices, we will provide the rigorous proofs for the two main theorems above, along with the proof of the matrix completion theorems 10 and 5. These proofs are technically heavy, with many technical lemmas whose lengthy proofs can distract from the overall flow. These more technical proofs will be in Appendix C.

References

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics*, 23:493–507, 1952. ISSN 0003-4851. doi: 10.1214/aoms/1177729330. URL https://doi.org/10.1214/aoms/1177729330.

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.

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.

László Erdös and Horng-Tzer Yau. *A Dynamical Approach to Random Matrix Theory*. Courant Institute of Mathematical Sciences, New York University, 2017.

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.

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.

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.

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.

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.

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

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

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.

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.

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.

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.

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.

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

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.

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.

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.

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.

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.

Appendix A. Proof of main results

As mentioned, Theorem 17 is a corollary of 14 when the noise matrix is random. In actuality, Theorem 14 is a slightly simplified version of the full argument for the deterministic case and does not directly lead to the random case. However, the reader can be assured that the changes needed to make Theorem 14 imply Theorem 17 are trivial, and will be discussed when we prove the latter.

Proof structure. First, we will assume Theorem 14 and use it to prove Theorem 17, which directly implies Theorem 10. The proof contains a novel high-probability *semi-isotropic* bound for powers of a random matrix, which can be of further independent interest.

We will then discard the random noise context and prove Theorem 14. The proof adapts the contour integral technique in Tran and Vu (2024), but with highly non-trivial adjustments to handle the inifnity norm, ínstead of spectral norm as in Tran and Vu (2024). The proof roughly has two steps:

1. Rewrite the quantities on the left-hand sides of the bounds in Theorem 14 as a power series in terms of E, similar to a Taylor expansion.
2. Devise a bound that decays exponentially for each power term, and sum them up as a geometric series to obtain a bound on the quantities of interest. The final bound, Lemma 25, will be general enough to imply all three of bounds of Theorem 14.

The structure for this section will be:

$$\text{Theorem 17} \xleftarrow{\text{implied by}} \text{Theorem 14} \xleftarrow{\text{implied by}} \text{Lemma 25}.$$

A.1. The random version: Proof of Theorem 17

In this section, we prove Theorem 17, assuming Theorem 14. First, consider the term

$$\frac{\|E\|}{\sigma_S} \vee \frac{2r\|U^T EV\|_\infty}{\Delta_S} \vee \frac{\sqrt{2r}\|E\|}{\sqrt{\sigma_S \Delta_S}}$$

from the condition (13). Let us replace the terms related to E in the above with their respective high-probability bounds.

- $\|E\|$. There are tight bounds in the literature. For E following the Model (18), with the assumption $M \le (m+n)^{1/2}\log^{-5}(m+n)$, the moment argument in Vu (2007) can be used.
- $\|U^T E V\|_\infty = \max_{i,j} |u_i^T E v_j|$. These terms can be bounded with a simple Bernstein bound.
- $y = \frac{1}{2}\max_{i\neq j}(|u_i E E^T u_j| + |v_i E^T E v_j|)$. The terms inside the maximum function can be bounded with the moment method. The most saving occurs when E is a stochastic matrix, meaning its row norms and column norms have the same second moment. For the purpose of proving Theorem 17, the naive bound $\|E\|^2$ suffices.

Upper-bounding these three is routine, which we summarize in the lemma below.

Lemma 18 *Consider the objects in Setting 13. Let $E \in \mathbb{R}^{m\times n}$ be a random matrix satisfying Model (18) with parameters M and ς. Suppose $M \le (m+n)^{1/2}\log^{-3}(m+n)$. Then with probability $1 - O((m+n)^{-2})$, all of the following hold:*

$$\|E\| \le 1.9\varsigma\sqrt{m+n} \le 2\varsigma\sqrt{m+n}, \tag{24}$$

$$\max_{i\neq j}(|u_i E E^T u_j| + |v_i E^T E v_j|) \le 2\|E\|^2 \le 8\varsigma^2(m+n). \tag{25}$$

$$\max_{i,j} |u_i^T E v_j| \le 2\varsigma(\sqrt{\log(m+n)} + M\|U\|_\infty\|V\|_\infty \log(m+n)). \tag{26}$$

Proof Eq. (24) follows from the moment argument in Vu (2007). Eq. (25) follows from Eq. (24). It remains to check Eq. (26). Fix $i, j \in [r]$. Write

$$u_i^T E v_j = \sum_{k\in[m], h\in[n]} u_{ik} v_{jh} E_{kh} = \sum_{(k,h)\in[m]\times[n]} Y_{kh},$$

where we temporarily let $Y_{kh} := u_{ik}v_{jh}E_{kh}$ for convenience. We have $|Y_{kh}| \le \|U\|_\infty\|V\|_\infty|E_{kh}|$. Let $X_{kh} := Y_{kh}/(\varsigma\|U\|_\infty\|V\|_\infty)$, then $\{X_{kh} : (k,h) \in [m]\times[n]\}$ are independent random variables and for each $(k,h) \in [m]\times[n]$,

$$\mathbf{E}[X_{kh}] = 0, \quad \mathbf{E}\left[|X_{kh}|^2\right] \le 1, \quad \mathbf{E}\left[|X_{kh}|^l\right] \le M^{l-2} \text{ for all } l \in \mathbb{N}.$$

We also have

$$\sum_{k,h} \mathbf{E}\left[|X_{kh}|^2\right] = \frac{\sum_{k,h} u_{ik}^2 v_{jh}^2 \mathbf{E}\left[|E_{kh}|^2\right]}{\varsigma^2\|U\|_\infty^2\|V\|_\infty^2} \le \frac{\varsigma^2\sum_{k,h} u_{ik}^2 v_{jh}^2}{\|U\|_\infty^2\|V\|_\infty^2} = \frac{1}{\|U\|_\infty^2\|V\|_\infty^2}$$

By Bernstein's inequality Chernoff (1952), we have for all $t > 0$

$$\mathbf{P}\left(\left|\sum_{k,h} X_{kh}\right| \ge t\right) \le \exp\left(\frac{-t^2}{\sum_{k,h}\mathbf{E}\left[|X_{kh}|^2\right] + \frac{2}{3}Mt}\right) \le \exp\left(\frac{-t^2}{\|U\|_\infty^{-2}\|V\|_\infty^{-2} + \frac{2}{3}Mt}\right).$$

We rescale $Y_{kh} = \varsigma\|U\|_\infty\|V\|_\infty X_{kh}$ and replace t with $t/(\varsigma\|U\|_\infty\|V\|_\infty)$, the above becomes

$$\mathbf{P}\left(\left|\sum_{k,h} Y_{kh}\right| \ge t\right) \le \exp\left(\frac{-t^2}{\varsigma^2 + \frac{2}{3}M\|U\|_\infty\|V\|_\infty t}\right).$$

Let $N = m + n$ and $t = 2\varsigma(\sqrt{\log N} + M\|U\|_\infty\|V\|_\infty \log N)$, we have

$$t^2 \geq 4\varsigma^2 \log N, \qquad t^2 \geq 2M\|U\|_\infty\|V\|_\infty t \log N,$$

thus

$$t^2 \geq \frac{12}{7}\left(\varsigma^2 + \frac{2}{3}M\|U\|_\infty\|V\|_\infty t\right)\log N.$$

Combining everything above, we get

$$\mathbf{P}\left(|u_i^T E v_j| \geq 2\varsigma(\sqrt{\log N} + M\|U\|_\infty\|V\|_\infty \log N)\right) \leq N^{-12/7}.$$

By a union bound over $(i,j) \in [r] \times [r]$, the proof of Eq. (26) and the lemma is complete. ■

Now all that remains is computing τ_1 and τ_2. More precisely, since both are random, we compute a good choice of high-probability upper bounds for them. This, however, is likely intractable since the appearance of powers of $\|E\|$ in the denominator makes it hard to analyze the right-hand sides of Eq. (12). To overcome this, notice that the argument in Theorem 14 works in the same way if, instead of being rigidly refined by Eq. (12), τ_1 and τ_2 are any real numbers satisfying

$$\begin{aligned}
\tau_1 &\geq \max_{a\in[[0,10\log(m+n)]]} \frac{1}{\sqrt{r}} \max\left\{\frac{\left\|(EE^T)^a U\right\|_{2,\infty}}{\mathcal{H}^{2a}}, \frac{\left\|(EE^T)^a EV\right\|_{2,\infty}}{\mathcal{H}^{2a+1}}\right\},\\
\tau_2 &\geq \max_{a\in[[0,10\log(m+n)]]} \frac{1}{\sqrt{r}} \max\left\{\frac{\left\|(E^TE)^a V\right\|_{2,\infty}}{\mathcal{H}^{2a}}, \frac{\left\|(E^TE)^a E^T U\right\|_{2,\infty}}{\mathcal{H}^{2a+1}}\right\},
\end{aligned} \tag{27}$$

for some upper bound $\mathcal{H} \geq \|E\|$.

From this point, we will discard Eq. (12) and treat $(\tau_1, \tau_2, \mathcal{H})$ as any tuple that satisfies Eq. (27). Specifically, we will choose $\tau_0(U)$, $\tau_1(U)$, $\tau_0(V)$, $\tau_1(V)$ such that

$$\forall a \in [[0, 10\log(m+n)]]: \ \tau_0(U) \geq \frac{1}{\sqrt{r}}\frac{\left\|(EE^T)^a U\right\|_{2,\infty}}{\mathcal{H}^{2a}}, \quad \tau_1(U) \geq \frac{1}{\sqrt{r}}\frac{\left\|(E^TE)^a E^T U\right\|_{2,\infty}}{\mathcal{H}^{2a+1}}$$

and symmetrically for $\tau_0(V)$ and $\tau_1(V)$, with E and E^T swapped. We can then simply let $\tau_1 = \tau_0(U) + \tau_1(V)$ and $\tau_2 = \tau_1(U) + \tau_0(V)$.

This is equivalent to bounding terms of the form

$$\|e_{m,k}^T (EE^T)^a U\|, \quad \|e_{m,k}^T (EE^T)^a EV\|, \quad \|e_{n,l}^T (E^TE)^a V\|, \quad \|e_{n,l}^T (E^TE)^a E^T U\|,$$

uniformly over all choices for $k \in [m]$, $l \in [n]$ and $0 \leq a \leq 10\log(m+n)$. We call them **semi-isotropic bounds** of powers of E, due to one side of them involving generic unit vectors (isotropic part) and the other side involving standard basis vectors.

To the best of our knowledge, there has been no well-known semi-isotropic bounds, entry bounds, or isotropic bounds of powers of a random matrix in the literature. The lemma below, which establishes such semi-isotropic bounds and gives values to $\tau_0(U)$, $\tau_1(U)$, $\tau_0(V)$, $\tau_1(V)$ and $\mathcal{H}$, is thus another noteworthy contribution of this paper and has potentials for further applications.

Lemma 19 *Let M and ς be positive real numbers and E be a $m \times n$ random matrix with independent entries following Model (18) with parameters M and ς. Let*

$$\mathcal{H} := 1.9\varsigma\sqrt{m+n}.$$

For each $p > 0$, define

$$\tau_0(U,p) := \frac{p\|U\|_{2,\infty}}{\sqrt{r}}, \quad \tau_1(U,p) := \frac{Mp^3\|U\|_{2,\infty}}{\sqrt{r(m+n)}} + \frac{p^{3/2}}{\sqrt{m+n}}. \tag{28}$$

There are universal constants C and c such that, for any $t > 0$, if $M \leq \frac{c\sqrt{m+n}}{t^2 \log^2(m+n)}$, then for each fixed $k \in [m]$, with probability $1 - O(\log^{-C}(m+n))$,

$$\max_{0 \leq \alpha \leq t\log(m+n)} \frac{\|e_{m,k}^T (EE^T)^a U\|}{\mathcal{H}^{2a}\sqrt{r}} \leq \tau_0(U, \log\log(m+n)) \tag{29}$$

for each fixed $k \in [n]$, with probability $1 - O(\log^{-C}(m+n))$,

$$\max_{0 \leq \alpha \leq t\log(m+n)} \frac{\|e_{n,k}^T (E^T E)^a E^T U\|}{\mathcal{H}^{2a+1}\sqrt{r}} \leq \tau_1(U, \log\log(m+n)) \tag{30}$$

If the stronger bound $M \leq \frac{c\sqrt{m+n}}{t^2 \log^5(m+n)}$ holds, then with probability $1 - O((m+n)^{-2})$,

$$\max_{0 \leq \alpha \leq t\log(m+n)} \max_{k \in [m]} \frac{\|e_{m,k}^T (EE^T)^a U\|}{\mathcal{H}^{2a}\sqrt{r}} \leq \tau_0(U, \log(m+n)), \tag{31}$$

$$\max_{0 \leq \alpha \leq t\log(m+n)} \max_{k \in [n]} \frac{\|e_{n,k}^T (E^T E)^a E^T U\|}{\mathcal{H}^{2a+1}\sqrt{r}} \leq \tau_1(U, \log(m+n)) \tag{32}$$

Analogous bounds hold for V, with E and E^T swapped.

We only use Eq. (31) and Eq. (32) to prove Theorem 17, but for completeness, we still include Eqs. (29) and (30), which have a better bound at the cost of being non-uniform in k. They may have potential for other applications.

To prove this lemma, we will use the moment method, with a walk-counting argument inspired by the coding scheme in Vu (2007), to bound these terms. We put the full proof in Section C.2.

Let us prove Theorem 17 using these lemmas.

Proof [Proof of Theorem 17] Consider the objects from Setting 13. We aim to apply Theorem 14. By Lemma 18, with probability $1 - O((m+n)^{-1})$, we can replace condition (13) in Theorem 14

$$\frac{\|E\|}{\sigma_S} \vee \frac{2r\|U^T EV\|_\infty}{\Delta_S} \vee \frac{\sqrt{2r}\|E\|}{\sqrt{\sigma_S \Delta_S}} \leq \frac{1}{8}$$

with condition (20) in Theorem 17

$$\frac{\varsigma\sqrt{N}}{\sigma_S} \vee \frac{r\varsigma(\sqrt{\log N} + M\|U\|_\infty \|V\|_\infty \log N)}{\Delta_S} \vee \frac{\varsigma\sqrt{rN}}{\sqrt{\Delta_S \sigma_S}} \leq \frac{1}{16}.$$

Assume (20) holds, then (13) also hold and we can now apply Theorem 14. Define

$$\tau_1 = \tau_0(U, \log(m+n)) + \tau_1(V, \log(m+n)), \quad \tau_2 = \tau_0(V, \log(m+n)) + \tau_1(U, \log(m+n)),$$

where $\tau_0(U,\cdot), \tau_1(U,\cdot)$ and $\tau_0(V,\cdot), \tau_1(V,\cdot)$ are from Lemma 19. These terms match exactly with τ_1 and τ_2 from the statement of Theorem 17. If they also matched τ_1 and τ_2 in Theorem 14, the proof would be complete. However, they do not.

Let $\mathcal{H} := 2\varsigma\sqrt{m+n}$, then $\mathcal{H} \geq \|E\|$ by Lemma 18. Per the discussion around the condition (27) above, if we can show that τ_1, τ_2 and $\mathcal{H}$ satisfy this condition, then the argument in Theorem 14 still works. By Lemma 19 for $t = 10$, (27) holds with probability $1 - O((m+n)^{-2})$, so the proof is complete. ■

In the next section, we prove Theorem 14. The proof is an adaptation of the main argument in Tran and Vu (2024) for the SVD. While this adaptation is easy, it has several important adjustments, sufficient to make Theorem 14 independent result rather than a simple corollary. For instance, the adjustment to adapt the argument for the infinity and 2-to-infinity norms necessitates the semi-isotropic bounds, a feature not required in the original results for the operator norm. For this reason, we present the proof in its entirety.

A.2. The deterministic version: Proof of Theorem 14

In this section, we provide the proof of Theorem 14.

Given A and $\tilde{A} = A + E$, there are three terms we need to bound, correspoding to Eqs. (14), (15) and (16):

$$\|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\|_\infty, \quad \|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\|_{2,\infty}, \quad \|\tilde{A}_s - A_s\|_\infty.$$

The strategy of bounding all three are almost identical, and is an extension to the SVD case of the strategy for the eigendecomposition in Tran and Vu (2024).

In fact, there are only two subtractions to analyze, namely $\tilde{V}_S\tilde{V}_S^T - V_S V_S^T$ and $\tilde{A}_s - A_s$. As an example, consider the former. If one views A and thus U and V as fixed, the above can be viewed as a function $f(E)$ satisfying $f(0) = 0$. The difficulty comes from the fact that we cannot (yet) express this function as an arithmetic combination of basic functions, which is often what is needed to analyze it in depth.

One basic idea to rewrite this function in a tractable form is to find a tractable form for the function $g : A \mapsto VV^T$, and write

$$\tilde{V}_S\tilde{V}_S^T - V_S V_S^T = g(\tilde{A}) - g(A) = g(A+E) - g(A).$$

If E is a square matrix (i.e. $m = n$) with some "favorable" properties, such as being a diagonal matrix, one can hope to rewrite the last expression as a Taylor series

$$\sum_{\gamma=1}^{\infty} \frac{g^{(\gamma)}(A)}{\gamma!} E^\gamma,$$

given the derivatives of g are well-defined at A. The crucial point is how to come up with the function g and an analogy for the Taylor series that works for a general matrix E. This is still hard, at first glance, since, just like f, g seems to be inexpressible in terms of simple functions.

The authors of Tran and Vu (2024) came up with a clever idea. Imagine first, for simplicity, that both A and E are square symmetric matrices, and that V and $\tilde{V}$ contain the eigenvector, rather than singular vectors, of their respective matrices. In other words, $U = V$ and the numbers σ_i are temporarily viewed as eigenvalues. Instead of measuring the difference $g(\tilde{A}) - g(A)$ directly, they considered the difference of the *Stieltjes transforms*, and obtained the expansion:

$$(zI - \tilde{A})^{-1} - (zI - A)^{-1} = \sum_{\gamma=1}^{\infty} [(zI - A)^{-1} E]^{\gamma} (zI - A)^{-1}. \tag{33}$$

It is easy to show that this identity hold whenever the right-hand side converges. Conveniently, the convergence is also guaranteed by the condition (13) of Theorem 14, as we will see later. To obtain $\tilde{V}_S \tilde{V}_S^T$ and $V_S V_S^T$, rewrite the left-hand side of Eq. (33) as

$$\sum_{i=1}^{n} \frac{\tilde{v}_i \tilde{v}_i^T}{z - \tilde{\sigma}_i} - \sum_{i=1}^{n} \frac{v_i v_i^T}{z - \sigma_i}.$$

If one can find a contour Γ_S that encircles precisely the set $\{\sigma_i, \tilde{\sigma}_i\}_{i \in S}$ while satisfying that the right-hand side of the expansion converges for every point on that contour, one will be able to integrate over Γ_S and obtain the power series expansion

$$\begin{aligned}
\tilde{V}_S \tilde{V}_S^T - V_S V_S^T &= \sum_{\gamma=1}^{\infty} \oint_{\Gamma_S} [(zI - A)^{-1} E]^{\gamma} (zI - A)^{-1} \\
&= \sum_{\gamma=1}^{\infty} \oint_{\Gamma_S} \left[\Big(\sum_{i \in [r]} \frac{v_i v_i^T}{z - \sigma_i} + \frac{I - VV^T}{z} \Big) E \right]^{\gamma} \Big(\sum_{i \in [r]} \frac{v_i v_i^T}{z - \sigma_i} + \frac{I - VV^T}{z} \Big).
\end{aligned}$$

The precise details on how to choose this contour can be found in Tran and Vu (2024). The final steps to bound the left-hand side will be:

1. Expand the right-hand side into sums involving products of E and $v_i v_i^T$ and $Q = I - VV^T$.

2. Bound each product by estimating the scalar contour integral and the norm of each factor.

Back to the context in this paper, where we handle the SVD instead of the eigendecomposition. In Tran and Vu (2024), the author used this expansion to obtain a bound on the spectral norm of the left-hand side by bounding each term in the series. We make appropriate adjustments to their argument to adapt it to the SVD, while also proving a novel *semi-isotropic* bound on powers of random matrices to extend the result to the infinity norm.

A.2.1. The power series expansion for the SVD case

Firstly, let us introduce the symmetrization trick, which translates the SVD into an eigendecomposition. If A has the SVD: $A = \sum_{i\in[r]} \sigma_i u_i^T v_i^T$, then we have the following eigendecomposition for the *symmetrized version* of A:

$$A_{\text{sym}} := \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix} = \sum_{i=1}^{r} \frac{1}{2}\sigma_i \left(\begin{bmatrix} u_i \\ v_i \end{bmatrix} \begin{bmatrix} u_i^T, & v_i^T \end{bmatrix} - \begin{bmatrix} u_i \\ -v_i \end{bmatrix} \begin{bmatrix} u_i^T, & -v_i^T \end{bmatrix} \right)$$

For each $i \in [r]$, let

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

The unit vectors $\{w_i : |i| \in [r]\}$ are orthogonal, and thus we can write

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

as an eigendecomposition of A_{sym}. We have

$$\begin{bmatrix} U_S U_S^T & 0 \\ 0 & V_S V_S^T \end{bmatrix} = \sum_{|i|\in S} w_i w_i^T.$$

Since the pair $(i, -i)$ always go together when we use A_{sym} to analyze A, we will use a different set of notation for A_{sym} and W, which supersede the conventional notation for spectral entities:

- W_S is the matrix whose columns are $\{w_i : |i| \in S\}$. Note that the conventional notation would just be $\{w_i : i \in S\}$.
- $(A_{\text{sym}})_S = W_S \Lambda W_S^T = \sum_{|i|\in S} \sigma_i w_i w_i^T$. This way $(A_{\text{sym}})_S = (A_S)_{\text{sym}}$. The conventional notation would only involve half of the sum.
- For $s \in [r]$, let $W_s := W_{[s]}$ and $(A_{\text{sym}})_s := (A_{\text{sym}})_{[s]}$. Technically, $(A_{\text{sym}})_s$ will then be the best rank-$2s$ approximation of A_{sym}, as opposed to the conventional meaning of the notation.
- For convenience, let $\sigma_0 := 0$.

We define $\tilde{\sigma}_i$, $\tilde{w}_i$ and $\tilde{W}_s$, $\tilde{W}_S$, $\tilde{A}_s$, $\tilde{A}_S$ similarly for $\tilde{A} = A + E$. From Eq. (33) for the symmetric case, we have the expansion

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

which is equivalent to

$$\sum_i \frac{\tilde{w}_i \tilde{w}_i^T}{z - \tilde{\sigma}_i} - \sum_{|i|\in[r]} \frac{w_i w_i^T}{z - \sigma_i} = \sum_{\gamma=1}^{\infty} \left[(zI - A_{\text{sym}})^{-1} E_{\text{sym}} \right]^{\gamma} (zI - A_{\text{sym}})^{-1}. \tag{34}$$

Let Γ_S denote a contour in $\mathbb{C}$ that encircles $\{\pm\sigma_i, \pm\tilde{\sigma}_i\}_{i\in S}$ and none of the other eigenvalues of $\tilde{W}$ and W, satisying that the right-hand side of Eq. (34) converges for every z on the contour. Integrating over Γ_S of both sides and dividing by $2\pi i$, we have

$$\begin{aligned}\begin{bmatrix}\tilde{U}_S\tilde{U}_S^T - U_S U_S^T & 0 \\ 0 & \tilde{V}_S\tilde{V}_S^T - V_S V_S^T\end{bmatrix} &= \tilde{W}_S\tilde{W}_S - W_S W_S^T \\ &= \sum_{\gamma=1}^{\infty}\oint_{\Gamma_S}\frac{\mathrm{d}z}{2\pi i}\left[(zI - A_{\text{sym}})^{-1}E_{\text{sym}}\right]^{\gamma}(zI - A_{\text{sym}})^{-1}.\end{aligned} \tag{35}$$

Suppose one aims to bound $\|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\|_\infty$. The simplest approach is to fix two entries $j, k \in [n]$ and obtain a bound for the jk-entry that holds regardless of j and k. Noting that

$$(\tilde{V}_S\tilde{V}_S^T - V_S V_S^T)_{jk} = (\tilde{W}_S\tilde{W}_S^T - W_S W_S^T)_{(j+m)(k+m)},$$

we have the expansion

$$(\tilde{V}_S\tilde{V}_S^T - V_S V_S^T)_{jk} = \sum_{\gamma=1}^{\infty}\oint_{\Gamma_S}\frac{\mathrm{d}z}{2\pi i}e_{m+n,m+j}^T\left[(zI - A_{\text{sym}})^{-1}E_{\text{sym}}\right]^{\gamma}(zI - A_{\text{sym}})^{-1}e_{m+n,m+k}, \tag{36}$$

where $e_{N,l}$ denotes the l^{th} standard basis vector in N dimensions.

From this point onwards, our proof diverges from the argument in Tran and Vu (2024). The goal is still the same, but our expansion will be different from Tran and Vu (2024), with the goal of creating powers of E_{sym}, rather than alternating products like $E_{\text{sym}}QE_{\text{sym}}Q\ldots E_{\text{sym}}$. To ease the notation, we denote

$$P_i := w_i w_i^T, \quad \text{for } i = \pm 1, \pm 2, \ldots, \pm r.$$

The resolvent of A_{sym}, which is a function of a complex variable z, can now be written as:

$$(zI - A_{\text{sym}})^{-1} = \sum_{|i|\in[r]}\frac{P_i}{z-\sigma_i} + \frac{I - \sum_{|i|\in[r]}P_i}{z} = \sum_{|i|\in[r]}\frac{\sigma_i P_i}{z(z-\sigma_i)} + \frac{I}{z}.$$

Plugging into Eq. (36), the term with power γ becomes

$$\oint_{\Gamma_S}\frac{\mathrm{d}z}{2\pi i}e_{m+n,m+j}^T\left[\Big(\sum_{|i|\in[r]}\frac{\sigma_i P_i}{z(z-\sigma_i)} + \frac{I}{z}\Big)E_{\text{sym}}\right]^{\gamma}\Big(\sum_{|i|\in[r]}\frac{\sigma_i P_i}{z(z-\sigma_i)} + \frac{I}{z}\Big)e_{m+n,m+k}. \tag{37}$$

When expanding the above, we get monomials of the form

$$\begin{aligned}&\oint_{\Gamma_S}\frac{\mathrm{d}z}{2\pi i}e_{m+n,m+j}^T\Big(\underbrace{\frac{I}{z}E_{\text{sym}}\ldots\frac{I}{z}E_{\text{sym}}}_{\alpha_0\text{ times}}\Big)\Big(\underbrace{\frac{\sigma_? P_?}{z(z-\sigma_?)}E_{\text{sym}}\ldots\frac{\sigma_? P_?}{z(z-\sigma_?)}E_{\text{sym}}}_{\beta_1\text{ times}}\Big) \\ &\ldots\Big(\underbrace{\frac{\sigma_? P_?}{z(z-\sigma_?)}E_{\text{sym}}\ldots E_{\text{sym}}\frac{\sigma_? P_?}{z(z-\sigma_?)}}_{(\beta_h-1)\ E_{\text{sym}}\text{ factors}}\Big)\Big(\underbrace{E_{\text{sym}}\frac{I}{z}\ldots E_{\text{sym}}\frac{I}{z}}_{\alpha_h\text{ times}}\Big)e_{m+n,m+k},\end{aligned}$$

where the question marks stand for different indices i's. Rearranging, we get the form

$$\begin{aligned}&\left[\oint_{\Gamma_S}\frac{\mathrm{d}z}{2\pi i}\frac{1}{z^{\alpha_0+\beta_0+\alpha_1+\ldots+\beta_{h-1}+\alpha_h}}\overbrace{\frac{\sigma_?}{z-\sigma_?}\frac{\sigma_?}{z-\sigma_?}\cdots\frac{\sigma_?}{z-\sigma_?}}^{\beta_1+\beta_2+\ldots+\beta_h\text{ factors}}\right]\\&e^T_{m+n,m+j}E^{\alpha_0}_{\text{sym}}\Big(\underbrace{P_?E_{\text{sym}}P_?E_{\text{sym}}\ldots P_?E_{\text{sym}}}_{\beta_1\text{ factors}}\Big)E^{\alpha_1}_{\text{sym}}\ldots\Big(\underbrace{P_?E_{\text{sym}}P_?E_{\text{sym}}\ldots P_?}_{(\beta_h-1)\ E_{\text{sym}}\text{ factors}}\Big)E^{\alpha_h}_{\text{sym}}e_{m+n,m+k},\end{aligned}\tag{38}$$

At this point, one can see how several terms in Theorem 14, especially the incoherence parameters τ and τ', appear in the final bounds. The long matrix product can be rearranged as

$$\begin{aligned}&\left(e^T_{m+n,m+j}E^{\alpha_0}_{\text{sym}}w_?\right)\Big(\underbrace{w_?^TE_{\text{sym}}w_?\ldots w_?^TE_{\text{sym}}w_?}_{(\beta_1-1)\ E_{\text{sym}}\text{ factors}}\Big)\left(w_?^TE^{\alpha_1+1}_{\text{sym}}w_?\right)\\&\ldots\ \left(w_?^TE^{\alpha_{h-1}+1}_{\text{sym}}w_?\right)\Big(\underbrace{w_?^TE_{\text{sym}}w_?\ldots w_?^TE_{\text{sym}}w_?}_{(\beta_h-1)\ E_{\text{sym}}\text{ factors}}\Big)\left(w_?^TE^{\alpha_h}_{\text{sym}}e_{m+n,m+k}\right)\end{aligned}\tag{39}$$

As a sneak peek of the proof:

- The two terms at the beginning and ending of the product give rise to τ and τ'.
- The terms $w_?^TE_{\text{sym}}w_?$ give rise to the term $\|U^TEV\|_\infty$ in Eq. (13).
- The terms $w_?^TE^{\alpha_i+1}_{\text{sym}}w_?$ mostly give rise to the term $\|E\|$, but in the special cases where $\alpha_i=1$ for all i will be more strongly bounded with the term y in R_3.

To further analyze these products and their sum and turn this argument into the proof, we need to formalize them with proper notation.

A.2.2. Notation and roadmap

Setting 20 *The following list also summarizes the notation used in the proof.*

- *For all matrices B, define* $B_{\text{sym}}:=\begin{pmatrix}0&B\\B^T&0\end{pmatrix}$.
- *Consider A. For each $i\in[r]$, let*

$$\sigma_{-i}=-\sigma_i,\ u_{-i}=-u_i,\ v_{-i}=-v_i,\quad\textit{and}\quad w_i=\frac{1}{\sqrt{2}}\begin{bmatrix}u_i\\v_i\end{bmatrix}$$

  Define $\mathbf{\Lambda}:=\{\sigma_i\}_{i\in[[-r,r]]}$ *(which includes $\sigma_0=0$) and* $W:=\left[w_i\right]_{i\in[\pm r]}$, *where* $[\pm r]:=\{i:|i|\in[r]\}$ *(which does not include 0).*
- *Define $\tilde{w}_i$ similarly, with* $\operatorname{rank}\tilde{A}$ *instead of r.*
- *Let $e_{N,k}$ be the k^{th} vector of the standard basis in $\mathbb{R}^N$.*
- *Γ_S is a contour encircling precisely the set $\{\sigma_i,\tilde{\sigma}_i:|i|\in S\}$ and no other eigenvalues, such that the right-hand side of Eq. (34) converges absolutely for all z on it.*

- *For each h, let $\Pi_h(\gamma)$ be the set of all pairs of $\boldsymbol{\alpha} = [\alpha_k]_{k=0}^h$ and $\boldsymbol{\beta} = [\beta_k]_{k=1}^h$ such that:*

$$\begin{aligned} &\bullet \quad \alpha_0, \alpha_h \geq 0, \quad \text{and } \alpha_k \geq 1 \text{ for } 1 \leq k \leq h-1, \\ &\bullet \quad \beta_k \geq 1 \text{ for } 1 \leq k \leq h, \\ &\bullet \quad \alpha + \beta = \gamma + 1, \quad \text{where } \alpha := \textstyle\sum_{k=0}^h \alpha_k, \quad \text{and } \beta := \textstyle\sum_{k=1}^h \beta_k. \end{aligned} \tag{40}$$

  Note that the conditions above imply $2h - 1 \leq \gamma + 1$, so the maximum value for h is $\lfloor \gamma/2 \rfloor + 1$.

- *For each $\boldsymbol{\beta}$ above satisfying each $\beta_k \geq 1$, we use $\mathbf{I} = [i_1, i_2, \ldots, i_\beta]$ for an element of $[\pm r]^\beta$. Together, the triple $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{I})$ define uniquely a monomial of the form (38). Define $\mathbf{I}_{a:b}$ as the subsequence $[i_a, i_{a+1}, \ldots, i_b]$.*

- *For each $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \Pi_h(\gamma)$ and $\mathbf{I} \in [\pm r]^\beta$, define*

$$\mathcal{C}(\mathbf{I}) := \oint_{\Gamma_S} \frac{\mathrm{d}z}{2\pi i} \frac{1}{z^{\gamma+1}} \prod_{j \in [\beta]} \frac{\sigma_{i_j}}{z - \sigma_{i_j}},$$

$$\mathcal{M}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{I}) := E_{\text{sym}}^{\alpha_0} \Big(\prod_{j=1}^{\beta_1} P_{i_j} E_{\text{sym}} \Big) E_{\text{sym}}^{\alpha_1} \ldots \Big(\prod_{j=\beta_1+\ldots+\beta_{h-1}}^{\beta-1} P_{i_j} E_{\text{sym}} \Big) P_{i_{\beta-1}} E_{\text{sym}}^{\alpha_h},$$

  where $P_i := w_i w_i^T$ for each $i \in [\pm r]$. We call the first, scalar, term the integral coefficient *and the second the* monomial matrix.

- *Define the following terms:*

$$\mathcal{T}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{\mathbf{I} \in [\pm r]^\beta} \mathcal{C}(\mathbf{I}) \mathcal{M}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{I}), \qquad \mathcal{T}^{(\gamma,h)} = \sum_{(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \Pi_h(\gamma)} \mathcal{T}(\boldsymbol{\alpha}, \boldsymbol{\beta}),$$

$$\mathcal{T}^{(\gamma)} = \sum_{h=1}^{\lfloor \gamma/2 \rfloor + 1} \mathcal{T}^{(\gamma,h)}, \qquad \mathcal{T} = \sum_{\gamma \geq 1} \mathcal{T}^{(\gamma)}.$$

From Eqs. (36), (37) and (38), we have

$$(\tilde{V}_S \tilde{V}_S^T - V_S V_S^T)_{jk} = e_{m+n,m+j}^T \mathcal{T} e_{m+n,m+k}. \tag{41}$$

At this point, we look at the larger context of Theorem 14. Consider Eq. (15). To bound $\|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\|_{2,\infty}$, we can fix one index j and find a bound for its j^{th} row that holds with probability close enough to 1 to beat the n factor from the union bound. We have

$$(\tilde{V}_S \tilde{V}_S^T - V_S V_S^T)_{j,\cdot} = e_{m+n,m+j}^T \mathcal{T}. \tag{42}$$

Therefore, we will introduce a Lemma to bound $M^T \mathcal{T} M'$ for generic matrices M and M' (both with $m + n$ rows), and apply it to obtain both Eq. (14) and (15).

Finally, consider Eq. (16). Following the same train of thought, we want to bound the (j, k)-entry of $\tilde{A}_s - A_s$ for a fixed $j \in [m]$ and $k \in [n]$. The series $\mathcal{T}$ as defined in Setting 20 will not be directly helpful here. Instead, we will modify it slightly, particularly at the integral coefficient, to obtain the power series for $(\tilde{A}_s - A_s)_{jk}$. The remaining steps will be identical to the proofs of (14) and (15). The details will be given later, when we prove (16).

A.2.3. Bounding the change in singular subspace expansions

Let us prove Eqs. (14) and (15) here. We aim to upper bound $\|M^T\mathcal{T}M'\|$, with $\|\cdot\|$ being the spectral norm, which generalizes both the absolute value of a scalar and the L2 norm of a vector. In fact, the proof works for any sub-multiplicative norm that is invariant under transposition. We can plug in different choices for M and M' to obtain (14) and (15).

We start off with bounds on the integral coefficient and the monomial matrix.

Lemma 21 (Bound on integral coefficients) *Consider the objects defined in Setting 20. Let $\mathbf{I} = \{i_k : k \in \beta\} \in [\pm r]^\beta$ and denote the following:*

$$\sigma_S(\mathbf{I}) := \min\{|\sigma_{i_k}| : |i_k| \in S\},$$
$$\Delta_S(\mathbf{I}) := \min\{|\sigma_{i_k} - \sigma_{i_l}| : |i_k| \in S, |i_l| \notin S\}.$$

We have,

$$|\mathcal{C}(\mathbf{I})| \le \frac{2^{\gamma+\beta-1}}{\sigma_S(\mathbf{I})^{\gamma+1-\beta}\Delta_S(\mathbf{I})^{\beta-1}} \le \frac{2^{\gamma+\beta-1}}{\sigma_S^{\gamma+1-\beta}\Delta_S^{\beta-1}}. \tag{43}$$

In the steps that follow, we will mainly use the second bound of Eq. (43), with one exception where the first, more precise, bound is needed. It thus makes sense to keep both.

Lemma 22 (Bound on monomial matrices) *Consider the objects defined in Setting 20. Fix γ, h and $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \Pi_h(\gamma)$ and $\mathbf{I} = \{i_k : k \in \beta\} \in [\pm r]^\beta$. Then*

$$\|M^T\mathcal{M}(\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{I})M'\| \le \|E\|^{\alpha-\alpha_0-\alpha_h+h-1} \cdot \|W^T E_{\text{sym}} W\|_\infty^{\beta-h} \cdot \|w_{i_1}^T E_{\text{sym}}^{\alpha_0} M\| \cdot \|w_{i_\beta}^T E_{\text{sym}}^{\alpha_h} M'\|. \tag{44}$$

Assuming both bounds above hold, we have the following bounds for each level in the sum $M^T\mathcal{T}M'$. The first is a bound on $M^T\mathcal{T}_\nu(\boldsymbol{\alpha},\boldsymbol{\beta})M'$.

Lemma 23 (Bound on $\mathcal{T}(\boldsymbol{\alpha},\boldsymbol{\beta})$) *Consider objects in Setting 20. Fix γ and h such that $1 \le h \le \gamma/2 + 1$, and $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \Pi_h(\gamma)$, and define the following terms*

$$\tau(M) = \max_{0 \le \alpha \le 10\log(m+n)} \frac{1}{2r} \sum_{|i| \in [r]} \frac{\|w_i^T E_{\text{sym}}^\alpha M\|}{\|E_{\text{sym}}\|^\alpha}, \qquad \text{and analogously for } \tau(M'). \tag{45}$$

$$R_1 := \frac{\|E\|}{\sigma_S} \vee \frac{2r\|W^T E_{\text{sym}} W\|_\infty}{\Delta_S}, \quad R_2 := \frac{\sqrt{2r}\|E\|}{\sqrt{\sigma_S \Delta_S}}, \quad R_3 := \frac{2r\max_{|i|\neq|j|}|w_i E_{\text{sym}}^2 w_j|}{\sigma_S \Delta_S} \tag{46}$$

and assume that

$$R := R_1 \vee R_2 < 1/4.$$

Suppose that $1 \le \gamma \le 10\log(m+n)$. We have

$$\|M^T\mathcal{T}(\boldsymbol{\alpha},\boldsymbol{\beta})M'\| \le \begin{cases} r\tau(M)\tau(M')2^{\gamma+\beta}R_1R^{\gamma-1} & \text{if } 1 \le h < \gamma/2+1, \\ 16r\tau(M)\tau(M')(4R)^{\gamma-2}(R_3 + R_1^2) & \text{if } h = \gamma/2+1. \end{cases} \tag{47}$$

When $10\log(m+n) < \gamma$, an analogous version of the above holds with $\|M\|$ and $\|M'\|$ replacing $\tau(M)$ and $\tau(M')$, respectively.

Summing up the bounds above over all $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \Pi_h(\gamma)$ and all $1 \le h \le \gamma/2 + 1$, we get the following lemma.

Lemma 24 (Bound on each power term in $\mathcal{T}$) *Consider the objects in Setting 20 and R, R_1, R_2 and R_3 from Lemma 23. For each $1 \le \gamma \le 10\log(m+n)$, we have*

$$\left\| M^T \mathcal{T}^{(\gamma)} M' \right\| \le r\tau(M)\tau(M') \left[9R_1(6R)^{\gamma-1} + \mathbf{1}\{\gamma \text{ even}\} \cdot 16(4R)^{\gamma-2}(R_3 + R_1^2) \right].$$

When $10\log(m+n) < \gamma$, an analogous version of the above holds with $\|M\|$ and $\|M'\|$ replacing $\tau(M)$ and $\tau(M')$, respectively.

Summing up the bounds above over all $\gamma \ge 1$, we get the final bound for the power series:

Lemma 25 (Bound on the whole $\mathcal{T}$) *Consider the objects in Setting 20 and R, R_1, R_2 and R_3 from Lemma 23. Suppose $R \le 1/4$. Then the $\mathcal{T}$ converges in the metric $\|\cdot\|$ and satisfies, for a universal constant C,*

$$\left\| M^T \mathcal{T} M' \right\| \le Cr\Big[\tau(M)\tau(M') + \|M\|\|M'\|(m+n)^{-2.5} \Big](R_1 + R_3).$$

Let us remark on the meanings of the new terms, which are simply translation of terms from Theorem 14 into the language of Setting 20.

- The term $\|M\|\|M'\|(m+n)^{-2.5}$ is small, and will be absorbed into the term $\tau(M)\tau(M')$ for our applications.
- When translating back from the symmetric setting with A_{sym} and W back to A and U, V, the terms R, R_1, R_2 and R_3 satisfy

$$R = R_1 \vee R_2 = \frac{\|E\|}{\sigma_S} \vee \frac{2r\|U^T E V\|_\infty}{\Delta_S} \vee \frac{\sqrt{2r}\|E\|}{\sqrt{\sigma_S \Delta_S}},$$

  and

$$R_1 + R_3 \le 2\left(\frac{\|E\|}{\sigma_S} + \frac{r\|U^T E V\|_\infty}{\Delta_S} + \frac{ry}{\Delta_S \sigma_S} \right).$$

- Similarly, recall the definitions of τ_1 and τ_2 in Eq. (12). As a function of M, τ satisfies

$$\tau(e_{m+n,k}) = \tau_1 \text{ for } k \le m, \quad \tau(e_{m+n,k}) = \tau_2 \text{ for } m+1 \le k \le m+n, \quad \text{and} \quad \tau(I) \le 1, \tag{48}$$

To summarize, the logical structure is:

Lemma 23 $\xrightarrow{\text{implies}}$ Lemma 24 $\xrightarrow{\text{implies}}$ Lemma 25 $\xrightarrow{\text{implies}}$ Eqs. (14), (15) in Theorem 14

We will finish the last step, which is the proof of (14) and (15) here. The proofs of Lemmas 23, 24 and 25 will be postpone to Section A.3.

Proof [Proof of Theorem 14 part I] Consider the objects defined in Theorem 14 and the additional objects in Setting 20. By the remark above, the condition (13) in Theorem 14 is equivalent to $R_1 \vee R_2 \leq 1/4$, so we can apply the lemmas in this section.

Let us prove Eq. (14). Consider arbitrary $j, k \in [n]$. From Eq. (41), $(\tilde{V}_S\tilde{V}_S^T - V_SV_S^T)_{jk}$ is $M^T\mathcal{T}M'$ for $M = e_{m+n,j+m}$ and $M' = e_{m+n,k+m}$. We apply the bound Lemma 25, while replacing both $\tau(M)$ and $\tau(M')$ with τ_2 (permissible by Eq. (48)), to get

$$\left|(\tilde{V}_S\tilde{V}_S^T - V_SV_S^T)_{jk}\right| \leq Cr(R_1+R_3)\left(\tau_2^2 + \frac{\|M\|\|M'\|}{(m+n)^{2.5}}\right) \leq 3Cr\tau_2^2(R_1+R_3),$$

where the last inequality is due to the facts $\|M\| = \|M'\| = 1$ and $\tau_1, \tau_2 \geq (m+n)^{-1/2}$. This holds over all $j, k \in [n]$, so it extends to the infinity norm, proving Eq. (14).

Let us prove Eq. (15). Consider an arbitrary $j \in [n]$. By Eq. (42), $(\tilde{V}_S\tilde{V}_S^T - V_SV_S^T)_{j,\cdot} = M^T\mathcal{T}M'$ for the choices $M = e_{m+n,j+m}$ and $M' = I_{m+n}$. We repeat the previous calculations, but this time Eq. (48) tells us to replace $\tau(M)$ with τ_2 and $\tau(M')$ with 1, to get

$$\left\|(\tilde{V}_S\tilde{V}_S^T - V_SV_S^T)_{j,\cdot}\right\| \leq 3Cr\tau_2(R_1+R_3),$$

which holds uniformly over $j \in [n]$, proving Eq. (15). ■

Next, we will finish proving Theorem 14 by proving Eq. (16). The argument is identical, but there is a small but important change in the integral coefficient, enough to separate the proof into the next part.

A.2.4. Bounding the change in low rank approximations

Throughout this part, we assume $S = [s]$ for a fixed $s \in [r]$. Consider Eq. (34) again. We already know that integrating both sides gives $\tilde{W}_s\tilde{W}_s^T - W_sW_s^T$ on the left-hand side. Since we are aiming to bound $\tilde{A}_s - A_s$, we need $\tilde{W}_s\tilde{\Lambda}_s\tilde{W}_s^T - W_s\Lambda_sW_s^T$ on the left-hand side instead. This can be achieved by multiplying both sides with z before integrating, taking advantage of the fact

$$\oint_\Gamma \frac{z\mathrm{d}z}{z-\sigma} = \sigma$$

for every contour Γ encircling σ. Therefore, the analogy of Eq. (35) is

$$\begin{aligned}
(\tilde{A}_s - A_s)_{\text{sym}} &= (\tilde{A}_{\text{sym}})_s - (A_{\text{sym}})_s = \sum_{|i|\in[s]}\left(\frac{z\tilde{w}_i\tilde{w}_i^T}{z-\tilde{\sigma}_i} - \frac{zw_iw_i^T}{z-\sigma_i}\right) \\
&= \sum_{\gamma=1}^{\infty}\oint_{\Gamma_s}\frac{z\mathrm{d}z}{2\pi i}\left[(zI - A_{\text{sym}})^{-1}E_{\text{sym}}\right]^\gamma (zI - A_{\text{sym}})^{-1}. \qquad (49)\\
&= \sum_{\gamma=1}^{\infty}\oint_{\Gamma_s}\frac{z\mathrm{d}z}{2\pi i}\left[\Big(\sum_{|i|\in[r]}\frac{\sigma_iP_i}{z(z-\sigma_i)} + \frac{I}{z}\Big)E_{\text{sym}}\right]^\gamma\Big(\sum_{|i|\in[r]}\frac{\sigma_iP_i}{z(z-\sigma_i)} + \frac{I}{z}\Big).
\end{aligned}$$

Therefore, we can replace the integral coefficient $\mathcal{C}(\mathbf{I})$ from Setting 20 with

$$\mathcal{C}_1(\mathbf{I}) := \oint_{\Gamma_S} \frac{z\mathrm{d}z}{2\pi i} \frac{1}{z^{\gamma+1}} \prod_{j\in[\beta]} \frac{\sigma_{i_j}}{z-\sigma_{i_j}}. \tag{50}$$

Respectively define $\mathcal{M}_1(\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{I})$, $\mathcal{T}_1(\boldsymbol{\alpha},\boldsymbol{\beta})$, $\mathcal{T}_1^{(\gamma,h)}$, $\mathcal{T}_1^{(h)}$ and $\mathcal{T}_1$ analogously to $\mathcal{M}(\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{I})$, $\mathcal{T}(\boldsymbol{\alpha},\boldsymbol{\beta})$, $\mathcal{T}^{(\gamma,h)}$, $\mathcal{T}^{(h)}$ and $\mathcal{T}$ from Setting 20.

The only piece we need to modify in the proofs of Eqs. (14) and (15) is the integral coefficient bound, namely Lemma 21. We have this bound for $\mathcal{C}_1(\mathbf{I})$:

Lemma 26 (Bound on integral coefficients) *Consider the objects in Setting 20 and Lemma 21 and $\mathcal{C}_1$ defined in Eq. (50). We have,*

$$|\mathcal{C}_1(\mathbf{I})| \le \frac{2^{\gamma+\beta-2}}{\sigma_s(\mathbf{I})^{\gamma-\beta}\Delta_s(\mathbf{I})^{\beta-1}} \le \frac{2^{\gamma+\beta-2}}{\sigma_s^{\gamma-\beta}\Delta_s^{\beta-1}} = \frac{\sigma_s}{2}\cdot\frac{2^{\gamma+\beta-1}}{\sigma_s^{\gamma+1-\beta}\Delta_s^{\beta-1}}. \tag{51}$$

The purpose of the last transformation is to highlight that the bound on the new integral coefficient is simply scaled up by a factor $\sigma_s/2$ compared to the old bound.

We remark that this bound does not hold for all choices of β if the power of z in Eq. (50) is larger than 1, or when S does not contain exactly the first s singular values. Therefore, one can neither extend Eq. (16) to a general S nor to quantities like $\tilde{A}_s^2 - A_s^2$, at least not in a simple way.

Proof [Proof of Theorem 14 part II] We prove Eq. (16). Fix $j \in [m]$ and $k \in [n]$. By Eq. (49), $(\tilde{A}_s - A_s)_{jk} = M^T\mathcal{T}_1 M'$ for $M = e_{m+n,j}$ and $M' = e_{m+n,m+k}$. The bound on $M^T\mathcal{T}_1M'$ will simply be the same bound for $M^T\mathcal{T}M'$ scaled up by $\sigma_s/2$. By Eq. (48), we can also replace $\tau(M)$ with τ_1 and $\tau(M')$ with τ_2. Therefore we obtain

$$\left|(\tilde{A}_s - A_s)_{jk}\right| \le Cr(R_1+R_3)\left(\tau_1\tau_2 + \frac{\|M\|\|M'\|}{(m+n)^{2.5}}\right) \le 3Cr\tau_1\tau_2(R_1+R_3),$$

where the last inequality holds due to $\tau_1, \tau_2 \ge (m+n)^{-1/2}$ and $\|M\| = \|M'\| = 1$. After passing to the inifinity norm, the proofs of Eq. (16) and of Theorem 14 are complete. ■

Now it remains to prove the lemmas in Sections A.2.3 and A.2.4. We will prove Lemmas 22, 23, 24 and 25. The proofs of the bounds on the integral coefficients (Lemmas 21 and 26) will be postponed to Section C due to their lengths.

A.3. Bounding the generic series

Let us prove Lemma 22.

Proof [Proof of Lemma 22] Consider a monomial matrix $\mathcal{M}(\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{I})$ has the form

$$\mathcal{M}(\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{I}) := E_{\text{sym}}^{\alpha_0}\Big(\prod_{j=1}^{\beta_1} P_{i_j}E_{\text{sym}}\Big)E_{\text{sym}}^{\alpha_1}\dots\Big(\prod_{j=\beta-\beta_h}^{\beta-1} P_{i_j}E_{\text{sym}}\Big)P_{i_{\beta-1}}E_{\text{sym}}^{\alpha_h}. \tag{52}$$

From Eq. (39), we can rearrange this to get

$$M^T\mathcal{M}(\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{I})M' = \left(M^T E_{\text{sym}}^{\alpha_0} w_{i_1}\right)\left(\prod_{j=1}^{\beta_1-1} w_{i_j}^T E_{\text{sym}} w_{i_{j+1}}\right)\left(w_{i_{\beta_1}}^T E_{\text{sym}}^{\alpha_1+1} w_{i_{\beta_1+1}}\right)$$
$$\dots \ \left(w_{i_{\beta-\beta_h}}^T E_{\text{sym}}^{\alpha_{h-1}+1} w_{i_{\beta-\beta_h+1}}\right)\left(\prod_{j=\beta-\beta_h+1}^{\beta-1} w_{i_j}^T E_{\text{sym}} w_{i_{j+1}}\right)\left(w_{i_\beta}^T E_{\text{sym}}^{\alpha_h} M'\right)$$

Let us break down this product into the following types:

1. $M^T E_{\text{sym}}^{\alpha_0} w_{i_1}$ and $w_{i_\beta}^T E_{\text{sym}}^{\alpha_h} M'$: bounded by their respective norms.
2. $w_{i_j}^T E_{\text{sym}} w_{i_{j+1}}$ for each $j \in [\beta-1]$: bounded by $\|W^T E_{\text{sym}} W\|_\infty$, and their number is $(\beta_1-1)+(\beta_2-1)+\ldots+(\beta_h-1)=\beta-h$.
3. $w_{i_j}^T E_{\text{sym}}^{\alpha+1} w_{i_{j+1}}$ for $j=\beta_1+\ldots+\beta_l$ and $\alpha=\alpha_l$ for some l: bounded by $\|E\|^{\alpha_l+1}$, and their total power is $(\alpha_1+1)+(\alpha_2+1)+\ldots+(\alpha_{h-1}+1)=\alpha-\alpha_0-\alpha_h+h-1$.

Due to the fact $\|\cdot\|$ is sub-multiplicative, the proof is complete. ■

We continue with proving Lemma 23.

Proof [Proof of Lemma 23] For simplicity, let $X = W^T E_{\text{sym}} W$. Since $\mathcal{T}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{\mathbf{I}\in[2r]^\beta}\mathcal{C}(\mathbf{I})\mathcal{M}(\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{I})$, we obtain

$$\|M^T\mathcal{T}(\boldsymbol{\alpha},\boldsymbol{\beta})M'\| \le \|X\|_\infty^{\beta-h}\|E\|^{\alpha-\alpha_0-\alpha_h+h-1}\sum_{\mathbf{I}\in[\pm r]^\beta}|\mathcal{C}(\mathbf{I})|\left\|M^T E_{\text{sym}}^{\alpha_0} w_{i_1}\right\|\left\|w_{i_{h\beta_h}}^T E_{\text{sym}}^{\alpha_h} M'\right\|.$$

Applying the second part of the bound (43) on $\mathcal{C}(\mathbf{I})$ in Lemma 21, we get

$$\begin{aligned}
\|M^T\mathcal{T}(\boldsymbol{\alpha},\boldsymbol{\beta})M'\| &\le \|X\|_\infty^{\beta-h}\|E\|^{\alpha-\alpha_0-\alpha_h+h-1}\frac{2^{\gamma+\beta-1}}{\sigma_S^\alpha\Delta_S^{\beta-1}}\sum_{\mathbf{I}\in[\pm r]^\beta}\left\|M^T E_{\text{sym}}^{\alpha_0} w_{i_1}\right\|\left\|w_{i_{h\beta_h}}^T E_{\text{sym}}^{\alpha_h} M'\right\| \\
&= \|X\|_\infty^{\beta-h}\|E\|^{\alpha-\alpha_0-\alpha_h+h-1}\frac{2^{\gamma+\beta-1}(2r)^{\beta-2}}{\sigma_S^\alpha\Delta_S^{\beta-1}}\sum_{i\in[\pm r]}\left\|w_i^T E_{\text{sym}}^{\alpha_0} M\right\|\sum_{i\in[\pm r]}\left\|w_i^T E_{\text{sym}}^{\alpha_h} M'\right\| \\
&= \|X\|_\infty^{\beta-h}\|E\|^{\alpha+h-1}\frac{2^{\gamma+\beta-1}(2r)^{\beta}}{\sigma_S^\alpha\Delta_S^{\beta-1}}\sum_{i\in[\pm r]}\frac{\left\|w_i^T E_{\text{sym}}^{\alpha_0} M\right\|}{2r\|E\|^{\alpha_0}}\sum_{i\in[\pm r]}\frac{\left\|w_i^T E_{\text{sym}}^{\alpha_h} M'\right\|}{2r\|E\|^{\alpha_h}} \\
&\le \tau(M)\tau(M')\|X\|_\infty^{\beta-h}\|E\|^{\alpha+h-1}\frac{2^{\gamma+\beta-1}(2r)^{\beta}}{\sigma_S^\alpha\Delta_S^{\beta-1}}.
\end{aligned} \tag{53}$$

After rearrangements, we get

$$\|M^T\mathcal{T}(\boldsymbol{\alpha},\boldsymbol{\beta})M'\| \le r\tau(M)\tau(M')2^{\gamma+\beta-1}\left[\frac{2r\|X\|_\infty}{\Delta_S}\right]^{\beta-h}\left[\frac{\|E\|}{\sigma_S}\right]^{\alpha-h+1}\left[\frac{\sqrt{2r}\|E\|}{\sqrt{\sigma_S\Delta_S}}\right]^{2(h-1)}.$$

By the definitions of R, R_1 and R_2, we can replace the first two powers with R_1 and the third with R_2 to get

$$\|M^T\mathcal{T}(\boldsymbol{\alpha},\boldsymbol{\beta})M'\| \le r\tau(M)\tau(M')2^{\gamma+\beta-1}R_1^{\gamma-2h+2}R_2^{2(h-1)}.$$

Suppose $h < \gamma/2 + 1$, then $\gamma - 2h + 2 \geq 1$, so we further have the bound

$$\|M^T \mathcal{T}(\boldsymbol{\alpha}, \boldsymbol{\beta}) M'\| \leq r\tau(M)\tau(M')2^{\gamma+\beta-1} R_1 R^{\gamma-2h+1+2(h-1)} = r\tau(M)\tau(M')2^{\beta} R_1 (2R)^{\gamma-1}.$$

We get the first case of Eq. (47). Now consider the case $h = \gamma/2 + 1$, which only happens when γ is even. The previous bound becomes

$$\|M^T \mathcal{T}(\boldsymbol{\alpha}, \boldsymbol{\beta}) M'\| \leq r\tau(M)\tau(M')2^{\gamma+\beta-1} R_2^{2(h-1)} = r\tau(M)\tau(M')2^{\beta-1}(2R_2)^{\gamma}. \tag{54}$$

If we are content with this bound, continuing the rest of the proof will lead to the final bound

$$\|M^T \mathcal{T} M'\| \leq Cr\tau(M)\tau(M')(R_1 + R_2^2),$$

which is fine, but slightly less efficient than the target

$$\|M^T \mathcal{T} M'\| \leq Cr\tau(M)\tau(M')(R_1 + R_3),$$

since it is trivial that $R_3 \leq R_2^2$, and can be much smaller in some cases (see Remark 15).

To reach the target, we need to extract at least one factor of R_1 or R_3 from the bound, rather than having R_2^{γ}, hence a more delicate argument is needed.

If $\gamma = 2h - 2$, then $\alpha_0 = \alpha_h = 0$ and $\alpha_1 = \ldots = \alpha_{h-1} = \beta_1 = \ldots = \beta_h = 1$, thus $\beta = h$. Let $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ denote the corresponding tuple. Plugging into Eq. (52) and simplifying, we have

$$M^T \mathcal{T}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) M' = \sum_{\mathbf{I} \in [\pm r]^h} \mathcal{C}(\mathbf{I}) \left(M^T w_{i_1}\right) \left(w_{i_h}^T M'\right) \prod_{k=1}^{h-1} w_{i_k}^T E_{\text{sym}}^2 w_{i_{k+1}},$$

Consider the long product at the end of the right-hand side. For the purpose of this proof, let $y := \max_{|i| \neq |j|} |w_i^T E_{\text{sym}}^2 w_j|$ (the term in R_3's definition). Note that this is smaller than the term y in Theorem 14. Our goal is to extract at least one factor y out from the product, which should give rise to R_3. Therefore, consider two subcases for $\mathbf{I}$:

(1) There is k so that $|i_k| \neq |i_{k+1}|$, Then $\left|w_{i_k}^T E_{\text{sym}}^2 w_{i_{k+1}}\right| \leq y$ and we are good. The rest of the product can be bounded by $\|E\|^2$. The total contribution of this subcase is at most

$$r\tau(M)\tau(M')2^{\gamma+\beta-1} R_3 R_2^{\gamma-2} = r\tau(M)\tau(M')2^{3\gamma/2} R_3 R_2^{\gamma-2},$$

since we can simply replace a factor of R_2^2 in Eq. (54) with R_3.

(2) $|i_k| = i$ for all $k \in [h-1]$, for some $i \in [r]$. If $i \notin S$, then it is trivial from the definition of $\mathcal{C}$ in (20) that $\mathcal{C}(\mathbf{I}) = 0$. Suppose $i \in S$, it is time for us to apply the first, stronger bound in Lemma 21. The key improvement is the fact $\Delta_S(\mathbf{I}) = \sigma_i \geq \sigma_S$, instead of $\Delta_S(\mathbf{I}) \geq \Delta_S$ in the normal cases, so we get

$$|\mathcal{C}(\mathbf{I})| \leq \frac{2^{\gamma+\beta-1}}{\sigma_S(\mathbf{I})^{\gamma+1-\beta} \Delta_S(\mathbf{I})^{\beta-1}} \leq \frac{2^{3\gamma/2}}{\sigma_S^{\gamma}}.$$

The monomial matrix total contribution of this subcase is at most

$$\tau(M)\tau(M') \sum_{i \in S} \sum_{\mathbf{I} \in \{\pm i\}^h} \frac{2^{3\gamma/2} \|E\|^{2(h-1)}}{\sigma_S^{\gamma}} = r\tau(M)\tau(M') \frac{2^{3\gamma/2+h} \|E\|^{\gamma}}{\sigma_S^{\gamma}} \leq 2r\tau(M)\tau(M')(4R_1)^{\gamma}.$$

Therefore, the contribution of the case $h = \gamma/2 + 1$ is at most

$$r\tau(M)\tau(M')\left[2^{3\gamma/2}R_3R_2^{\gamma-2} + 2(4R_1)^{\gamma}\right] \le 16r\tau(M)\tau(M')(4R)^{\gamma-2}\left(R_3 + R_1^2\right).$$

The proof is complete in the case $1 \le \gamma \le 10\log(m+n)$. For the case $\gamma > 10\log(m+n)$, consider Eq. (53) again. We cannot use $\tau(M)$ and $\tau(M')$ anymore, but we can use the trivial upper bounds

$$\sum_{i\in[\pm r]} \frac{\left\|w_i^T E_{\text{sym}}^{\alpha_0} M\right\|}{2r\|E\|^{\alpha_0}} \le \|M\|, \qquad \sum_{i\in[\pm r]} \frac{\left\|w_i^T E_{\text{sym}}^{\alpha_h} M'\right\|}{2r\|E\|^{\alpha_h}} \le \|M'\|$$

in place of $\tau(M)$ and $\tau(M')$, which complete the proof. ■

Let us proceed with the proof of Lemma 24, which simply involve summing up the bounds in Lemma 23 over all choices of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

Proof [Proof of Lemma 24] Let us consider the case $\gamma \le 10\log(m+n)$ first. Recall that

$$M^T\mathcal{T}^{(\gamma)}M' = \sum_{h=1}^{\lfloor\gamma/2\rfloor+1} \sum_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\Pi_h(\gamma)} M^T\mathcal{T}(\boldsymbol{\alpha},\boldsymbol{\beta})M'. \tag{55}$$

Consider the easy case where γ is odd. Then $h < \gamma/2 + 1$, and we have, by Lemma 23,

$$\begin{aligned}\left\|M^T\mathcal{T}^{(\gamma)}M'\right\| &\le \sum_{h=1}^{\lfloor\gamma/2\rfloor+1} \sum_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\Pi_h(\gamma)} r\tau(M)\tau(M')2^{\beta}R_1(2R)^{\gamma-1} \\ &= r\tau(M)\tau(M')R_1(2R)^{\gamma-1} \sum_{h=1}^{\lfloor\gamma/2\rfloor+1} \sum_{\beta=h}^{\gamma+2-h} 2^{\beta} \left|\{(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \Pi_h(\gamma) : \textstyle\sum_j \beta_j = \beta\}\right|\end{aligned} \tag{56}$$

The elements of the set at the end are just tuples $(\alpha_0, \ldots, \alpha_h, \beta_1, \ldots, \beta_h)$ such that

$$\beta_1, \ldots, \beta_h \ge 1, \quad \sum_{i=1}^{h} \beta_i = \beta, \quad \text{and} \quad \alpha_0, \alpha_h \ge 0, \quad \alpha_1, \ldots, \alpha_{h-1}, \quad \sum_{i=0}^{h} \alpha_i = \gamma + 1 - \beta.$$

The number of ways to choose such a tuple is $\binom{\beta-1}{h-1}\binom{\gamma+2-\beta}{h}$. Plugging into Eq. (56), we obtain

$$\begin{aligned}\left\|M^T\mathcal{T}_{\nu}^{(\gamma)}M'\right\| &\le r\tau(M)\tau(M')R_1(2R)^{\gamma-1} \sum_{h=1}^{\lfloor\gamma/2\rfloor+1} \sum_{\beta=h}^{\gamma+2-h} \binom{\beta-1}{h-1}\binom{\gamma+2-\beta}{h}2^{\beta} \\ &= r\tau(M)\tau(M')R_1(2R)^{\gamma-1} \sum_{\beta=1}^{\gamma+1} 2^{\beta} \sum_{h=1}^{\beta\wedge(\gamma+2-\beta)} \binom{\beta-1}{h-1}\binom{\gamma+2-\beta}{h} \\ &\le r\tau(M)\tau(M')R_1(2R)^{\gamma-1} \sum_{\beta=1}^{\gamma+1} 2^{\beta}\binom{\gamma+1}{\beta} = 9r\tau(M)\tau(M')R_1(6R)^{\gamma-1}.\end{aligned} \tag{57}$$

Now consider the case γ is even. The only extra term will be in the case $h = \gamma/2 + 1$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are both all 1s. Therefore, in total we have

$$\begin{aligned}\left\|M^T \mathcal{T}_\nu^{(\gamma)} M'\right\| &\le 9r\tau(M)\tau(M')R_1(6R)^{\gamma-1} + \mathbf{1}\{\gamma \text{ even}\} \cdot 16r\tau(M)\tau(M')(4R)^{\gamma-2}(R_3 + R_1^2) \\ &\le r\tau(M)\tau(M')\left[9R(6R)^{\gamma-1} + 16(4R)^{\gamma-2}\mathbf{1}\{\gamma \text{ even}\}(R_1^2 + R_3)\right]\end{aligned}$$

For the remaining case, $\gamma > 10\log(m+n)$, we can simply replace $\tau(M)$ with $\|M\|$ and similarly for M'. The proof is complete. ■

Now we finish the bound on the entire power series.

Proof [Proof of Lemma 25] For convenience, let $k = \lfloor 10\log(m+n)\rfloor$. Applying Lemma 24, we have

$$\begin{aligned}\sum_{\gamma=1}^{k}\left\|M^T \mathcal{T}_\nu^{(\gamma)} M'\right\| &\le r\tau(M)\tau(M')\left[9\sum_{\gamma=1}^{\infty} R_1(6R)^{\gamma-1} + 16(R_3 + R_1^2)\sum_{\gamma=1}^{\infty}(4R)^{2\gamma-2}\right] \\ &\le r\tau(M)\tau(M')\left[\frac{9R_1}{1-6R} + \frac{16(R_3+R_1^2)}{1-16R^2}\right] \le Cr\mathrm{L}_\nu\tau\tau'(R_1+R_3),\end{aligned}$$

and

$$\begin{aligned}&\sum_{\gamma=k+1}^{\infty}\left\|M^T \mathcal{T}_\nu^{(\gamma)} M'\right\| \le r\|M\|\|M'\|\left[9\sum_{\gamma=k+1}^{\infty} R_1(6R)^{\gamma-1} + 16(R_3 + R_1^2)\sum_{\gamma=\lceil (k+1)/2\rceil}^{\infty}(4R)^{2\gamma-2}\right] \\ &\le r\|M\|\|M'\|\left[\frac{9R_1(6R)^k}{1-6R} + \frac{16(4R)^{k-1}(R_3+R_1^2)}{1-16R^2}\right] \le \frac{Cr\|M\|\|M'\|(R_1+R_3)}{(m+n)^{2.5}}.\end{aligned}$$

The convergence is guaranteed by the geometrically vanishing bounds on the $\|\cdot\|$-norms of the terms. Summing up the two parts, we obtain, by the triangle inequality

$$\left\|M^T \mathcal{T}_\nu M'\right\| \le Cr\left(\tau(M)\tau(M') + \frac{\|M\|\|M'\|}{(m+n)^{2.5}}\right)(R_1 + R_3).$$

The proof is complete. ■

Appendix B. Full proof of the matrix completion theorems

B.1. Proof of Theorem 5

In this section, we prove Theorem 5, first in its basic form, so that the readers can grasp the technical intuitions, then with the sampling density condition (8) replacing (5). We will first assume Theorem 10 as a black box, then prove Theorem 10 in the next subsection.

Proof [Proof of Theorem 5 using Theorem 10] For convenience, let $K := K_{A,Z} = K_A + K_Z$. Recall that in the discussion, we defined the following terms:

$$\rho := \frac{\hat{p}}{p}, \quad \tilde{A} = \rho\hat{A} = p^{-1}A_{\Omega,Z}, \quad E = \tilde{A} - A,$$

and E is a random matrix with mean 0, independent entries. Consider the moments of E_{ij}. For each $l \geq 2$, we have

$$\begin{aligned}\mathbf{E}\left[|E_{ij}|^l\right] &= \frac{\mathbf{E}\left[|A_{ij}(1-p)+Z_{ij}|^l\right]}{p^{l-1}} + (1-p)|A_{ij}|^l \leq \frac{(K_A(1-p)+K_Z)^l}{p^{l-1}} + (1-p)K_A^l \\ &\leq \frac{1}{p^{l-1}}\left(\sum_{k=0}^{l-1}\binom{l}{k}K_A^k(1-p)^kK_Z^{l-k} + K_A^l(1-p)^l + p^{l-1}(1-p)K_A^l\right) \\ &\leq \frac{1}{p^{l-1}}\left(\sum_{k=0}^{l-1}\binom{l}{k}K_A^kK_Z^{l-k} + K_A^l\right) = \frac{(K_A+K_Z)^l}{p^{l-1}}.\end{aligned} \tag{58}$$

Thus E satisfies $\mathbf{E}\left[|E_{ij}|^l\right] = p^{1-l}K^l$, making Theorem 10 applicable here.

Recall that we have the SVD $\hat{A} = \sum_i \hat{\sigma}_i\hat{u}_i\hat{v}_i^T$. Similarly, denote the SVD of $\tilde{A}$ by

$$\tilde{A} = \sum_{i}^{\min\{m,n\}} \tilde{\sigma}_i\tilde{u}_i\tilde{v}_i^T.$$

We have the relation $\tilde{u}_i = \hat{u}_i$, $\tilde{v}_i = \hat{v}_i$ and $\tilde{\sigma}_i = \rho^{-1}\hat{\sigma}_i$ for each i.

From the sampling density assumption, a standard application of concentration bounds Hoeffding (1963); Chernoff (1952) guarantees that, with probability $1-O(N^{-2})$,

$$0.9 \leq 1 - \frac{1}{\sqrt{N}} \leq 1 - \frac{\log N}{\sqrt{pmn}} \leq \rho \leq 1 + \frac{\log N}{\sqrt{pmn}} \leq 1 + \frac{1}{\sqrt{N}} \leq 1.1. \tag{59}$$

Furthermore, an application of well-established bounds on random matrix norms gives

$$\|E\| \leq 2K\sqrt{N/p}, \tag{60}$$

with probability $1-O(N^{-1})$. See Bandeira and van Handel (2016); Vu (2007), (Tran and Vu, 2025, Lemma A.7) or Bandeira and van Handel (2016) for detailed proofs. Therefore we can assume both Eqs. (59) and (60) at the cost of an $O(N^{-1})$ exceptional probability. The sampling density condition (Eq. (5) is equivalent to the conjunction of two conditions:

$$p \geq C\left(\frac{1}{m}+\frac{1}{n}\right)\log^{10} N, \tag{61}$$

$$p \geq \frac{Cr^3K^2}{\varepsilon^2}\left(1+\frac{\mu_0^2}{\log^2 N}\right)\left(\frac{1}{m}+\frac{1}{n}\right)\log^6 N. \tag{62}$$

Before entering the three steps outlined in the proof sketch, we show that the SVD step is guaranteed to choose a valid $s \in [r]$ such that $\hat{\delta}_s \geq 20K\sqrt{r_{\max}N/\hat{p}}$. Choose an index $l \in [r]$ such that $\delta_l \geq \sigma_1/r$, which exists since $\sum_{l\in[r]}\delta_l = \sigma_1$. We have, by Weyl's inequality,

$$\tilde{\delta}_l \geq \delta_l - \|E\| \geq \frac{\sigma_1}{r} - 2K\sqrt{\frac{N}{p}} \geq (100r_{\max}^{1/2} - 4)K\sqrt{\frac{N}{p}} \geq 90K\sqrt{\frac{r_{\max}N}{p}}.$$

Therefore

$$\hat{\delta}_l \geq \rho^{-1}\tilde{\delta}_l \geq 90\rho^{-1}K\sqrt{\frac{r_{\max}N}{p}} \geq 80\rho^{-1/2}K\sqrt{\frac{r_{\max}N}{p}} = 80K\sqrt{\frac{r_{\max}N}{\hat{p}}},$$

so the cutoff point s is guaranteed to exist. To see why $s \in [r]$, note that, again by Weyl's inequality,

$$\tilde{\delta}_{r+1} \leq \tilde{\sigma}_{r+1} \leq \sigma_{r+1} + \|E\| = \|E\| \leq 2K\sqrt{N/p}.$$

Therefore,

$$\hat{\delta}_{r+1} = \rho^{-1}\tilde{\delta}_{r+1} \leq 2\rho^{-1}K\sqrt{\frac{N}{p}} \leq 3\rho^{-1/2}K\sqrt{\frac{N}{p}} = 3K\sqrt{\frac{N}{\hat{p}}} < 20K\sqrt{\frac{r_{\max}N}{\hat{p}}}.$$

We want to show that the first three steps of Approximate-and-Round 2 recover A up to an absolute error ε, namely $\|\hat{A}_s - A\|_\infty \leq \varepsilon$. We will now follow the three steps in the sketch.

1. *Bounding* $\|A - \rho^{-1}A\|_\infty$. We have

$$\|A - \rho^{-1}A\|_\infty = \left|\rho^{-1} - 1\right| \|A\|_\infty \leq \frac{K_A}{.9\sqrt{N}} < \varepsilon/4.$$

2. *Bounding* $\|A_s - A\|_\infty$. Firstly, we bound σ_{s+1}. We have, since $\tilde{\delta}_s$ is the last singular value gap of $\tilde{A}$ such that $\tilde{\delta}_s \geq 20\rho^{-1/2}K\sqrt{\frac{r_{\max}N}{p}}$, then

$$\sigma_{s+1} \leq \tilde{\sigma}_{s+1} + \|E\| \leq 20\rho^{-1/2}rK\sqrt{\frac{r_{\max}N}{p}} + 2K\sqrt{\frac{N}{p}} \leq 22rK\sqrt{\frac{r_{\max}N}{p}}. \tag{63}$$

   For each fixed indices j, k, we have

$$|(A_s - A)_{jk}| = \left|U_{j,\cdot}^T \Sigma_{[s+1,r]} V_{k,\cdot}\right| \leq \sigma_{s+1}\|U\|_{2,\infty}\|V\|_{2,\infty} \leq 22rK\sqrt{\frac{r_{\max}N}{p}}\frac{r\mu_0}{\sqrt{mn}}$$
$$= \sqrt{\frac{22^2 r^4 r_{\max}\mu_0^2 K^2}{p}\left(\frac{1}{m} + \frac{1}{n}\right)} \leq \sqrt{\frac{22^2 r r_{\max}}{C\log^4 N}} \leq \varepsilon/4,$$

   where the last inequality comes from Eq. (62) and the assumption that $r_{\max} \leq \log^2 N$, if C is large enough. Since this holds for all pairs (j, k), we have $\|A_s - A\|_\infty \leq \varepsilon/4$.

3. *Bounding* $\|\tilde{A}_s - A_s\|_\infty$*, with probability* $1 - O(N^{-1})$. The condition (61) guarantees that we can apply Theorem 10. We will get, by Eq. (9)

$$\|\tilde{A}_s - A_s\|_\infty = C'\frac{(\mu_0 + \log N)\log^2 N}{\sqrt{mn}} \cdot rK\left(\sqrt{\frac{N}{p}} + \frac{\log N}{p}\frac{\sigma_s}{\delta_s}\right). \tag{64}$$

Under the assumption $r_{\max} \le \log^2 N$, we can further simplify. We have

$$\delta_s \ge \tilde{\delta}_s - 2\|E\| \ge 20\rho^{-1/2}\sqrt{\frac{r_{\max}N}{p}} - 4K\sqrt{\frac{N}{p}} \ge 18\sqrt{\frac{r_{\max}N}{p}}.$$

Therefore, by Eq. (63), $\sigma_{s+1} < 2\delta_s$, so

$$\frac{\sigma_s}{\delta_s} = 1 + \frac{\sigma_{s+1}}{\delta_s} < 1 + 2r \le 3r.$$

Back to Eq. (64), consider the second factor of the right-hand side, we have

$$\frac{\log N}{p}\frac{\sigma_s}{\delta_s} \le \frac{3r\log N}{p} \le 3r\sqrt{\frac{N}{p}}\frac{\log N}{\sqrt{pN}} \le \frac{3r}{\log^4 N}\cdot\sqrt{\frac{N}{p}} < \sqrt{\frac{N}{p}}.$$

Therefore Eq. (64) becomes

$$\|\tilde{A}_s - A_s\|_\infty \le 2C'\frac{(\mu_0 + \log N)rK\sqrt{N}\log^2 N}{\sqrt{pmn}}.$$

We want this term to be at most $\varepsilon/4$. We have

$$\begin{aligned} 2C'\frac{(\mu_0 + \log N)rK\sqrt{N}\log^2 N}{\sqrt{pmn}} \le \frac{\varepsilon}{4} \;&\Leftrightarrow\; p \ge (2C')^2 r^2 K^2(\mu_0 + \log N)^2\cdot\frac{N}{mn}\log^4 N \\ &\Leftrightarrow\; p \ge (2C')^2 r^2 K^2\left(1 + \frac{\mu_0}{\log N}\right)^2\left(\frac{1}{m} + \frac{1}{n}\right)\log^6 N. \end{aligned}$$

This is satisfied by the condition (62), when C is large enough. The third step is complete.

Now we combine the three steps. We have, by triangle inequality,

$$\begin{aligned} \|\hat{A}_s - A\|_\infty = \|\rho^{-1}\tilde{A}_s - A\|_\infty &\le \|A - \rho^{-1}A\|_\infty + \rho^{-1}\left(\|\tilde{A}_s - A_s\|_\infty + \|A_s - A\|_\infty\right) \\ &\le \frac{\varepsilon}{4} + 1.2\left(\frac{\varepsilon}{4} + \frac{\varepsilon}{4}\right) < .9\varepsilon. \end{aligned}$$

The total exceptional probability is $O(N^{-1})$. The proof is complete. ■

Proof [Proof of the full Theorem 5] Let $C_2 = 1/c$ for the constant c in Theorem 17. We rewrite the assumptions below:

1. *Signal-to-noise:* $\sigma_1 \ge 100r\kappa\sqrt{r_{\max}N}$.

2. *Sampling density:* this is equivalent to the conjunction of three conditions:

$$p \ge \frac{Cr^4 r_{\max}\mu_0^2 K_{A,Z}^2}{\varepsilon^2}\left(\frac{1}{m} + \frac{1}{n}\right), \tag{65}$$

$$p \ge C\left(\frac{1}{m} + \frac{1}{n}\right)\log^{10} N, \tag{66}$$

$$p \ge \frac{Cr^3 K_{A,Z}^2}{\varepsilon^2}\left(1 + \frac{\mu_0^2}{\log^2 N}\right)\left(1 + \frac{r^3\log N}{N}\right)\left(\frac{1}{m} + \frac{1}{n}\right)\log^6 N. \tag{67}$$

Let $\rho := \hat{p}/p$. From the sampling density assumption, a standard application of concentration bounds Hoeffding (1963); Chernoff (1952) guarantees that, with probability $1-O(N^{-2})$.

$$0.9 \leq 1 - \frac{1}{\sqrt{N}} \leq 1 - \frac{\log N}{\sqrt{pmn}} \leq \rho \leq 1 + \frac{\log N}{\sqrt{pmn}} \leq 1 + \frac{1}{\sqrt{N}} \leq 1.1. \tag{68}$$

Furthermore, an application of well-established bounds on random matrix norms gives

$$\|E\| \leq 2\kappa\sqrt{N}, \tag{69}$$

with probability $1 - O(N^{-1})$. See Bandeira and van Handel (2016); Vu (2007), (Tran and Vu, 2025, Lemma A.7) or Bandeira and van Handel (2016) for detailed proofs. Therefore we can assume both Eqs. (68) and (69) at the cost of an $O(N^{-1})$ exceptional probability.

Let $C_0 := 40$. The index s chosen in the SVD step of Approximate-and-Round 2 is the largest such that

$$\hat{\delta}_s \geq C_0 K_{A,Z}\sqrt{r_{\max}N/\hat{p}} = C_0\rho^{-1/2}\kappa\sqrt{r_{\max}N}.$$

Firstly, we show that SVD step is guaranteed to choose a valid $s \in [r]$. Choose an index $l \in [r]$ such that $\delta_l \geq \sigma_1/r \geq 100\kappa\sqrt{r_{\max}N}$, we have

$$\hat{\delta}_l \geq \rho^{-1/2}\tilde{\delta}_l \geq \rho^{-1/2}(\delta_l - 2\|E\|) \geq (100r_{\max}^{1/2} - 4)\rho^{-1/2}\kappa\sqrt{N} \geq 2C_0\rho^{-1/2}\kappa\sqrt{r_{\max}N},$$

so the cutoff point s is guaranteed to exist. To see why $s \in [r]$, note that

$$\hat{\delta}_{r+1} \leq \rho^{-1/2}\tilde{\sigma}_{r+1} \leq \rho^{-1/2}\|E\| \leq 2\rho^{-1/2}\kappa\sqrt{r_{\max}N} < C_0\rho^{-1/2}\kappa\sqrt{r_{\max}N}.$$

We want to show that the first three steps of Approximate-and-Round 2 recover A up to an absolute error ε, namely $\|\hat{A}_s - A\|_\infty \leq \varepsilon$, we will first show that $\|\tilde{A}_s - A\|_\infty \leq \varepsilon/2$ (with probability $1 - O(N^{-1})$). We proceed in two steps:

1. We will show that $\|A_s - A\|_\infty \leq \varepsilon/4$ when C is large enough. To this end, we establish:

$$\sigma_{s+1} \leq r\delta_{s+1} \leq r(\tilde{\delta}_{s+1} + 2\|E\|) \leq r(C_0\rho^{-1/2}\sqrt{r_{\max}} + 4)\kappa\sqrt{N} \leq 2rC_0K_{A,Z}\sqrt{r_{\max}N/p}. \tag{70}$$

For each fixed indices j, k, we have

$$|(A_s - A)_{jk}| = \left|U_{j,\cdot}^T\Sigma_{[s+1,r]}V_{k,\cdot}\right| \leq \sigma_{s+1}\|U\|_{2,\infty}\|V\|_{2,\infty} \leq 2rC_0K_{A,Z}\sqrt{\frac{r_{\max}N}{p}}\frac{r\mu_0}{\sqrt{mn}}$$
$$= \sqrt{\frac{4C_0^2r^4r_{\max}\mu_0^2K_{A,Z}^2}{p}\left(\frac{1}{m} + \frac{1}{n}\right)} \leq \varepsilon/4.$$

where the last inequality comes from the assumption (65) if C is large enough. Since this holds for all pairs (j, k), we have $\|A_s - A\|_\infty \leq \varepsilon/4$.

2. Secondly, we will show that $\|\tilde{A}_s - A_s\|_\infty \leq \varepsilon/4$ with probability $1 - O(N^{-1})$. We aim to use Theorem 17, so let us translate its terms into the current context. By the sampling density condition, we have the following lower bounds for δ_s and σ_s:

$$\sigma_s \geq \delta_s \geq \tilde{\delta}_s - 2\|E\| \geq C_0\rho^{-1/2}\kappa\sqrt{r_{\max}N} - 2\|E\| \geq .9C_0\kappa\sqrt{r_{\max}N}. \tag{71}$$

Consider the condition (20). If it holds, then we can apply Theorem 17. We want

$$\frac{\kappa\sqrt{N}}{\sigma_s} \vee \frac{r\kappa(\sqrt{\log N} + K\|U\|_\infty\|V\|_\infty \log N)}{\delta_s} \vee \frac{\kappa\sqrt{rN}}{\sqrt{\delta_s\sigma_s}} \leq \frac{1}{16}$$

By Eq. (71), we can replace all three denominators above with $.9C_0\kappa\sqrt{r_{\max}N}$. Additionally, $\|U\|_\infty \leq \|U\|_{2,\infty} \leq \sqrt{\frac{r\mu_0}{m}}$ and $\|V\|_\infty \leq \|V\|_{2,\infty} \leq \sqrt{\frac{r\mu_0}{n}}$, so we can replace them with these upper bounds. We also replace K with $p^{-1/2}$ (its definition). We want

$$\frac{\kappa\sqrt{N} \ \vee \ \kappa\sqrt{rN} \ \vee \ r\kappa(\sqrt{\log N} + \frac{r\mu_0}{\sqrt{pmn}} \log N)}{.9C_0\kappa\sqrt{r_{\max}N}} \leq \frac{1}{16},$$

which is equivalent to

$$\frac{1 \ \vee \ \sqrt{r} \ \vee \ r(\sqrt{\frac{\log N}{N}} + \frac{r\mu_0}{\sqrt{pmnN}} \log N)}{.9C_0\sqrt{r_{\max}}} \leq \frac{1}{16}$$

which easily holds. Therefore we can apply Theorem 17. We get, for a constant C_1,

$$\|\tilde{A}_s - A_s\|_\infty \leq C_1\tau_{UV}\tau_{VU} \cdot r\sigma_s R_s + \frac{1}{N}.$$

Let us simplify the first term in the product, $\tau_{UV}\tau_{VU}$.

$$\begin{aligned}
\tau_{UV} &= \frac{K\|U\|_{2,\infty}\log^3 N}{\sqrt{rN}} + \frac{\log^{3/2} N}{\sqrt{N}} + \frac{\|V\|_{2,\infty}\log N}{\sqrt{r}} \\
&\leq \frac{\sqrt{\mu_0}\log^3 N}{\sqrt{pmN}} + \frac{\log^{3/2} N}{\sqrt{N}} + \frac{\sqrt{\mu_0}\log N}{\sqrt{n}} \leq \frac{\log^{3/2} N}{\sqrt{N}} + \frac{\sqrt{2\mu_0}\log N}{\sqrt{n}},
\end{aligned}$$

where the first inequality comes from (66) if C is large enough. Similarly,

$$\tau_{VU} \leq N^{-1/2}\log^{3/2} N + m^{-1/2}\sqrt{2\mu_0}\log N.$$

Therefore,

$$\begin{aligned}
\tau_{UV}\tau_{VU} &\leq \frac{\log^3 N}{N} + \frac{\sqrt{\mu_0}\log^{5/2} N}{\sqrt{N}} \cdot \frac{\sqrt{2m}+\sqrt{2n}}{\sqrt{mn}} + \frac{2\mu_0\log^2 N}{\sqrt{mn}} \\
&\leq \log^2 N \frac{\log N + 4\sqrt{\mu_0}\sqrt{\log N} + 4\mu_0}{2\sqrt{mn}} \leq \log^2 N \frac{\log N + 4\mu_0}{\sqrt{mn}}.
\end{aligned}$$

For the second term, we have the following upper bound:

$$\begin{aligned}
r\sigma_s R_s &\leq r\sigma_s\left(\frac{\kappa\sqrt{N}}{\sigma_s} + \frac{r\kappa(\sqrt{\log N} + \frac{r\mu_0}{\sqrt{mn}}K\log N}{\delta_s} + \frac{r\kappa^2 N}{\delta_s\sigma_s}\right) \\
&= r\left(\kappa\sqrt{N} + \frac{r\kappa\sigma_s}{\delta_s}\left(\sqrt{\log N} + \frac{r\mu_0\log N}{\sqrt{pmn}}\right) + \frac{r\kappa^2 N}{\delta_s}\right) \\
&\leq r\left(\kappa\sqrt{N} + r^2\kappa\left(\sqrt{\log N} + \frac{r\mu_0\log N}{\sqrt{pmn}}\right) + \frac{r\kappa^2 N}{.9C_0\kappa\sqrt{rN}}\right) \\
&\leq r^{3/2}\kappa\left(\sqrt{2N} + r^{3/2}\left(\sqrt{\log N} + \frac{r\mu_0\log N}{\sqrt{pmn}}\right)\right).
\end{aligned}$$

Under the condition (67), we have

$$pmn \geq Cr^3\mu_0^2 \log^4 N \implies \frac{r\mu_0 \log N}{\sqrt{pmn}} < .1\sqrt{\log N},$$

so the above is simply upper bounded by

$$\frac{\sqrt{2}r^{3/2}K_{A,Z}}{\sqrt{p}}\left(\sqrt{N} + r^{3/2}\sqrt{\log N}\right).$$

Multiplying the two terms, we have by Theorem 17,

$$\begin{aligned} \|\tilde{A}_s - A_s\|_\infty &\leq \log^2 N \cdot \frac{\log N + 4\mu_0}{\sqrt{mn}} \cdot \frac{\sqrt{2}r^{3/2}K_{A,Z}}{\sqrt{p}}\left(\sqrt{N} + r^{3/2}\sqrt{\log N}\right) \\ &\leq \sqrt{\frac{2r^3K_{A,Z}^2\log^6 N}{p}\left(1 + \frac{4\mu_0^2}{\log^2 N}\right)\left(1 + \frac{r^3\log N}{N}\right)\left(\frac{1}{m} + \frac{1}{n}\right)} \leq \varepsilon/4. \end{aligned} \tag{72}$$

where the last inequality comes from the condition (67) if C is large enough.

After the two steps above, we obtain $\|\tilde{A}_s - A\|_\infty \leq \varepsilon/2$ with probablity $1 - O(N^{-1})$. Finally, we get, using Fact (68) and the triangle inequality,

$$\|\hat{A}_s - A\|_\infty = \left\|\rho^{-1}\tilde{A}_s - A\right\|_\infty \leq \frac{1}{\rho}\|\tilde{A}_s - A\|_\infty + \left|\frac{1}{\rho} - 1\right| \|A\|_\infty \leq \frac{\varepsilon/2}{.9} + \frac{K_A}{.9\sqrt{N}} < \varepsilon.$$

This is the desired bound. The total exceptional probability is $O(N^{-1})$. The proof is complete. ■

B.2. Proof of Theorem 10 using Theorem 17

Proof [Proof of Theorem 10] Let $\varsigma = K/\sqrt{p}$ and $M = 1/\sqrt{p}$. Then for C sufficiently large, $p \geq C(m^{-1} + n^{-1})\log^{10} N$ implies $M \leq c\sqrt{N}\log^{-5} N$, meaning we can apply Theorem 17, specifically Eq. (23) for this choice of ς and M if the condition (20) holds. We check it for $S = [s]$. Given that $\sigma_s \geq \delta_s \geq 40K\sqrt{rN/p}$, we have

$$\frac{\varsigma\sqrt{N}}{\sigma_S} = \frac{K}{\sigma_S}\sqrt{\frac{rN}{p}} \leq \frac{1}{40\sqrt{r}} < \frac{1}{16}, \qquad \frac{\varsigma\sqrt{rN}}{\sqrt{\delta_s\sigma_s}} \leq \frac{K\sqrt{rN}}{\sqrt{p}\cdot 40rK\sqrt{rN/p}} \leq \frac{1}{40} < \frac{1}{16},$$

and, using the fact $\mu_0 \leq N$ and the assumption $r \leq \log^2 N$,

$$\begin{aligned} &\frac{r\varsigma(\sqrt{\log N} + M\|U\|_\infty\|V\|_\infty \log N}{\delta_S} \leq \frac{rK\sqrt{\log N}}{\delta_s\sqrt{p}} + \frac{r^2K\mu_0\log N}{\delta_s p\sqrt{mn}} \\ &\leq \frac{\sqrt{r}\log N}{40\sqrt{N}} + \frac{r^{3/2}\mu_0\log N}{\sqrt{pmnN}} \leq \frac{\sqrt{r}\log N}{40\sqrt{N}} + \frac{r^{3/2}\mu_0\log N}{\sqrt{C}N\log^5 N} \leq \frac{1}{\log N} < \frac{1}{16}. \end{aligned}$$

It remains to transform the right-hand side of Eq. (23) to the right-hand side of Eq. (9). We have

$$\tau_1 \leq \frac{\sqrt{\mu_0}\log^3 N}{\sqrt{pmN}} + \frac{\log^{3/2} N}{\sqrt{N}} + \frac{\sqrt{\mu_0}\log N}{\sqrt{n}} \leq \frac{\log^{3/2} N}{\sqrt{N}} + \frac{\sqrt{2\mu_0}\log N}{\sqrt{n}}.$$

Combining with the symmetric bound for τ_2, we get

$$\tau_1\tau_2 \le \frac{\log^3 N}{N} + \frac{\sqrt{\mu_0}\log^{5/2} N}{\sqrt{N}} \cdot \frac{\sqrt{2m}+\sqrt{2n}}{\sqrt{mn}} + \frac{2\mu_0 \log^2 N}{\sqrt{mn}} \le 4\log^2 N \frac{\log N + \mu_0}{\sqrt{mn}},$$

which is the first factor on the right-hand side of Eq. (9).

Consider the term R_s. From the above, we have

$$R_s \le \frac{K}{\sigma_S}\sqrt{\frac{rN}{p}} + \frac{rK\sqrt{\log N}}{\delta_s\sqrt{p}} + \frac{r^2 K \mu_0 \log N}{\delta_s p \sqrt{mn}} + \frac{K^2 rN}{p\delta_s\sigma_s}.$$

Since $\delta_s \ge 40K\sqrt{rN/p}$, the fourth term is at most $1/40$ of the first term. Removing it recovers exactly the second factor on the right-hand side of Eq. (9). The proof is complete. ■

Appendix C. Proofs of technical lemmas

C.1. Proof of bound for contour integrals of polynomial reciprocals

In this section, we prove Lemmas 21 and 26, which provide the necessary bounds on the integral coefficients in the proof of Theorem 14. Recall that the integrals we are interested in have the form

$$\mathcal{C}(\mathbf{I}) = \oint_{\Gamma_S} \frac{z^\nu \mathrm{d}z}{2\pi i}\frac{1}{z^{\gamma+1}}\prod_{k=1}^{\beta}\frac{\sigma_{i_k}}{z-\sigma_{i_k}}, \qquad \mathcal{C}_1(\mathbf{I}) = \oint_{\Gamma_S} \frac{z\mathrm{d}z}{2\pi i}\frac{1}{z^{\gamma+1}}\prod_{k=1}^{\beta}\frac{\sigma_{i_k}}{z-\sigma_{i_k}},$$

where $\beta \le \gamma+1$. We can combine them into the common form below:

$$\mathcal{C}_\nu(\mathbf{I}) := \oint_{\Gamma_S} \frac{z^\nu \mathrm{d}z}{2\pi i}\frac{1}{z^{\gamma+1}}\prod_{k=1}^{\beta}\frac{\sigma_{i_k}}{z-\sigma_{i_k}}, \quad \text{where } \nu \in \{0,1\} \text{ and } \beta \le \gamma+1. \tag{73}$$

Let the multiset $\{\sigma_{i_k}\}_{k\in[\beta]} = A \cup B$, where $A := \{a_i\}_{i\in[l]}$ and $B := \{b_j\}_{j\in[k]}$, where each $a_i \in S$ and each $b_j \notin S$, having multiplicities m_i and n_j respectively. We can rewrite the above into

$$\mathcal{C}_\nu(\mathbf{I}) = \prod_{i=1}^{l} a_i^{m_i}\prod_{j=1}^{k} b_j^{n_j} C(n_0; A, \mathbf{m}; B, \mathbf{n}), \tag{74}$$

where

$$C(n_0; A, \mathbf{m}; B, \mathbf{n}) := \oint_{\Gamma_A} \frac{\mathrm{d}z}{2\pi i}\frac{1}{z^{n_0}}\prod_{j=1}^{k}\frac{1}{(z-b_j)^{n_j}}\prod_{i=1}^{l}\frac{1}{(z-a_i)^{m_i}}, \tag{75}$$

where $n_0 = \gamma+1-\nu$. The m_i's and n_j's satisfy $\sum_i m_i + \sum_j n_j \le \gamma+1$. We can remove the set S and simply denote the contour by Γ_A without affecting its meaning. The next three results will build up the argument to bound these sums and ultimately prove the target lemmas.

Lemma 27 *Let $A = \{a_i\}_{i\in[l]}$ and $B = \{b_j\}_{j\in[k]}$ be disjoint set of complex non-zero numbers and $\mathbf{m} = \{m_i\}_{i\in[l]}$ and n_0 and $\mathbf{n} = \{n_j\}_{j\in[k]}$ be nonnegative integers such that $m+n+n_0 \geq 2$, where $m = \sum_i m_i$ and $n := \sum_{i\geq 1} n_i$. Let Γ_A be a contour encircling all numbers in A and none in $B \cup \{0\}$. Let $a, d > 0$ be arbitrary such that:*

$$d \leq a, \qquad a \leq \min_i |a_i|, \qquad d \leq \min_{i,j} |a_i - b_j|. \tag{76}$$

Suppose that $0 \leq m_i' \leq m_i$ for each $i \in [l]$ and that $m' := \sum_{i=1}^k m_i' \leq n_0$. Then for $C(n_0; A, \mathbf{m}; B, \mathbf{n})$ defined Eq. (75), we have

$$|C(n_0; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m+n+n_0-2}{m-1} \frac{1}{a^{n_0-m'} d^{m+n-1}} \prod_{i=1}^{l} \frac{1}{|a_i|^{m_i'}} \tag{77}$$

Proof Firstly, given the sets A and B and the notations and conditions in Lemma 27, the weak bound below holds

$$|C(n_0; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m+n+n_0-2}{m-1} \frac{1}{d^{m+n+n_0-1}}. \tag{78}$$

We omit the details of the proof, which is a simple induction argument. We now use Eq. (78) to prove the following:

$$|C(n_0; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m+n+n_0-2}{m-1} \frac{1}{a^{n_0} d^{m+n-1}}. \tag{79}$$

We proceed with induction. Let $P_1(N)$ be the following statement: "For any sets A and B, and the notations and conditions described in Lemma 27, such that $m + n + n_0 = N$, Eq. (79) holds."

Since $m+n+n_0 \geq 2$, consider $N = 2$ for the base case. The only case where the integral is non-zero is when $m = 1$ and $n + n_0 = 1$, meaning $A = \{a_1\}$, $m_1 = 1$ and either $B = \varnothing$ and $n_0 = 1$, or $B = \{b_1\}$ and $n_1 = 1$, $n_0 = 0$. The integral yields a_1^{-1} in the former case and $(a_1 - b_1)^{-1}$ in the latter, confirming the inequality in both.

Consider $n \geq 3$ and assume $P_1(n-1)$. If $m = 0$, the integral is again 0. If $n_0 = 0$, Eq. (79) automatically holds by being the same as Eq. (78). Assume $m, n_0 \geq 1$. There must then be some $i \in [l]$ such that $m_i \geq 1$, without loss of generality let 1 be that i. We have

$$C(n_0; A, \mathbf{m}; B, \mathbf{n}) = \frac{1}{a_1}\Big[C(n_0 - 1; A, \mathbf{m}; B, \mathbf{n}) - C(n_0; A, \mathbf{m}^{(1)}; B, \mathbf{n})\Big] \tag{80}$$

where $\mathbf{m}^{(i)}$ is the same as $\mathbf{m}$ except that the i-entry is $m_i - 1$.

Consider the first integral on the right-hand side. Applying $P_1(N-1)$, we get

$$|C(n_0 - 1; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m+n+n_0-3}{m-1} \frac{1}{a^{n_0-1} d^{m+n-1}}. \tag{81}$$

Analogously, we have the following bound for the second integral:

$$\left|C(n_0; A, \mathbf{m}^{(1)}; B, \mathbf{n})\right| \leq \binom{m+n+n_0-3}{m-2} \frac{1}{a^{n_0} d^{m+n-2}} \leq \binom{m+n+n_0-3}{m-2} \frac{1}{a^{n_0-1} d^{m+n-1}}. \tag{82}$$

Notice that the binomial coefficients in Eqs. (81) and (82) sum to the binomial coefficient in Eq. (79), we get $P_1(N)$, which proves Eq. (79) by induction.

Now we can prove Eq. (77). The logic is almost identical, with Eq. (79) playing the role of Eq. (78) in its own proof, handling an edge case in the inductive step. Let $P_2(n)$ be the statement: "For any sets A and B, and the notations and conditions described in Lemma 27, such that $m + n + n_0 = N$, Eq. (77) holds."

The cases $N = 1$ and $N = 2$ are again trivially true. Consider $N \geq 3$ and assume $P_2(N-1)$. Fix any sequence $m'_1, m'_2, \ldots, m'_l$ satisfying $0 \leq m'_i \leq m_i$ for each $i \in [k]$ and $n_0 \geq m'_1 + \ldots + m'_k$. If $m'_1 = m'_2 = \ldots = m'_k = 0$, we are done by Eq. (79). By symmetry among the indices, assume $m'_1 \geq 1$. This also means $n_0 \geq 1$. Consider Eq. (80) again. For the first integral on the right-hand side, applying $P_2(N-1)$ for the parameters $n_0 - 1, n_1, \ldots, n_k$, $m_1, \ldots, m_l$ and $m'_1 - 1, m'_2, \ldots, m'_k$ yields the bound

$$|C(n_0 - 1; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m + n + n_0 - 3}{m - 1} \frac{1}{a^{n_0 - m'} d^{m+n-1}} \frac{1}{|a_1|^{m'_1 - 1}} \prod_{i=2}^{l} \frac{1}{|a_i|^{m'_i}}. \tag{83}$$

Applying $P_2(N-1)$ for the parameters $n_0, n_1, \ldots, n_k$, $m_1 - 1, \ldots, m_l$ and $m'_1 - 1, m'_2, \ldots, m'_k$, we get the following bound for the second integral on the right-hand side of Eq. (80):

$$\begin{aligned}
\left|C(n_0; A, \mathbf{m}^{(1)}; B, \mathbf{n})\right| &\leq \binom{m + n + n_0 - 3}{m - 2} \frac{1}{a^{n_0 - m' + 1} d^{m+n-2}} \frac{1}{|a_1|^{m'_1 - 1}} \prod_{i=2}^{l} \frac{1}{|a_i|^{m'_i}} \\
&\leq \binom{m + n + n_0 - 3}{m - 2} \frac{1}{a^{n_0 - m'} d^{m+n-1}} \frac{1}{|a_1|^{m'_1 - 1}} \prod_{i=2}^{l} \frac{1}{|a_i|^{m'_i}}.
\end{aligned}$$

Summing up the bounds by summing the binomial coefficients, we get exactly $P_2(N)$, so Eq. (77) is proven by induction. ∎

Lemma 28 *Let A, B, $\mathbf{m}$, $\mathbf{n}$, n_0, Γ_A and a, d be the same, with the same conditions as in Lemma 27. Suppose that $0 \leq m'_i \leq m_i$ and $0 \leq n'_j \leq n_j$ for each $i, j \geq 1$ and*

$$m' + n' \leq n_0 \quad \text{for} \quad m' := \sum_i m'_i, \quad n' := \sum_j n'_j.$$

Then for $C(n_0; A, \mathbf{m}; B, \mathbf{n})$ defined in Eq. (75), we have

$$|C(n_0; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{n + n_0 - n' + m - 2}{m - 1} \frac{(1 + d/a)^{n'}}{a^{n_0 - m' - n'} d^{m+n-1}} \prod_{i=1}^{l} \frac{1}{|a_i|^{m'_i}} \prod_{j=1}^{k} \frac{1}{|b_j|^{n'_j}}. \tag{84}$$

Proof We have the expansion

$$\frac{1}{z^{n_0}}\prod_{j=1}^{k}\frac{b_j^{n'_j}}{(z-b_j)^{n_j}}\prod_{i=1}^{l}\frac{1}{(z-a_i)^{m_i}}=\frac{1}{z^{n_0-n'}}\prod_{j=1}^{k}\frac{1}{(z-b_j)^{n_j-n'_j}}\prod_{j=1}^{k}\left(\frac{1}{z}-\frac{1}{z-b_j}\right)^{n'_j}\prod_{i=1}^{l}\frac{1}{(z-a_i)^{m_i}}$$

$$=\frac{1}{z^{n_0-n'}}\prod_{j=1}^{k}\frac{1}{(z-b_j)^{n_j-n'_j}}\sum_{0\le r_j\le n'_j\forall j}\frac{(-1)^{r_1+\ldots+r_k}}{z^{n'-r_1-\ldots-r_k}}\prod_{j=1}^{k}\binom{n'_j}{r_j}\frac{1}{(z-b_j)^{r_j}}\prod_{i=1}^{l}\frac{1}{(z-a_i)^{m_i}}$$

$$=\sum_{0\le r_j\le n'_j\forall j}\frac{(-1)^{r_1+\ldots+r_k}}{z^{n_0-r_1-\ldots-r_k}}\prod_{j=1}^{k}\binom{n'_j}{r_j}\frac{1}{(z-b_j)^{r_j+n_j-n'_j}}\prod_{i=1}^{l}\frac{1}{(z-a_i)^{m_i}}.$$

Integrating both sides over Γ_A, we have

$$C(n_0;A,\mathbf{m};B,\mathbf{n})\prod_{j=1}^{k}b_j^{n'_j}=\sum_{0\le r_j\le n'_j\forall j}(-1)^{\sum_j r_j}\binom{n'_j}{r_j}C\left(n_0-\textstyle\sum_j r_j;A,\mathbf{m};B,\mathbf{r}+\mathbf{n}-\mathbf{n}'\right),$$

where the j-entry of $\mathbf{r}+\mathbf{n}-\mathbf{n}'$ is simply $r_j+n_j-n'_j$. Applying Lemma 27 for each summand on the right-hand side and rearranging the powers, we get

$$\left|C\left(n_0-\textstyle\sum_j r_j;A,\mathbf{m};B,\mathbf{r}+\mathbf{n}-\mathbf{n}'\right)\right|\le\binom{m+n+n_0-n'-2}{m-1}\frac{(a/d)^{\sum_j r_j}}{a^{n_0-m'}d^{n-n'+m-1}}\prod_{i=1}^{l}\frac{1}{|a_i|^{m'_i}}.$$

Summing up the bounds, we get

$$\left|C(n_0;A,\mathbf{m};B,\mathbf{n})\prod_{j=1}^{k}b_j^{n'_j}\right|\le\binom{m+n+n_0-n'-2}{m-1}\frac{\prod_{i=1}^{l}|a_i|^{-m'_i}}{a^{n_0-m'}d^{n-n'+m-1}}\sum_{0\le r_j\le n'_j\forall j}\prod_{j=1}^{k}\binom{n'_j}{r_j}\frac{a^{r_j}}{d^{r_j}}$$

$$=\binom{m+n+n_0-n'-2}{m-1}\frac{\prod_{i=1}^{l}|a_i|^{-m'_i}}{a^{n_0-m'}d^{n-n'+m-1}}\left(\frac{a}{d}+1\right)^{n'}.$$

Rearranging the term, we get precisely the desired inequality. ■

With the lemma above, we are ready to prove both Lemmas 21 and 26.
Proof [Proof of Lemmas 21 and 26] First rewrite the integral into the forms of (73), then (74) and (75). Let us consider two cases for $\mathcal{C}$:

1. $\nu=0$, so $n_0=\gamma+1$. Let $a=\sigma_S(\mathbf{I})$, $d=\Delta_S(\mathbf{I})$, $m=\beta_S(\mathbf{I})$, $n=n'=\beta_{S^c}(\mathbf{I})$, $m'_i=m_i$ and $n'_j=n_j$ for all i,j, then $m'+n'=\beta\le\gamma+1=n_0$, so we can apply Lemma 28 to get

$$|C(n_0;A,\mathbf{m};B,\mathbf{n})|\le\binom{n_0+m-2}{m-1}\frac{(1+d/a)^{n'}}{a^{n_0-m-n}d^{m+n-1}}\prod_{i=1}^{l}\frac{1}{|a_i|^{m_i}}\prod_{j=1}^{k}\frac{1}{|b_j|^{n_j}},$$

or equivalently,

$$|\mathcal{C}_0(\mathbf{I})|\le\left(1+\frac{\Delta_S(\mathbf{I})}{\sigma_S(\mathbf{I})}\right)^{\beta_{S^c}(\mathbf{I})}\binom{\gamma+\beta_S(\mathbf{I})-1}{\beta_S(\mathbf{I})-1}\frac{1}{\sigma_S(\mathbf{I})^{\gamma+1-\beta}\Delta_S(\mathbf{I})^{\beta-1}}.$$

Since $\Delta_S(\mathbf{I}) \le \sigma_S(\mathbf{I})$ and the binomial coefficient is at most $2^{\gamma+\beta_S(\mathbf{I})-1}$, we get the final bound

$$|\mathcal{C}_0(\mathbf{I})| \le \frac{2^{\gamma+\beta_S(\mathbf{I})-1+\beta_{S^c}(\mathbf{I})}}{\sigma_S(\mathbf{I})^{\gamma+1-\beta}\Delta_S(\mathbf{I})^{\beta-1}} = \frac{2^{\gamma+\beta-1}}{\sigma_S(\mathbf{I})^{\gamma+1-\beta}\Delta_S(\mathbf{I})^{\beta-1}} \le \frac{2^{\gamma+\beta-1}}{\sigma_S^{\gamma+1-\beta}\Delta_S^{\beta-1}},$$

where the last inequality holds due to $\sigma_S(\mathbf{I}) \ge \sigma_S$ and $\Delta_S(\mathbf{I}) \ge \Delta_S$. The proof of Lemma 21 is complete.

2. $\nu = 1$ and $S = [s]$ for some $s \in [r]$. This is the special case for Lemma 26. Note that $n_0 = \gamma$ in this case. Without loss of generality, assume $|a_1| = \sigma_s(\mathbf{I})$, then we are guaranteed $m_1 \ge 1$. Applying Lemma 28 for the same parameters as in the previous case, except that $m_1' = m_1 - 1$, we get

$$|C(n_0; A, \mathbf{m}; B, \mathbf{n})| \le |a_1| \binom{n_0 + m - 2}{m - 1} \frac{(1+d/a)^{n'}}{a^{n_0-m+1-n}d^{m+n-1}} \prod_{i=1}^{l} \frac{1}{|a_i|^{m_i}} \prod_{j=1}^{k} \frac{1}{|b_j|^{n_j}},$$

which translates to

$$|\mathcal{C}_1(\mathbf{I})| \le \binom{\gamma + \beta_s(\mathbf{I}) - 2}{\beta_s(\mathbf{I}) - 1} \left(1 + \frac{\Delta_s(\mathbf{I})}{\sigma_s(\mathbf{I})}\right)^{\beta_{S^c}(\mathbf{I})} \frac{\sigma_s(\mathbf{I})}{\sigma_s(\mathbf{I})^{\gamma+1-\beta}\Delta_s(\mathbf{I})^{\beta-1}} \le \frac{2^{\gamma+\beta-2}}{\sigma_s(\mathbf{I})^{\gamma-\beta}\Delta_s(\mathbf{I})^{\beta-1}}.$$

Now, it may seem that we can simply replace $\sigma_s(\mathbf{I})$ and $\Delta_s(\mathbf{I})$ respectively with σ_s and Δ_s to get the final bound. This is true in most cases, but the situtation is more complicated when $\beta = \gamma + 1$, since the inequality $\sigma_s(\mathbf{I})^{\gamma-\beta} \ge \sigma_s^{\gamma-\beta}$ would be reversed. This is where the fact $S = [s]$ comes into play. Consider the case $\beta = \gamma + 1$. We have

$$\frac{2^{\gamma+\beta-2}}{\sigma_s(\mathbf{I})^{\gamma-\beta}\Delta_s(\mathbf{I})^{\beta-1}} \le \frac{2^{\gamma+\beta-2}}{\sigma_s^{\gamma-\beta}\Delta_s^{\beta-1}} \Leftrightarrow \frac{\sigma_s(\mathbf{I})}{\Delta_s(\mathbf{I})^{\gamma}} \le \frac{\sigma_s}{\Delta_s^{\gamma}}.$$

Since $\gamma \ge 1$, we have

$$\frac{1}{\Delta_s(\mathbf{I})^{\gamma-1}} \le \frac{1}{\Delta_s^{\gamma-1}}.$$

It suffices to show $\sigma_s(\mathbf{I})/\Delta_s(\mathbf{I}) \le \sigma_s/\delta_s$ to complete the last step. Choose $t \in [s]$ where $\sigma_t = \sigma_S(\mathbf{I})$, then $\Delta_S(\mathbf{I}) \ge \sigma_t - \sigma_{s+1}$, thus

$$\frac{\sigma_S(\mathbf{I})}{\Delta_S(\mathbf{I})} \le \frac{\sigma_t}{\sigma_t - \sigma_{s+1}} \le \frac{\sigma_s}{\sigma_s - \sigma_{s+1}} = \frac{\sigma_s}{\delta_s}.$$

This completes the final step, proving Lemma 26. Note that the inequality above does not hold if S does not contain a continguous chunk of the largest singular values.

∎

C.2. Proof of semi-isotropic bounds for powers of random matrices

In this section, we prove Lemma 19, which gives semi-itrosopic bounds for powers of E_{sym} in the second step of the main proof strategy.

The form of the bounds naturally implies that we should handle the even and odd powers separately. We split the two cases into the following lemmas.

Lemma 29 *Let $m, r \in \mathbb{N}$ and $U \in \mathbb{R}^{m\times r}$ be a matrix whose columns $u_1, u_2, \ldots, u_r$ are unit vectors. Let E be a $m \times n$ random matrix following Model (18) with parameters M and $\varsigma = 1$, meaning E has independent entries and*

$$\mathbf{E}\left[E_{ij}\right] = 0, \quad \mathbf{E}\left[\|E\|_{ij}^2\right] \leq 1, \quad \mathbf{E}\left[\|E\|_{ij}^p\right] \leq M^{p-2} \quad \textit{for all } \; p.$$

For any $a \in \mathbb{N}$, $k \in [n]$, for any $D > 0$, for any $p \in \mathbb{N}$ such that

$$m + n \geq 2^8 M^2 p^6 (2a+1)^4,$$

we have, with probability at least $1 - (2^5/D)^{2p}$,

$$\left\|e_{n,k}^T (E^T E)^a E^T U\right\| \leq D r^{1/2} p^{3/2} \sqrt{2a+1}\left(16 p^{3/2}(2a+1)^{3/2} M \frac{\|U\|_{2,\infty}}{\sqrt{r}} + 1\right)[2(m+n)]^a.$$

Lemma 30 *Let E be a $m \times n$ random matrix following the model in Lemma 29. For any matrix $V \in \mathbb{R}^{m\times l}$ with unit columns $v_1, v_2, \ldots, v_l$, any $a \in \mathbb{N}$, $k \in [n]$, any $D > 0$, and any $p \in \mathbb{N}$ such that*

$$m + n \geq 2^8 M^2 p^6 (2a)^4,$$

we have, with probability at least $1 - (2^4/D)^{2p}$,

$$\left\|e_{n,k}^T (E^T E)^a V\right\| \leq D p \|V\|_{2,\infty} [2(m+n)]^a.$$

Let us prove the main objective of this section, Lemma 19, before delving into the proof of the technical lemmas.

Proof [Proof of Lemma 19] We follow two steps:

1. *Assuming M*

Consider the analogue of Eq. (29) for V (we wrote the proof for V before the final edit, and wanted to save the energy of changing to U) and Eq. (30), and assume $M \leq \log^{-2-\varepsilon}(m + n)\sqrt{m+n}$. Fix $k \in [n]$. It suffices to prove the following two bounds uniformly over all $a \in [\lfloor t \log(m+n) \rfloor]$:

$$\left\|e_{n,k}^T (E^T E)^a E^T U\right\| \leq C \tau_1(U, \log\log(m+n))(1.9\varsigma\sqrt{m+n})^{2a+1}\sqrt{r} \tag{85}$$

$$\left\|e_{n,k}^T (E^T E)^a V\right\| \leq C \tau_0(V, \log\log(m+n))(1.9\varsigma\sqrt{m+n})^{2a}\sqrt{r}. \tag{86}$$

Fix $a \in [\lfloor t \log(m+n) \rfloor]$. Let $p = \log\log(m+n)$. We can assume p is an integer for simplicity without any loss. This choice ensures

$$M^2 p^6 (2a)^4 < M^2 p^6 (2a+1)^4 \leq \frac{(m+n) t^4 \log^4(m+n) \log^6 \log(m+n)}{\log^{4+2\varepsilon}(m+n)} = o(m+n),$$

so we can apply both Lemmas 29 and 30.

Let us prove Eq. (85) for a. Applying Lemma 29 for the random matrix E/ς and $D = 2^{13}$ gives, with probability $1 - \log^{-4.04}(m+n)$,

$$\frac{\|e_{n,k}^T (E^T E)^a E^T U\|}{(1.9\varsigma\sqrt{m+n})^{2a+1}} \le \frac{Dr^{1/2}p^{3/2}\sqrt{2a+1}}{1.9\sqrt{m+n}} \left(16p^{3/2}(2a+1)^{3/2}M\frac{\|U\|_{2,\infty}}{\sqrt{r}} + 1\right)\left(\frac{2}{3.61}\right)^a$$
$$\le \frac{Dr^{1/2}p^{3/2}}{\sqrt{m+n}}\left(16p^{3/2}M\frac{\|U\|_{2,\infty}}{\sqrt{r}} + 1\right) \le 2^{17}\sqrt{r}\left(\frac{Mp^3\|U\|_{2,\infty}}{\sqrt{r(m+n)}} + \frac{p^{3/2}}{\sqrt{m+n}}\right),$$

where the second inequality is due to $\alpha \le (\sqrt{2}/1.9)^{\alpha}$. A union bound over all $a \in [\lfloor t\log(m+n)\rfloor]$ makes the bound uniform, with probability at least $1 - \log^{-3}(m+n)$. The term inside parentheses in the last expression is less than $D_{U,V,\log\log(m+n)}$, so Eq. (85), and thus Eq. (30) follows.

Let us prove Eq. (86). Applying Lemma 29 for the random matrix E/ς and $D = 2^{10}$ gives, with probability $1 - \log^{-8}(m+n)$,

$$\frac{\|e_{n,k}^T (E^T E)^a V\|}{(1.9\varsigma\sqrt{m+n})^{2a+1}} \le Dp\|V\|_{2,\infty}\left(\frac{2}{3.61}\right)^a \le 2^{10}p\|V\|_{2,\infty} \le 2^{10}\sqrt{r}D_{U,V,p},$$

proving Eq. (86) and thus Eq. (29) after a union bound, similar to the previous case.

Let us now prove Eqs (31) and (32), focusing on the former first. Since the 2-to-∞ norm is the the largest norm among the rows, it suffices to prove Eq. (29) holds uniformly over all $k \in [n]$ for $p = \log(m+n)$. Substituting this new choice of p into the previous argument, for a fixed k, we have Eq. (29), but with probability at least $1 - (m+n)^{-4.04}$. Applying another union bound over $k \le [n]$ gives Eq. (31) with probability at least $1 - (m+n)^{-3}$. The proof of (32) is analogous. The proof of Lemma 19 is complete. ∎

Now let us handle the technical lemmas 29 and 30. The odd case (Lemma 29) is more difficult, so we will handle it first to demonstrate our technique. The argument for the even case (Lemma 30) is just a simpler version of the same technique.

C.2.1. Case 1: odd powers

Proof Without loss of generality, let $k = 1$. Let us fix $p \in \mathbb{N}$ and bound the $(2p)^{th}$ moment of the expression of concern. We have

$$\mathbf{E}\left[\left\|e_{n,1}^T (E^T E)^a E^T U\right\|^{2p}\right] = \mathbf{E}\left[\left(\sum_{l=1}^{r}(e_{n,1}^T (E^T E)^a E^T u_l)^2\right)^p\right]$$
$$= \sum_{l_1,\dots,l_p \in [r]} \mathbf{E}\left[\prod_{h=1}^{p}(e_{n,1}^T (E^T E)^a E^T u_{l_h})^2\right]. \tag{87}$$

Temporarily let $\mathcal{W}$ be the set of walks $W = (j_0 i_0 j_1 i_1 \dots i_a)$ of length $2a+1$ on the complete bipartite graph $M_{m,n}$ such that $j_0 = 1$. Here the two parts of M are $I = \{1', 2', \dots, m'\}$ and $J = \{1, 2, \dots, n\}$, where the prime symbol serves to distinguish two vertices on different parts with the same number. Let $E_W = E_{i_0 j_0} E_{i_0 j_1} \dots E_{i_{a-1} j_a} E_{i_a j_a}$. We can rewrite the final expression in the above as

$$\sum_{l_1, l_2, \dots, l_p \in [r]} \sum_{W_{11}, W_{12}, W_{21}, \dots, W_{p2} \in \mathcal{W}} \mathbf{E} \left[\prod_{h=1}^{p} E_{W_{h1}} E_{W_{h2}} u_{l_h i_{(h1)a}} u_{l_h i_{(h2)a}} \right],$$

where we denote $W_{hd} = (j_{(hd)0}, i_{(hd)0}, \dots, i_{(hd)a})$. We can swap the two summation in the above to get

$$\sum_{W_{11}, W_{12}, W_{21}, \dots, W_{p2} \in \mathcal{W}} \mathbf{E} \left[\prod_{h=1}^{p} E_{W_{h1}} E_{W_{h2}} \right] \sum_{l_1, l_2, \dots, l_p \in [r]} \prod_{h=1}^{p} u_{l_h i_{(h1)a}} u_{l_h i_{(h2)a}}.$$

The second sum can be recollected in the form of a product, so we can rewrite the above as

$$\sum_{W_{11}, W_{12}, W_{21}, \dots, W_{p2} \in \mathcal{W}} \mathbf{E} \left[\prod_{h=1}^{p} E_{W_{h1}} E_{W_{h2}} \right] \prod_{h=1}^{p} U^T_{\cdot, i_{(h1)a}} U_{\cdot, i_{(h2)a}}$$

Define the following notation:

1. $\mathcal{P}$ is the set of all *star*, i.e. tuples of walks $P = (P_1, \dots, P_{2p})$ on the complete bipartite graph $M_{m,n}$, such that each walk $P_r \in \mathcal{W}$ and each edge appears at least twice.

   Rename each tuple $(W_{h1}, W_{h2})_{h=1}^{p}$ as a star P with $W_{hd} = P_{2h-2+d}$.

   For each P, let $V(P)$ and $E(P)$ respectively be the set of vertices and edges involved in P.

   Define the partition $V(P) = V_I(P) \cup V_J(P)$, where $V_I(P) := V(P) \cap I$ and $V_J(P) := V(P) \cap J$.

2. $E_P := E_{P_1} E_{P_2} \dots E_{P_{2p}}$.

3. $P^{\text{end}} := (i_{1a}, i_{2a}, \dots, i_{(2p)a})$, which we call the *boundary* of P. Then $u_Q := \prod_{r=1}^{2p} u_{q_r}$ for any tuple $Q = (q_1, \dots, q_r)$.

4. $\mathcal{S}$ is the subset of "shapes" in $\mathcal{P}$. A shape is a tuple of walks $S = (S_1, \dots, S_{2p})$ such that all S_r start with 1 and for all $r \in [2p]$ and $s \in [0, a]$, if i_{rs} appears for the first time in $\{i_{r's'} : r' \le r, s' \le s\}$, then it is stricly larger than all indices before it, and similarly for j_{rs}. We say a star $P \in \mathcal{P}$ has shape $S \in \mathcal{S}$ if there is a bijection from $V(P)$ to $[|V(P)|]$ that transforms P into S. The notations $V(S)$, $V_I(S)$, $V_J(S)$, $E(S)$ are defined analogously. Observe that the shape of P is unique, and $\mathcal{S}$ forms a set of equivalent classes on $\mathcal{P}$.

5. Denote by $\mathcal{P}(S)$ the class associated with the shape S, namely the set of all stars P having shape S.

We can rewrite the previous sum as:

$$\sum_{P \in \mathcal{P}} \mathbf{E}\left[E_P\right] \prod_{h=1}^{p} U^T_{\cdot, i_{(2h-1)a}} U_{\cdot, i_{(2h)a}}$$

Using triangle inequality and the sub-multiplicity of the operator norm, we get the following upper bound for the above:

$$\sum_{P\in\mathcal{P}} |\mathbf{E}\left[E_P\right]| \prod_{h=1}^{p} \|U_{\cdot,i_{(2h-1)a}}\|\|U_{\cdot,i_{(2h)a}}\| = r^p \sum_{P\in\mathcal{P}} u_{P^{\text{end}}} |\mathbf{E}\left[E_P\right]| = r^p \sum_{S\in\mathcal{S}} \sum_{P\in\mathcal{P}(S)} u_{P^{\text{end}}} |\mathbf{E}\left[E_P\right]|, \tag{88}$$

where the vector u is given by $u_i = r^{-1/2}\|U_{\cdot,i}\|$ for $i \in [m]$. Observe that

$$\|u\| = 1 \quad \text{and} \quad \|u\|_\infty = r^{-1/2}\|U\|_{2,\infty}.$$

Fix $P \in \mathcal{P}$. Let us bound $\mathbf{E}\left[E_P\right]$. For each $(i,j) \in E(P)$, let $\mu_P(i,j)$ be the number of times (i,j) is traversed in P. We have

$$|\mathbf{E}\left[E_P\right]| = \prod_{(i,j)\in E(P)} \mathbf{E}\left[|E_{ij}|^{\mu_P(i,j)}\right] \le \prod_{(i,j)\in E(P)} M^{\mu_P(i,j)-2} = M^{2p(2a+1)-2|E(P)|}.$$

Since the entries u_i are related by the fact their squares sum to 1, it will be better to bound their symmetric sums rather than just a product $u_{P^{\text{end}}}$. Fix a shape S, we have

$$\begin{aligned}\sum_{P\in\mathcal{P}(S)} |u_{P^{\text{end}}}| &= \sum_{f:V(S)\hookrightarrow[m]} \prod_{k=1}^{|V(S^{\text{end}})|} |u_{f(k)}|^{\mu_{S^{\text{end}}}(k)} \le m^{|V_I(S)|-|V(S^{\text{end}})|} n^{|V_J(S)|-1} \prod_{k=1}^{|V(S^{\text{end}})|} \sum_{i=1}^{m} |u_i|^{\mu_{S^{\text{end}}}(k)} \\ &= m^{|V_I(S)|-|V(S^{\text{end}})|} n^{|V_J(S)|-1} \prod_{k=1}^{|V(S^{\text{end}})|} \|u\|_{\mu_{S^{\text{end}}}(k)}^{\mu_{S^{\text{end}}}(k)},\end{aligned}$$

where we slightly abuse notation by letting $\mu_Q(k)$ be the number of time k appears in Q.

Consider $\|u\|_l^l$ for an arbitrary $l \in \mathbb{N}$. When $l = 1$, $\|u\|_l^l \le \sqrt{m}$ by Cauchy-Schwarz. When $l \ge 2$, we have $\|u\|_l^l \le \|u\|_\infty^{l-2}\|u\|_2^2 = \|u\|_\infty^{l-2}$. Thus

$$\sum_{P\in\mathcal{P}(S)} |u_{P^{\text{end}}}| \le \prod_{k=1}^{|V(S)|} \|u\|_{\mu_{S^{\text{end}}}(k)}^{\mu_{S^{\text{end}}}(k)} \le \prod_{k\in V_2(S)} \|u\|_\infty^{\mu_{S^{\text{end}}}(k)-2} (\sqrt{m})^{|V_1(S^{\text{end}})|} = \|u\|_\infty^{2p-\nu(S)} m^{|V_1(S^{\text{end}})|/2},$$

where, we define $V_1(Q)$ as the set of vertices appearing in Q exactly once and $V_2(Q)$ as the set of vertices appearing at least twice, and to shorten the notation, we let $\nu(S) := |V_1(S^{\text{end}})| + 2|V_2(S^{\text{end}})|$. Combining the bounds, we get the upper bound below for (88):

$$\begin{aligned}&M^{2p(2a+1)} \sum_{S\in\mathcal{S}} M^{-2|E(S)|} m^{|V_I(S)|-|V(S^{\text{end}})|} n^{|V_J(S)|-1} \|u\|_\infty^{2p-\nu(S)} m^{|V_1(S^{\text{end}})|/2} \\ &= M^{2p(2a+1)+2} \sum_{S\in\mathcal{S}} M^{-2|V(S)|} m^{|V_I(S)|-\nu(S)/2} n^{|V_J(S)|-1} \|u\|_\infty^{2p-\nu(S)}.\end{aligned}$$

Suppose we fix $|V_1(S^{\text{end}})| = x$, $|V_2(S^{\text{end}})| = y$, $|V_I(S)| = z$, $|V_J(S)| = t$. Let $\mathcal{S}(x,y,z,t)$ be the subset of shapes having these quantities. To further shorten the notation, let $M_1 := M^{2p(2a+1)}\|u\|_\infty^{2p}$. Then we can rewrite the above as:

$$M_1 \sum_{x,y,z,t\in\mathcal{A}} M^{-2(z+t)} m^{z-x/2-y} n^{t-1} \|u\|_\infty^{-x-2y} |\mathcal{S}(x,y,z,t)|, \tag{89}$$

where $\mathcal{A}$ is defined, somewhat abstractly, as the set of all tuples (x,y,z,t) such that $\mathcal{S}(x,y,z,t) \neq \varnothing$. We first derive some basic conditions for such tuples. Trivially, one has the following initial bounds:

$$0 \leq x, y, \qquad 1 \leq x+y \leq z, \qquad x+2y \leq 2p, \qquad 0 \leq z, t, \qquad z+t \leq p(2a+1)+1,$$

where the last bound is due to $z+t = |V(S)| \leq |E(S)|+1 \leq p(2a+1)+1$, since each edge is repeated at least twice. However, it is not strong enough, since we want the highest power of m and n combined to be at most $2ap$, so we need to eliminate a quantity of p.

Claim 31 *When each edge is repeated at least twice, we have $z - x/2 - y + t - 1 \leq 2ap$.*

Proof [Proof of Claim 31] Let $S = (S_1, \ldots, S_{2p})$, where $S_r = j_{r0}i_{r0}j_{r1}i_{r1}\ldots j_{ra}i_{ra}$. We have $j_{r0} = 1$ for all r. It is tempting to think (falsely) that when each edge is repeated at least twice, each vertex appears at least twice too. If this were to be the case, then each vertex in the set

$$A(S) := \{i_{rs} : 1 \leq r \leq 2p, 0 \leq s \leq a-1\} \cup \{j_{rs} : 1 \leq r \leq 2p, 1 \leq s \leq a\} \cup V_1(S^{\text{end}})$$

appears at least twice. The sum of their repetitions is $4ap + x$, so the size of this set is at most $2ap + x/2$. Since this set covers every vertex, with the possible exceptions of $1 \in I$ and $V_2(S^{\text{end}})$, its size is at least $z - y + t - 1$, proving the claim. In general, there will be vertices appearing only once in S. However, we can still use the simple idea above. Temporarily let $A_1(S)$ be the set of vertices appearing once in S and $f(S)$ be the sum of all edges' repetitions in S. Let $S^{(0)} := S$. Suppose for $k \geq 0$, $S^{(k)}$ is known and satisfies $|A(S^{(k)})| = |A(S)| - k$, $f(S^{(k)}) = 4pa + x - 2k$ and each edge appears at least twice in $S^{(k)}$. If $A_1(S^{(k)}) = \varnothing$, then by the previous argument, we have

$$2(z - y + t - 1 - k) \leq 4pa + x - 2k \implies z - x/2 - y + t - 1 \leq 2pa,$$

proving the claim. If there is some vertex in $A_1(S^{(k)})$, assume it is some i_{rs}, then we must have $s \leq a-1$ and $j_{rs} = j_{r(s+1)}$, otherwise the edge $j_{rs}i_{rs}$ appears only once. Create $S^{(k+1)}$ from $S^{(k)}$ by removing i_{rs} and identifying j_{rs} and $j_{r(s+1)}$, we have $|A(S^{(k+1)})| = |A(S)| - (k+1)$ and $|f(S^{(k)}) = 4pa + x - 2(k+1)$. Further, since i_{rs} is unique, $j_{rs}i_{rs} \equiv i_{rs}j_{r(s+1)}$ are the only 2 occurences of this edge in $S^{(k)}$, thus the edges remaining in $S^{(k+1)}$ also appears at least twice. Now we only have $|A_1(S^{(k+1)})| \leq |A_1(S^{(k)})|$, with possible equality, since j_{rs} can be come unique after the removal, but since there is only a finite number of edges to remove, eventually we have $A_1(S^{(k)}) = \varnothing$, completing the proof of the claim. ■

Claim 31 shows that we can define the set $\mathcal{A}$ of *eligible sizes* as follows:

$$\mathcal{A} = \left\{(x,y,z,t) \in \mathbb{N}^4_{\geq 0} : \quad 1 \leq t; \quad 1 \leq x+y \leq z; \quad x+2y \leq 2p; \quad z - x/2 - y + t - 1 \leq 2ap\right\}. \tag{90}$$

Now it remains to bound $|\mathcal{S}(x,y,z,t)|$.

Claim 32 *Given a tuple $(x,y,z,t) \in \mathcal{A}$, where $\mathcal{A}$ is defined in Eq. (90), we have*

$$|\mathcal{S}(x,y,z,t)| \leq \frac{2^{l+1}(2p(a+1))!(2pa)!(l+1)^{2p(2a+1)-2l}}{(2p(2a+1)-2l)!l!z!(t-1)!}(16p(a+1) - 8l - 2)^{4p(a+1)-2l-1}.$$

Proof We use the following coding scheme for each shape $S \in \mathcal{S}(x,y,z,t)$: Given such an S, we can progressively build a codeword $W(S)$ and an associated tree $T(S)$ accoding to the following scheme:

1. Start with $V_J = \{1\}$ and $V_I = \varnothing$, $W = []$ and T being the tree with one vertex, 1.
2. For $r = 1, 2, \ldots, 2p$:
   (a) Relabel S_r as $1k_1k_2 \ldots k_{2a}$.
   (b) For $s = 1, 2, \ldots, 2a$:
      - If $k_s \notin V(T)$ then add k_s to T and draw an edge connecting k_{s-1} and k_s, then mark that edge with a $(+)$ in T, and append $(+)$ to W. We call its instance in S_r a *plus edge*.
      - If $k_s \in V(T)$ and the edge $k_{s-1}k_s \in E(T)$ and is marked with $(+)$: unmark it in T, and append $(-)$ to W. We call its instance in S_r a *minus edge*.
      - If $k_s \in V(T)$ but either $k_{s-1}k_s \notin E(T)$ or is unmarked, we call its instance in S_r a *neutral edge*, and append the symbol k_s to W.

This scheme only creates a *preliminary codeword* W, which does not yet uniquely determine the original S. To be able to trace back S, we need the scheme in Vu (2007) to add more details to the preliminary codewords. For completeness, we will describe this scheme later, but let us first bound the number of preliminary codewords.

Claim 33 *Let $\mathcal{PC}(x,y,z,t)$ denote the set of preliminary codewords generable from shapes in $\mathcal{S}(x,y,z,t)$. Then for $l := z+t-1$ we have*

$$|\mathcal{PC}(x,y,z,t)| \le \frac{2^l(2p(a+1))!(2pa)!(l+1)^{2p(2a+1)-2l}}{(2p(2a+1)-2l)!l!z!(t-1)!}.$$

Note that the bound above does not depend on x and y. In fact, for fixed z and t, the right-hand side is actually an upper bound for the sum of $|\mathcal{S}(x,y,z,t)|$ over all pairs (x,y) such that (x,y,z,t) is eligible. We believe there is plenty of room to improve this bound in the future.

Proof To begin, note that there are precisely z and $t-1$ plus edges whose right endpoint is respectively in I and J. Suppose we know u and v, the number of minus edges whose right endpoint is in I and J, respectively. Then

- The number of ways to place plus edges is at most $\binom{2p(a+1)}{z}\binom{2pa}{t-1}$.
- The number of ways to place minus edges, given the position of plus edges, is at most $\binom{2p(a+1)-z}{u}\binom{2pa-t+1}{v}$.
- The number of ways to choose the endpoint for each neutral edge is at most $z^{2p(a+1)-z-u}t^{2pa-t+1-v}$.

Combining the bounds above, we have

$$|\mathcal{S}(x,y,z,t)| \le \binom{2p(a+1)}{z}\binom{2pa}{t-1}\sum_{u+v=z+t-1}\binom{2p(a+1)-z}{u}\binom{2pa-t+1}{v}z^{f(z,u)}t^{g(t,v)}, \tag{91}$$

where $f(z,u) = 2p(a+1) - z - u$ and $g(u,v) = 2pa - t + 1 - v$. Let us simplify this bound. The sum on the right-hand side has the form

$$\sum_{i+j=k} \binom{N}{i}\binom{M}{j} z^i t^j,$$

where $k = 2(p(2a+1) - (z+t-1))$, $N = 2p(a+1) - z$, $M = 2pa - t + 1$. We have

$$\sum_{i+j=k} \binom{N}{i}\binom{M}{j} z^i t^j = \sum_{i+j=k} \frac{N!M!}{k!(N-i)!(M-j)!}\binom{k}{i} z^i t^j \leq \sum_{i+j=k} \frac{N!M!}{k!} \frac{(z+t)^k}{(N-i)(M-j)!}$$
$$\leq \frac{N!M!(z+t)^k}{k!(M+N-k)!} \sum_{i+j=k} \binom{M+N-k}{N-i} \leq \frac{2^{M+N-k} N!M!(z+t)^k}{k!(M+N-k)!}.$$

Replacing M, N and k with their definitions, we get

$$\sum_{u+v=z+t-1} \binom{2p(a+1)-z}{u}\binom{2pa-t+1}{v} z^{f(z,u)} t^{g(t,v)}$$
$$\leq \frac{2^{z+t-1}(2p(a+1)-z)!(2pa-t+1)!(z+t)^{2p(2a+1)-2(z+t-1)}}{(2p(2a+1)-2(z+t-1))!(z+t-1)!},$$

replacing $z+t-1$ with l, we prove the claim. ∎

Back to the proof of Claim 32, to uniquely determine the shape S, the general idea is the following. We first generated the preliminary codeword W from S, then attempt to decode it. If we encounter a plus or neutral edge, we immediately know the next vertex. If we see a minus edge that follows from a plus edge (u,v), we know that the next vertex is again u. Similarly, if there are chunks of the form $(++\ldots+--\ldots-)$ with the same number of each sign, the vertices are uniquely determined from the first vertex. Therefore, we can create a condensed codeword W^* repeatedly removing consecutive pairs of $(+-)$ until none remains. For example, the sections $(-+-+-)$ and $(-++--)$ both become $(-)$. Observe that the condensed codeword is always unique regardless of the order of removal, and has the form

$$W^* = [(+\ldots+) \text{ or } (-\ldots-)]\ (\text{neutral})\ [(+\ldots+) \text{ or } (-\ldots-)] \ldots (\text{neutral})\ [(+\ldots+) \text{ or } (-\ldots-)],$$

where we allow blocks of pure pluses and minuses to be empty. The minus blocks that remain in W^* are the only ones where we cannot decipher.

Recall that during decoding, we also reconstruct the tree $T(S)$, and the partial result remains a tree at any step. If we encounter a block of minuses in W^* beginning with the vertex i, knowing the right endpoint j of the last minus edge is enough to determine the rest of the vertices, which is just the unique path between i and j in the current tree. We call the last minus edge of such a block an *important edge*. There are two cases for an important edge.

1. If i and all vertices between i and j (excluding j) are only adjacent to at most two plus edges in the current tree (exactly for the interior vertices), we call this important

edge *simple* and just mark the it with a direction (left or right, in addition to the existing minus). For example, $(- - \dots -)$ becomes $(- - \dots (-dir))$ where dir is the direction.

2. If the edge is non-simple, we just mark it with the vertex j, so $(- - \dots -)$ becomes $(- - \dots (-j))$.

It has been shown in Vu (2007) that the fully codeword $\overline{W}$ resulting from W by marking important edges uniquely determines S, and that when the shape of S *is that of a single walk*, the cost of these markings is at most a multiplicative factor of $2(4N+8)^N$, where N is the number of neutral edges in the preliminary W. To adapt this bound to our case, we treat the star shape S as a single walk, with a neutral edge marked by 1 after every $2a+1$ edges. There are $2p-1$ additional neutral edges from this perspective, making $N = 4p(a+1)-2l-1$ in total. Combining this with the bound on the number of preliminary codewords (Claim 33) yields

$$|\mathcal{S}(x,y,z,t)| \leq \frac{2^{l+1}(2p(a+1))!(2pa)!(l+1)^{2p(2a+1)-2l}}{(2p(2a+1)-2l)!l!z!(t-1)!}(16p(a+1)-8l-2)^{4p(a+1)-2l-1},$$

where $l = z+t-1$. Claim 32 is proven. ■

Back to the proof of Lemma 29. Temporarily let

$$G_l := 2p(2a+1)-2l \quad \text{and} \quad F_l := \frac{2^{l+1}(l+1)^{G_l}}{G_l!l!}(4G_l+8p-2)^{G_l+2p-1}.$$

Note that $(2p(a+1))!(2pa)!F_l$ is precisely the upper bound on $|\mathcal{S}(x,y,z,t)|$ in Claim 32. Also let

$$M_2 = M_1(2p(a+1))!(2pa)! = M^{2p(2a+1)}(2p(a+1))!(2pa)!\|u\|_\infty^{2p}.$$

Replacing the appropriate terms in the bound in Claim 32 with these short forms, we get another series of upper bounds for the last double sum in Eq. (88):

$$\begin{aligned}
&M_2\sum_{x,y}\|u\|_\infty^{-x-2y}\sum_{l=x+y}^{\lfloor 2pa+x/2+y\rfloor} M^{-2(l+1)}F_l\sum_{z+t=l+1}\frac{m^{z-x/2-y}n^{t-1}}{z!(t-1)!}\\
&\leq M_2\sum_{x,y}\|u\|_\infty^{-x-2y}\sum_{l=x+y}^{\lfloor 2pa+x/2+y\rfloor}\frac{M^{-2(l+1)}F_l}{(l-\lfloor\frac{x}{2}\rfloor-y)!}\sum_{z+t=l+1}\binom{l-\lfloor\frac{x}{2}\rfloor-y}{z-\lfloor\frac{x}{2}\rfloor-y}m^{z-\lfloor\frac{x}{2}\rfloor-y}n^{t-1}\\
&\leq M_2\sum_{x,y}\|u\|_\infty^{-x-2y}\sum_{l=x+y}^{\lfloor 2pa+x/2+y\rfloor}\frac{M^{-2(l+1)}F_l}{(l-\lfloor\frac{x}{2}\rfloor-y)!}(m+n)^{l-\lfloor\frac{x}{2}\rfloor-y}.
\end{aligned}$$

Temporarily let C_l be the term corresponding to l in the sum above. For $l \geq x+y+1$, we have

$$\frac{C_l}{C_{l-1}} = \frac{2(m+n)(G_l+1)(G_l+2)}{M^2l^3(4G_l+8p-2)^2(l-\lfloor\frac{x}{2}\rfloor-y)}\left(1+\frac{1}{l}\right)^{G_l}\left(1-\frac{4}{2G_l+4p+3}\right)^{G_l+2p+1}.$$

The last power is approximately $e^{-2} \approx 0.135$, and for $p \geq 7$ a routine numerical check shows that it is at least $1/8$. The second to last power is at least 1. The fraction be bounded as below.

$$\frac{2(m+n)(G_l+1)(G_l+2)}{M^2 l^3 (4G_l + 8p - 2)^2 (l - \lfloor \frac{x}{2} \rfloor - y)} \geq \frac{2(m+n) \cdot 1 \cdot 2}{M^2 l^4 (8p-2)^2} \geq \frac{m+n}{16M^2 l^4 p^2} \geq \frac{m+n}{16M^2 p^6 (2a+1)^4}.$$

Therefore, under the assumption that $m+n \geq 256M^2p^6(2a+1)^4$, we have $C_l \geq 2C_{l-1}$ for all $l \geq 1$, so $\sum_l C_l \leq 2C_{l^*}$, where $l^* = \lfloor 2pa + x/2 + y \rfloor$, the maximum in the range. We have

$$\begin{aligned} 2C_{l^*} \leq 2(m+n)^{2pa} & \frac{(2M^{-2})^{2pa+\lfloor \frac{x}{2} \rfloor + y + 1}(2pa + \lfloor \frac{x}{2} \rfloor + y + 1)^{2(p - \lfloor \frac{x}{2} \rfloor - y)}}{(2(p - \lfloor \frac{x}{2} \rfloor - y))! \cdot (2pa + \lfloor \frac{x}{2} \rfloor + y)! \cdot (2pa)!} \\ & \cdot \left(16p - 8\left\lfloor \frac{x}{2} \right\rfloor - 8y - 2\right)^{4p - 2\lfloor \frac{x}{2} \rfloor - 2y - 1}. \end{aligned}$$

Temporarily let $d = p - (\lfloor \frac{x}{2} \rfloor + y)$ and $N = p(2a+1)$, we have

$$2C_{l^*} \leq 2(m+n)^{2pa} \frac{(2M^{-2})^{N-d+1}(N-d+1)^{2d}(8p+8d-2)^{2p+2d-1}}{(2pa)! \cdot (2d)! \cdot (N-d)!}.$$

For each d, there are at most $2(p-d)$ pairs (x,y) such that $d = p - (\lfloor \frac{x}{2} \rfloor + y)$, so overall we have the following series of upper bounds for the last double sum in Eq. (88):

$$\begin{aligned} & M_2(m+n)^{2pa} \sum_{d=0}^{p-1} 4(p-d)\|u\|_\infty^{-2(p-d)} \cdot \frac{(2M^{-2})^{N-d+1}(N-d+1)^{2d}(8p+8d-2)^{2p+2d-1}}{(2pa)! \cdot (2d)! \cdot (N-d)!} \\ & \leq M_3(m+n)^{2pa} \sum_{d=0}^{p-1} \|u\|_\infty^{2d} \cdot \frac{2^{-d}M^{2d}(N-d+1)^{2d}(8p+8d-2)^{2p+2d-1}}{(2d)! \cdot (N-d)!}, \end{aligned} \tag{92}$$

where

$$M_3 = 4p\frac{M_2\|u\|_\infty^{-2p}(2M^{-2})^{N+1}}{(2pa)!} = 2^{p(2a+1)+3}pM^{-2}(2p(a+1))!.$$

Let us bound the sum at the end of Eq. (92). Temporarily let A_d be the term corresponding to d and $x := 2^{-1/2}M\|u\|_\infty$. We have

$$A_d = \frac{x^{2d}(N-d+1)^{2d}}{(2d)!(N-d)!}(8p+8d-2)^{2p+2d-1} \leq \frac{x^{2d}N^{3d}}{(2d)!N!}\frac{(16p)^{2p+2d}}{8p}.$$

Therefore

$$\begin{aligned} \sum_{d=0}^{p-1} A_d & \leq \frac{(16p)^{2p}}{8pN!} \sum_{d=0}^{p-1} \frac{(16pN^{3/2}x)^{2d}}{(2d)!} \leq \frac{(16p)^{2p}}{8pN!} \sum_{d=0}^{p-1} \binom{2p}{2d} (16pN^{3/2}x)^{2d} \frac{e^{2d}}{(2p)^{2d}} \\ & = \frac{(16p)^{2p}}{8pN!}(8eN^{3/2}x+1)^{2p} \leq \frac{(16p)^{2p}}{8pN!}(16N^{3/2}M\|u\|_\infty + 1)^{2p}. \end{aligned}$$

Plugging this into Eq. (92), we get another upper bound for (88):

$$M_4(16N^{3/2}M\|u\|_\infty + 1)^{2p}(m+n)^{2ap},$$

where

$$M_4 := M_3 \frac{(16p)^{2p}}{8pN!} = 2^{p(2a+1)+3} p M^{-2} (2p(a+1))! \frac{(16p)^{2p}}{8p(2ap+p)!} \le \frac{2^{2ap} 2^{10p} p^{3p} (a+1)^p}{8M^2}.$$

To sum up, we have

$$\begin{aligned}
\mathbf{E}\left[\left\| e_{n,1}^T (E^T E)^a E^T U \right\|^{2p} \right] &\le r^p \sum_{S \in \mathcal{S}} \sum_{P \in \mathcal{P}(S)} u_{P^{\text{end}}} \left| \mathbf{E}\left[E_P \right] \right| \\
&\le \frac{r^p 2^{2ap} 2^{10p} p^{3p} (a+1)^p}{8M^2} (16 N^{3/2} M \|u\|_\infty + 1)^{2p} (m+n)^{2ap} \\
&\le \left(2^5 r^{1/2} p^{3/2} \sqrt{2a+1} (2^4 p^{3/2} (2a+1)^{3/2} M \|u\|_\infty + 1) \cdot [2(m+n)]^a \right)^{2p}.
\end{aligned}$$

Let $D > 0$ be arbitrary. By Markov's inequality, for any p such that $m + n \ge 2^8 M^2 p^6 (2a+1)^4$, the moment bound above applies, so we have

$$\left\| e_{n,1}^T (E^T E)^a E^T U \right\| \le D r^{1/2} p^{3/2} \sqrt{2a+1} (16 p^{3/2} (2a+1)^{3/2} M \|u\|_\infty + 1) [2(m+n)]^a$$

with probability at least $1 - (2^5/D)^{2p}$. Replacing $\|u\|_\infty$ with $\frac{1}{\sqrt{r}} \|U\|_{2,\infty}$, we complete the proof. ■

C.2.2. Case 2: even powers

Proof Without loss of generality, assume $k = 1$. We can reuse the first part and the notations from the proof of Lemma 29 to get the bound

$$\mathbf{E}\left[\left\| e_{n,1}^T (E^T E)^a V \right\|^{2p} \right] \le r^p \sum_{S \in \mathcal{S}} \sum_{P \in \mathcal{P}(S)} v_{P^{\text{end}}} \left| \mathbf{E}\left[E_P \right] \right|,$$

where $v_i = r^{-1/2} \|V_{\cdot,i}\|$. Again,

$$\|v\| = 1 \quad \text{and } \|v\|_\infty = r^{-1/2} \|V\|_{2,\infty},$$

and $\mathcal{S}$ is the set of shapes such that every edge appears at least twice, $\mathcal{P}(S)$ is the set of stars having shape S, and

$$E_P = \prod_{ij \in E(P)} E_{ij}^{m_P(ij)}, \quad \text{and} \quad v_Q = \prod_{j \in V(Q)} v_j^{m_Q(j)}.$$

Note that a shape for a star now consists of walks of length $2a$:

$$S = (S_1, S_2, \ldots, S_{2p}) \quad \text{where } S_r = j_{r0} i_{r0} j_{r1} i_{r1} \ldots j_{ra}.$$

We have, for any shape S and $P \in \mathcal{P}(S)$,

$$\mathbf{E}\left[E_P \right] \le M^{4pa - 2|E(S)|} \le M^{2pa - 2|V(S)| + 2}, \quad |v_{P^{\text{end}}}| \le \|v\|_\infty^{2p}, \quad \text{and } |\mathcal{P}(S)| \le m^{|V_I(S)|} n^{|V_J(S)| - 1},$$

where the power of n in the last inequality is due to 1 having been fixed in $V_J(S)$. Therefore

$$\sum_{S\in\mathcal{S}}\sum_{P\in\mathcal{P}(S)} v_{P^{\text{end}}}\left|\mathbf{E}\left[E_P\right]\right| \leq M_1 \sum_{S\in\mathcal{S}} M^{-2|V(S)|} m^{|V_I(S)|} n^{|V_J(S)|-1}, \quad \text{where } M_1 := M^{4pa+2}\|v\|_\infty^{2p}.$$

Let $\mathcal{S}(z,t)$ be the set of shapes S such that $|V_I(S)| = z$ and $|V_J(S)| = t$. Let $\mathcal{A}$ be the set of eligible indices:

$$\mathcal{A} := \Big\{(z,t) \in \mathbb{N}^2 : 0 \leq z,\ 1 \leq t, \quad \text{and } z+t \leq 2pa+1\Big\}.$$

Using the previous argument in the proof of Lemma 29 for counting shapes, we have for $(z,t)\in\mathcal{A}$:

$$|\mathcal{S}(z,t)| \leq \frac{[(2pa)!]^2 F_l}{z!\cdot(t-1)!} m^z n^{t-1}, \quad \text{where } l := z+t-1 \in [2pa],$$

where

$$G_l := 4ap - 2l \quad \text{and} \quad F_l := \frac{2^{l+1}(l+1)^{G_l}}{G_l! l!}(4G_l + 8p - 2)^{G_l+2p-1}.$$

We have

$$\begin{aligned}
\sum_{S\in\mathcal{S}}\sum_{P\in\mathcal{P}(S)} v_{P^{\text{end}}}\left|\mathbf{E}\left[E_P\right]\right| &\leq M_1 \sum_{l=0}^{2ap} M^{-2(l+1)}[(2ap)!]^2 F_l \sum_{z+t=l+1} \frac{m^z n^{t-1}}{z!\cdot(t-1)!} \\
&= M_2 \sum_{l=0}^{2ap} \frac{M^{-2l}F_l}{l!} \sum_{z+t=l+1} \binom{l}{z} m^z n^{t-1} = M_2 \sum_{l=0}^{2ap} \frac{M^{-2l}F_l}{l!}(m+n)^l,
\end{aligned}$$

where $M_2 := M_1[(2pa)!]^2 M^{-2} = M^{4ap}[(2pa)!]^2\|v\|_\infty^{2p}$. Let C_l be the term corresponding to l in the last sum above. An analogous calculation from the proof of Lemma 29 shows that under the assumption that $m+n \geq 256M^2p^6(2a)^4$, $C_l \geq 2C_{l-1}$ for each l, so $\sum_{l=0}^{2pa} C_l \leq 2C_{2pa}$, where

$$C_{2pa} = \frac{M^{-4ap}2^{2ap+1}(8p-2)^{2p-1}}{[(2ap)!]^2}(m+n)^{2ap}.$$

Therefore

$$\begin{aligned}
\mathbf{E}\left[\left\|e_{n,1}^T(E^TE)^aV\right\|^{2p}\right] &\leq r^p \sum_{S\in\mathcal{S}}\sum_{P\in\mathcal{P}(S)} v_{P^{\text{end}}}\left|\mathbf{E}\left[E_P\right]\right| \\
&\leq 2r^pM_2\frac{M^{-4ap}2^{2ap+1}(8p-2)^{2p-1}}{[(2ap)!]^2}(m+n)^{2ap} = 4\left(2^3pr^{1/2}\|v\|_\infty[2(m+n)]^a\right)^{2p}.
\end{aligned}$$

Pick $D>0$, by Markov's inequality, we have

$$\mathbf{P}\left(\left\|e_{n,1}^T(E^TE)^aV\right\| \geq Dpr^{1/2}\|v\|_\infty[2(m+n)]^a\right) \leq \left(\frac{16p}{D}\right)^{2p}.$$

Replacing $\|v\|_\infty$ with $r^{-1/2}\|V\|_{2,\infty}$, we complete the proof. ■