
Supplementary material to the paper “Fast exact recovery of noisy matrix from few entries: the infinity norm approach”

Anonymous Author(s)

Affiliation

Address

email

75 A A deeper comparison with the current methods

76 In this section, we will discuss the four approaches (nuclear norm minimization, alternating projec-
77 tions, low-rank approximation with GD, and single-step low-rank approximation) in the introduction
78 in more detail, and compare them with our method. Note that most of them are RMSE recovery
79 in the noisy setting, but we can still make comparison due to the fact that our infinity norm bound
80 automatically implies a RMSE bound with the same error margin.

81 A.1 Nuclear norm minimization and ISVT

82 This approach starts from the intuitive idea that if A is mathematically recoverable, it has to be the
83 matrix with the lowest rank agreeing with the observations at the revealed entries. Formally, one
84 would like to solve the following optimization problem:

$$\text{minimize } \text{rank } X \quad \text{subject to } X_{\Omega} = A_{\Omega}. \quad (8)$$

85 Unfortunately, this problem is NP-hard, and all existing algorithms take doubly exponential time in
86 terms of the dimensions of A [12]. To overcome this problem, Candes and Recht [13], motivated by
87 an idea from the *sparse signal recovery* problem in the field of *compressed sensing* [14, 15], proposed
88 to replace the rank with the nuclear norm of X , leading to

$$\text{minimize } \|X\|_* \quad \text{subject to } X_{\Omega} = A_{\Omega}. \quad (9)$$

89 The paper [13] was shortly followed by Candes and Tao [16], with both improvements and trade-offs,
90 and ultimately by Recht [17], who improved both previous results, proving that A is the unique
91 solution to (9), given the sampling size bound

$$|\Omega| \geq C \max\{\mu_0, \mu_1^2\} r N \log^2 N, \quad (10)$$

92 for $\mu_1 := \frac{1}{r} \sqrt{mn} \|UV^T\|_{\infty}$.

93 If one replaces μ_1 with its trivial upper bound $\mu_0 \sqrt{r}$, the RHS becomes $C \mu_0^2 r^2 N \log^2 N$. This attains
94 the optimal power of N while missing slightly from the optimal powers of r and $\log N$.

95 Candes and Plan [18] adapted to the noisy situation by relaxing the constraint on the observations,
96 leading to the following problem:

$$\text{minimize } \|X\|_* \quad \text{subject to } \|X_{\Omega} - A_{\Omega, Z}\|_F \leq \delta, \quad (11)$$

97 where δ is a known upper bound on $\|Z_{\Omega}\|_F$. The authors showed that, under the same sample size
98 condition in [17], with probability $1 - o(1)$, the optimal solution \hat{A} satisfies

$$\frac{1}{\sqrt{mn}} \|\hat{A} - A\|_F \leq C \|Z_\Omega\|_F \sqrt{\frac{\min\{m, n\}}{|\Omega|}}. \quad (12)$$

99 If one would like the RMSE to be at most ε , then one needs to require

$$|\Omega| \geq C \frac{\|Z_\Omega\|_F^2 \min\{m, n\}}{\varepsilon^2}, \quad (13)$$

100 which grows quadratically with $1/\varepsilon$.

101 For exact recovery, one needs to turn the approximation in the Frobenius norm into an approximation
 102 in the infinity norm; see subsection 1.1. This is a major mathematical challenge, and we do not know
 103 any efficient way to do this. The trivial bound that $\|M\|_\infty \leq \|M\|_F$ is too generous. In the common
 104 situation where the entries of Z have mean 0 and variance 1, if we use this and then use (12) to
 105 bound the RHS, then the corresponding bound on $|\Omega|$ in (13) becomes larger than mn , which is
 106 meaningless. This is the common situation with all Frobenius norm bounds discussed in this section.

107 Compared to our exact recovery method, they have a better sample size condition, where the powers
 108 of r and $\log N$ are lower. This disadvantage is made up for by the simplicity and efficiency of our
 109 algorithm. The standard method for solving Problem (9) is to convert it to a semidefinite program
 110 [13, 16], then solve it with interior point methods such as SDPT3 [19] and SeDuMi [20] (see [21]
 111 for more details). These methods can take up to $O(|\Omega|^2 N^2)$ FLOPs (floating point operations),
 112 even if one takes advantage of the sparsity of A_Ω . Indeed, as Ω is at least $CN \log N$ (by coupon
 113 collector), the number of operations is $\Omega(N^4)$, which is too large even for moderate N . In contrast,
 114 our algorithm, Approximate-and-Round 2, takes $O(|\Omega|r)$ FLOPs.

115 Our advantage is even clearer in the noisy case, where the bound in [18] becomes trivial in many
 116 common cases of Z (as discussed above), while ours is still close to the optimal bounds, up to powers
 117 of r and $\log N$. The only downside is our extra assumption that $\sigma_1 = \|A\|$ be large enough, which
 118 we have shown to be unavoidable for single-step SVD-based approaches; see Remark 2.4.

119 An *iterative singular value thresholding* method aiming to solve a regularized version of nuclear norm
 120 minimization, trading exactness for performance, has been proposed by Cai et al. [22]. However, this
 121 takes $O(|\Omega|r(\# \text{iterations}))$ FLOPs, and few guarantees on the number of iterations to achieve any
 122 type of error margin have been proposed.

123 A.2 Modified alternating projections

124 The intuition behind this approach is to fix the rank, then attempt to match the samples as much as
 125 possible. If we *know* $r = \text{rank } A$ precisely, it is natural to look at the following optimization problem

$$\text{minimize } \|(A - XY^T)_\Omega\|_F^2 \quad \text{over } X \in \mathbb{R}^{m \times r}, Y \in \mathbb{R}^{r \times n}. \quad (14)$$

126 This, unfortunately, like (8), is NP-hard [23]. There have been many studies proposing variants of
 127 *alternating projections*, all of which involve the following basic idea: suppose one already obtains an
 128 approximator $X^{(l)}$ of X at iteration l , then $Y^{(l)}$ and $X^{(l+1)}$ are defined by

$$Y^{(l)} := \operatorname{argmin}_{Y \in \mathbb{R}^{r \times n}} \|(A - X^{(l)}Y^T)_\Omega\|_F^2, \quad X^{(l+1)} := \operatorname{argmin}_{X \in \mathbb{R}^{m \times n}} \|(A - X(Y^{(l)})^T)_\Omega\|_F^2.$$

129 The survey [23] pointed out that these methods tend to outperform nuclear norm minimization in
 130 practice. On the other hand, there are few rigorous guarantees for recovery. The convergence and
 131 final output of the basic algorithm above also depends highly on the choice of $X^{(0)}$ [23].

132 Jain, Netrapalli and Sanghavi (2012) [24] developed one of the first alternating projections variants
 133 for matrix completion with rigorous recovery guarantees. They proved that, under the same setting in
 134 Section 1.2 and the sample size condition

$$|\Omega| \geq C\mu_0 r^{4.5} \left(\frac{\sigma_1}{\sigma_r}\right)^4 N \log N \log \frac{r}{\varepsilon},$$

the AP algorithm in [24] recovers A within an Frobenius norm error ε in $O(|\Omega|r^2 \log(1/\varepsilon))$ time with high probability. Since the Frobenius norm is larger than the infinity norm, this also gives us an exact recovery up to the error margin ε .

Compared to the previous approach, there are two new essential requirements here. First one needs to know the rank of A precisely. Second, there is a strong dependence on the condition number $\kappa := \sigma_1/\sigma_r$. Therefore, the result is effective only if κ is small.

The condition number factor was reduced to quadratic by Hardt [25] and again by Hardt and Wooters [26] to logarithmic, at the cost of an increase in the powers of r , μ_0 and $\log N$.

Concerning noisy recovery, a corollary of [26, Theorem 1] shows that we can obtain an approximation \hat{A} of rank r , where

$$\|\hat{A} - A\| \leq (2 + o(1))\|Z\| + \varepsilon\sigma_1, \quad (15)$$

given that (for sufficiently large universal constants C_1 and C_2),

$$p \geq C_1 \log^{C_2} N \left(\frac{1}{n} \left(1 + \frac{\|Z\|_F}{\varepsilon\sigma_1} \right) \right)^2.$$

The bound here is in the spectral norm, and one can translate into Frobenius norm by the fact that $\|M\|_F \leq \sqrt{\text{rank } M} \|M\|$. Again, it is not clear of how to obtain exact recovery from here.

Our approach fully removes the dependence on κ , while being simpler and even more efficient in some cases. Our time complexity is $O(|\Omega|r)$, which is worse than this approach in terms of the growth factor of $1/\varepsilon$ (observe that $|\Omega|$ grows with $1/\varepsilon^2$ in our sample size condition, but theirs has $\log 1/\varepsilon$), but this does not matter when $\varepsilon = \Theta(1)$, which is the case for most real-life datasets. For instance, the Netflix price data [27] has $\varepsilon = .5$. Our guarantees for recovery in the noisy setting is well-defined, with a well-established time complexity, while this approach, and most iterative approaches discussed here, do not have rigorous analysis on the number of iterations.

Remark A.1 (A problem with trying all possible ranks). In practice, usually we do not know the rank r exactly, but have some estimates (for instance, r is between known values r_{\min} and r_{\max}). It has been suggested (see, for instance, [28]) that one tries all integers in this range as the potential value of r . From the complexity view point, this only increases the running time by a small factor $r_{\max} - r_{\min}$, which is acceptable. However, the main trouble with this idea is that it is not clear that among the outputs, which one we should choose. If we go for exact recovery, then what should we do if there are two different outputs which agree on Ω ? We have not found a rigorous treatment of this issue in the literature.

A.3 Low rank approximation with Gradient descent

As discussed earlier, if one assumes the independent sampling model with probability p , then the rescaled sample matrix $p^{-1}A_\Omega$ can be viewed as a *random perturbation* of A . Since $\mathbf{E}[A_\Omega/p] = A$, this perturbation is unbiased, and the matrix $E := p^{-1}A_\Omega - A$ is a random matrix with mean zero.

Assuming that the rank r is known, Keshavan, Montanari and Oh [29] first use the best rank- r approximation of $p^{-1}A_\Omega$ to obtain an approximation of A . Next, they add a cleaning step, using optimization via gradient descent, to achieve exact recovery. Here is the description of their algorithm:

1. *Trimming*: first zero out all columns in A_Ω with more than $2|\Omega|/m$ entries, then zero out all rows with more than $2|\Omega|/n$ entries, producing a matrix \widetilde{A}_Ω .
2. *Low-rank approximation*: Compute the best rank- r approximation of \widetilde{A}_Ω via truncated SVD. Let $\mathbf{T}_r(\widetilde{A}_\Omega) = \widetilde{U}_r \widetilde{\Sigma}_r \widetilde{V}_r^T$ be the output.
3. *Cleaning*: Solve for X, Y, S in the following optimization problem:

$$\text{minimize } \|A_\Omega - (XSY^T)_\Omega\|_F^2 \quad \text{for } X \in \mathbb{R}^{m \times r}, Y \in \mathbb{R}^{n \times r}, S \in \mathbb{R}^{r \times r}, \quad (16)$$

using a gradient descent variant [29], starting with $X_0 = \widetilde{U}_r$, $Y_0 = \widetilde{V}_r$ and S_0 be the $r \times r$ matrix minimizing the objective function above given X_0 and Y_0 .

Let (X_*, Y_*, S_*) be the optimal solution. Output $X_* S_* Y_*^T$.

178 The last cleaning step resembles the optimization problem in alternating projections methods, but
 179 here the authors used gradient descent instead. They [29] showed that the algorithm returns an output
 180 arbitrarily close to A , given *enough* iterations in the cleaning step, provided the following sampling
 181 size condition:

$$|\Omega| \geq C \max \left\{ \mu_0 \sqrt{mn} \left(\frac{\sigma_1}{\sigma_r} \right)^2 r \log N, \quad \max\{\mu_0, \mu_1\}^2 r \min\{m, n\} \left(\frac{\sigma_1}{\sigma_r} \right)^6 \right\}. \quad (17)$$

182 It was pointed out [29] that the powers of r and $\log N$ are optimal by the coupon-collector limit,
 183 answering a question from [16]. On the other hand, the bound depends heavily on the condition
 184 number $\kappa := \sigma_1/\sigma_r$. Furthermore, similar to the situation in the previous subsection, one needs to
 185 know the rank r in advance; see Remark A.1.

186 In a later paper [28], Keshavan and Oh showed that one can compute r (with high probability) if
 187 the condition number satisfies $\kappa = O(1)$; see also Remark A.1. Thus, it seems that the critical extra
 188 assumption for this algorithm (apart from the three basic assumptions) to be efficient is that the
 189 singular values of A are of the same order of magnitude, i.e., $\kappa = O(1)$. This assumption is strong,
 190 and we do not know how often it holds in practice. In fact, Figure 1.3 shows that the centered data
 191 matrix made from the Yale face dataset has a large condition number.

192 From the complexity point of view, the first part (low rank approximation) of the algorithm is very
 193 fast, in both theory and practice, as it used truncated SVD only once. On the other hand, [29] did
 194 not provide a full convergence rate analysis of their (cleaning) gradient descent part. It only briefly
 195 mentioned that quadratic convergence is possible [29, Page 14].

196 Keshavan et al. [30] also extended their result from [29] to the noisy case, using the same algorithm.
 197 They proved that with the same sample size condition as (17), the output satisfies w.h.p.

$$\|\hat{A} - A\|_F \leq C \left(\frac{\sigma_1}{\sigma_r} \right)^2 \frac{r^{1/2} mn}{|\Omega|} \|Z_\Omega\|_{\text{op}}. \quad (18)$$

198 If one would like to have $\frac{1}{\sqrt{mn}} \|\hat{A} - A\|_F \leq \varepsilon$, this translates to the following sample size condition:

$$|\Omega| \geq C \frac{\sigma^2 r N}{\varepsilon^2} \left(\frac{\sigma_1}{\sigma_r} \right)^2, \quad (19)$$

199 where the dependence on ε is again quadratic.

200 Chatterjee [31] uses a spectral approach with a fixed truncation point independent from rank A , and
 201 thus does not require knowing it. He required that $p \geq CN^{-1} \log^6 N$ and achieved the bound

$$\mathbf{E} \left[\frac{1}{mn} \|\hat{A} - A\|_F^2 \right] \leq C \min \left\{ \sqrt{\frac{r}{p} \left(\frac{1}{m} + \frac{1}{n} \right)}, 1 \right\} + o(N).$$

202 The advantage over previous methods is the absence of the incoherence assumption. However, to
 203 translate the bound in expectation above to obtain $\frac{1}{\sqrt{mn}} \|\hat{A} - A\|_F \leq \varepsilon$ with probability at least $1 - \delta$,
 204 assuming Markov's inequality is used, one will need

$$p \geq \frac{Cr}{\varepsilon^4 \delta^2} \left(\frac{1}{m} + \frac{1}{n} \right),$$

205 The dependence on ε grows substantially faster than [30]. In fact, P. Tran's and Vu's recent result in
 206 random perturbation theory [32] can be used to prove a high-probability mean-squared-error bound,
 207 again without incoherence, requiring only a quadratic dependence on ε .

208 Again, our main advantages over their method is the simplicity, with a well-established and more
 209 optimal time complexity, the independence from κ (compared to [29, 30]), or the more optimal
 210 growth factor of $1/\varepsilon$ (compared to [33]), and the exactness of recovery. Our main drawback is the
 211 extra assumption on $\|A\|$, which is the best our method can do (Remark 2.4).

A.4 Single-step Low-rank approximation with rounding-off

In this approach, one exploits the fact that A has finite precision (which is a necessary assumption for exact recovery to make sense); see subsection 1.1. It is clear that 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 infinity norm. It has been shown, under different extra assumptions, that low rank approximation fulfils this purpose.

The first infinity norm result was obtained by Abbe, Fan, Wang, and Zhong [10]. They showed that the best rank- r approximation of $p^{-1}A_\Omega$ is close to A in the infinity norm [10, Theorem 3.4]. Technically, they proved that if $p \geq 6N^{-1} \log N$, then

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

for some universal constant C , provided $\sigma_r \geq C\kappa\|A\|_\infty\sqrt{\frac{N \log N}{p}}$, where $\kappa = \sigma_1/\sigma_r$ is the condition number.

If we turn this result into an algorithm (by simply rounding off the approximation), then we face the same two issues discussed in the previous subsection. The algorithm needs to know the rank r , and the condition number κ has to be small. As discussed before, this boils down to the strong assumption that the condition number is bounded by a constant ($\kappa = O(1)$).

Eliminating the condition number. Very recently, Bhardwaj and Vu [11] used different mathematical tools to analyze a slightly different algorithm. In their analysis, they do not need to know the rank of A . Next, their bound on $|\Omega|$ does not include the condition number κ . Thus, they completely eliminated the role of the condition number. However, the cost here is that they need a new assumption on the gaps between consecutive singular values.

As this work is the closest to our new result, let us state their result for matrices with integer entries (the precision $\varepsilon_0 = 1$). One can reduce the case of general case to this by scaling.

Algorithm A.2 (Approximate-and-Round (AR)).

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.

They showed that with probability $1 - o(1)$, before the rounding step, $\|\hat{A} - A\|_\infty < 1/2$, guaranteeing an exact recovery of A , under the following assumptions:

- *Low-rank:* $r = O(1)$.
- *Incoherence:* $\mu = O(1)$.
- *Sampling density:* $p \geq N^{-1} \log^{4.03} N$.
- *Bounded entries:* $\|A\|_\infty \leq K_A$ for a known constant K_A .
- *Gaps between consecutive singular values:* $\min_{i \in [s]} (\sigma_i - \sigma_{i+1}) \geq Cp^{-1} \log N$.

Aside from the first three basic assumptions, the new assumption that the entries are bounded is standard for real-life datasets. In the step of finding the threshold, it seems that one needs to know both r and μ , but a closer look at the analysis reveals that it is possible to relax to knowing only their upper bounds. (We will do exactly this in our algorithm, which is a variant of **AR**.)

For exact recovery, this is the only approach which adapts well to the noisy situation. As a matter of fact, the infinity norm bounds presented above hold in both noiseless and noisy case (with some modification). The reason is this: even in the noiseless case, one already views the (rescaled)

input matrix $p^{-1}A_\Omega$ as the sum of A and a random matrix E . Thus, adding a new noise matrix Z just changes E to $E + Z$. This changes few parameters in the analysis, but the key mathematical arguments remain valid. In fact, in the noisy case, both approaches above simply replace $\|A\|_\infty$ with $\|A\|_\infty + \|Z\|_\infty$ in their respective bounds.

As discussed, the main improvement of Bhardwaj and Vu [11] over the previous spectral approaches is the removal of the dependence on the condition number. One no longer need $\kappa = O(1)$. This removal was based on an entirely different mathematical analysis, which shows that the leading singular vectors of A and $p^{-1}A_\Omega$ are close in the infinity norm.

In the new assumption on the gaps, the the required bound for the gaps is relatively mild (weaker than what one requires for the application of Davis-Kahan theorem; see [11] for more discussion). However, we do not know how often matrices in practice satisfy it. In fact, Figure 1.3 in the main body shows that the Yale face database matrix [34] has several steep drops in singular value gaps.

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 it 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 has been resolved by our result in the main body of the paper.

Therefore, our advantage over [11] is the independence from the singular value gaps, and our advantage over [10] is the independence from the condition number. We retain the simplicity and efficiency of the single-step SVD approach in these papers.

A slight downside is we do have slightly worse factors of r and $\log N$ in the bounds.

A.5 A brief summary

We have the best rigorously analyzed time complexity, being on par with other single-step SVD approaches [10, 11]. Our approach yields exact recovery, which is more refined than the RMSE approaches (for relevant applications). Our assumptions for recovery does not depend on the condition number or the singular value gaps, as opposed to various past works. While there were works that have no such dependence before (e.g. by Chatterjee [31], Bhattacharya and Chatterjee [33]), our growth factors in the sample size assumption is noticeable better.

B The main matrix perturbation theorems

This section serves two purposes.

First, we will formally introduce the main technical theorems that form the backbone of our argument. The two main theorems are Theorem B.2, an extension of the classic Davis-Kahan theorem for the perturbation of low-rank approximations, in the infinity norm; and Theorem B.4, a semi-isotropic bound that serves an essential role in the contour integral method used to prove Theorem B.2. While both are novel, Theorem B.2 can be deduced easily with the argument in [32], while Theorem B.4 requires an entirely separate proof with the moment method. We will introduce their corollary, Theorem B.6, that serves as a “random” version of Theorem B.2. We tend to use this theorem directly in applications rather than the previous two. Along the way, we will argue that the bounds of Theorem B.6 are nearly optimal, up to a few $\log N$ and r factors.

Second, we will provide the full proofs of Theorems 2.3 and 3.1, in the following chain:

$$\text{Theorem B.6 (assumed)} \xrightarrow{\text{implies}} \text{Theorem 3.1} \xrightarrow{\text{implies}} \text{Theorem 2.3}.$$

B.1 Davis-Kahan in the infinity norm: the deterministic version

At this point, we can put aside the matrix completion problem and focus on the perturbation theory view point. Let us formally introduce the objects involved below.

Setting B.1 (Matrix perturbation). Consider two $m \times n$ matrices: the *original or pure matrix* A , and the *noise or perturbation matrix* E . Let $\tilde{A} := A + E$ be the *noisy or perturbed matrix*. Let A have the SVD $A = U\Sigma V^T = \sum_{i=1}^r \sigma_i u_i v_i^T$, where $\sigma_1 \geq \sigma_2 \geq \dots \sigma_r$. Define the following for A :

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

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

Define analogous notations $\tilde{\sigma}_i, \tilde{u}_i, \tilde{v}_i, \tilde{\delta}_k, \tilde{\Delta}_k, \tilde{\sigma}_S, \tilde{\Delta}_S, \tilde{V}_S, \tilde{U}_S$, and \tilde{A}_S for \tilde{A} . When $S = [s]$ for some $s \in [r]$, we also use V_s, U_s, A_s in place of the three above.

Some extra notation. To aid the presentation, for every $a, b \in \mathbb{R}$, let $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$. Let $[[a, b]] := \{x \in \mathbb{Z} : a \leq x \leq b\}$ and $[a] := [[1, a]]$.

As mentioned in the previous section, one of the most well-known results in perturbation theory is the **Davis-Kahan sin Θ theorem**, proven by Davis and Kahan [35], 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 [36], states that:

$$\|\tilde{U}_S \tilde{U}_S^T - U_S U_S^T\| \vee \|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\| \leq \frac{C\|E\|}{\Delta_S}. \quad (20)$$

A key observation is that the worst case (equality) only happens when there are special interactions between E and A . A series of papers by O'Rourke et al. [37], Tran and Vu [32] exploited the improbability of such interactions when E is random and A has low rank, and improved the bound significantly. The former proved the following:

$$\|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\| \leq C\sqrt{|S|} \left(\frac{\|E\|}{\sigma_S} + \frac{\sqrt{r}\|U^T E V\|_\infty}{\Delta_S} + \frac{\|E\|^2}{\Delta_S \sigma_S} \right), \quad (21)$$

with high probability, effectively turning the *noise-to-gap* on the right-hand side of Eq. (20) into the much smaller *noise-to-signal ratio*. The latter then improved the third term, at the cost of an extra factor of \sqrt{r} , which does not matter when $r = O(1)$. They showed that if

$$\frac{\|E\|}{\sigma_S} \vee \frac{2r\|U^T E V\|_\infty}{\Delta_S} \vee \frac{\sqrt{2r}\|E\|}{\sqrt{\Delta_S \sigma_S}} \leq \frac{1}{8}, \quad (22)$$

then

$$\|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\| \leq CrR_S, \text{ where } R_S := \frac{\|E\|}{\sigma_S} + \frac{2r\|U^T E V\|_\infty}{\Delta_S} + \frac{2ry}{\Delta_S \sigma_S}, \quad (23)$$

where

$$y := \frac{1}{2} \max_{i \neq j} (|u_i^T E E^T u_j| + |v_i^T E^T E v_j|) \quad (24)$$

Their key improvement is replacing $\|E\|^2$ in the previous result with the smaller term y , which can be much smaller in many cases, notably when E is *regular* [32].

Our first main theorem can be seen as the infinity norm version of this result.

Theorem B.2. Consider the objects in Setting B.1. Define the following terms:

$$\begin{aligned} \tau_1 &:= \max_{0 \leq a \leq 10 \log(m+n)} \frac{1}{\sqrt{r}} \max \left\{ \frac{\|(EE^T)^a U\|_{2,\infty}}{\|E\|^{2a}}, \frac{\|(EE^T)^a E V\|_{2,\infty}}{\|E\|^{2a+1}} \right\}, \\ \tau_2 &:= \max_{0 \leq a \leq 10 \log(m+n)} \frac{1}{\sqrt{r}} \max \left\{ \frac{\|(E^T E)^a V\|_{2,\infty}}{\|E\|^{2a}}, \frac{\|(E^T E)^a E^T U\|_{2,\infty}}{\|E\|^{2a+1}} \right\}. \end{aligned} \quad (25)$$

Suppose an arbitrary subset $S \subset [r]$ satisfies Eq. (22). Then for a universal constant C ,

$$\|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\|_\infty \leq C\tau_1^2 r R_S, \quad (26)$$

$$\|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\|_{2,\infty} \leq C\tau_1 r R_S, \quad (27)$$

where R_S is defined in Eq. (23). When $S = [s]$ for some $s \in [r]$, we also have

$$\|\tilde{A}_s - A_s\|_\infty \leq C\tau_1 \tau_2 \sigma_s r R_s, \text{ where } R_s := R_{[s]}. \quad (28)$$

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

We use Eq. (28) to prove Theorem 3.1. While Eqs. (26) and (27) are not directly used in this paper, they are the best known infinity norm estimates for these perturbations of spectral quantities and may be useful in other applications.

One can clearly see that the parameters τ_1 and τ_2 play the roles of the coherence parameters from Eq. (2). They are needed to extend the spectral norm bounds in [32] to the ∞ - and 2-to- ∞ -norm bounds. The best possible values for them are respectively $\|U\|_{2,\infty}/\sqrt{r}$ and $\|V\|_{2,\infty}/\sqrt{r}$. To estimate them, we need the semi-isotropic bounds of the theorem in the next part.

Let us end this part with a comment on the optimality of the term R_S in Eq. (23).

Remark B.3 (Sharpness of R_S). Consider the term R_S :

$$R_S = \frac{\|E\|}{\sigma_s} + \frac{2r\|U^T E V\|_\infty}{\delta_s} + \frac{2ry}{\delta_s \sigma_s}.$$

The discussion in [32] shows that R_S is an optimal bound for $\|\tilde{V}_S V_S^T - V_S V_S^T\|$, up to the factor r . The first term is clearly optimal due to the Davis-Kahan theorem in the case $r = 1$. The second and third terms can be shown to be non-removable as part of the power series expansion that we will demonstrate in the proof (Section C.2).

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 (22). This bound was discovered by O'Rourke et al. [37] and was the best-known until [32]. In many cases, notably when E is a *stochastic/regular random matrix*, namely, there is a common ς such that, for all $i \in [m]$ and $j \in [n]$,

$$\varsigma = \frac{1}{m} \sum_{k=1}^m \mathbf{E} [|E_{kj}|^2] = \frac{1}{n} \sum_{l=1}^n \mathbf{E} [|E_{il}|^2],$$

y can be much smaller than $\|E\|$ (see [32] for a detailed computation of y).

B.2 The semi-isotropic bounds on random matrix powers

Below is the main theorem of this part, the full form of the semi-isotropic bound in Section 3.2.

Theorem B.4. Suppose E is a random $m \times n$ matrix with independent entries satisfying:

$$\mathbf{E} [E_{ij}] = 0, \quad \mathbf{E} [|E_{ij}|^2] \leq \varsigma^2, \quad \mathbf{E} [E_{ij}^l] \leq M^{l-2} \varsigma^l \quad \text{for all } l \in \mathbb{N}_{\geq 2}. \quad (29)$$

Let $N = m + n$ and $\mathcal{H} := 1.9\varsigma\sqrt{N}$. For each $U \in \mathbb{R}^{m \times n}$ and $p > 0$, define

$$\tau_0(U, p) := \frac{p\|U\|_{2,\infty}}{\sqrt{r}}, \quad \tau_1(U, p) := \frac{Mp^3\|U\|_{2,\infty}}{\sqrt{rN}} + \frac{p^{3/2}}{\sqrt{N}}. \quad (30)$$

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

$$\max_{0 \leq a \leq t \log N} \|e_{m,k}^T (EE^T)^a U\| \leq \tau_0(U, \log \log N) \mathcal{H}^{2a} \sqrt{r}. \quad (31)$$

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

$$\max_{0 \leq a \leq t \log N} \|e_{n,k}^T (E^T E)^a E^T U\| \leq \tau_1(U, \log \log N) \mathcal{H}^{2a+1} \sqrt{r}. \quad (32)$$

If the stronger bound $M \leq ct^{-2}N \log^{-5} N$ holds, then with probability $1 - O(N^{-2})$,

$$\max_{0 \leq a \leq t \log N} \max_{k \in [m]} \|e_{m,k}^T (EE^T)^a U\| \leq \tau_0(U, \log N) \mathcal{H}^{2a} \sqrt{r}, \quad (33)$$

$$\max_{0 \leq a \leq t \log N} \max_{k \in [n]} \|e_{n,k}^T (E^T E)^a E^T U\| \leq \tau_1(U, \log N) \mathcal{H}^{2a+1} \sqrt{r} \quad (34)$$

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

To the best of our knowledge, there has been no well-known isotropic, semi-isotropic, or even entry-wise bounds of powers of a random matrix in the literature. This theorem is thus another noteworthy contribution of this paper and may be of independent interest.

We only use Eqs. (33) and (34) to prove Theorem 3.1, but for the sake of potential future applications, we still present Eqs. (31) and (32), whose bounds are better but non-uniform in k . More specifically, we use the following bounds, which directly results from the theorem.

$$\tau_1 \leq \frac{(\sqrt{\mu_0} + \sqrt{\log N}) \log N}{\sqrt{m}}, \quad \tau_2 \leq \frac{(\sqrt{\mu_0} + \sqrt{\log N}) \log N}{\sqrt{n}}.$$

As mentioned, τ_1 and τ_2 in Theorem B.2 play the roles of the coherence parameters in the matrix completion setting. In practice, one replaces them with upper bounds when applying Theorem B.2, as the theorem still works after such substitutions. Let us show that the choices of τ_1 and τ_2 in Theorem B.4 are nearly optimal upper bounds for them.

Remark B.5 (Sharpness of τ_1, τ_2). A trivial choice of upper bounds is $\tau_1 = \tau_2 = 1/\sqrt{r}$, since we have

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

and analogously for V . This is the best estimate in the worse case for a deterministic E .

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. (25) 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 Eq. (22), one gets, in philosophy, the following bounds from Theorem B.2 (when $r = O(1)$):

$$\begin{aligned} \|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\|_\infty &\leq C \frac{\mu(V)}{n} \|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\|, \\ \|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\|_{2,\infty} &\leq C \sqrt{\frac{\mu(V)}{n}} \|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\|, \\ \|\tilde{A}_s - A_s\|_\infty &\leq C \sqrt{\frac{\mu(U)\mu(V)}{mn}} \|\tilde{A}_s - A_s\|. \end{aligned} \tag{35}$$

These are the best possible bounds one can hope to produce with Theorem B.2. *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. (26) asserts that

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

On the other hand, we have

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

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. (26) sharp. The sharpness (in the best case) of Eq. (27) and Eq. (28) can be argued similarly. Notice that this also fully justifies the optimality of the bound (7) in Theorem 3.1, up to polylogarithmic factors.

In the next section, we will rigorously prove Theorems 2.3 and 3.1.

B.3 The random theorem

Combining Theorem B.4 with familiar bounds on $\|E\|$ and $\|U^T EV\|_\infty$, we get the following “random version” of Theorem B.2. Theorem 3.1 is a direct consequence of this theorem.

Theorem B.6. *Consider the objects in Setting B.1. Let $\varepsilon \in (0, 1)$ be arbitrary. Suppose E is a random $m \times n$ matrix with independent entries following the model (29) with parameters M and ς . Let $N = m + n$. Replace τ_1 from Eq. (25) with*

$$\tau_1 := \frac{\|U\|_{2,\infty} \log N}{\sqrt{r}} + \frac{M \|V\|_{2,\infty} \log^3 N}{\sqrt{rN}} + \frac{\log^{3/2} N}{\sqrt{N}}, \tag{36}$$

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

$$\frac{\varsigma\sqrt{N}}{\sigma_S} \vee \frac{r\varsigma(\sqrt{\log N} + M\|U\|_\infty\|V\|_\infty \log N)}{\Delta_S} \vee \frac{\varsigma\sqrt{rN}}{\sqrt{\Delta_S}\sigma_S} \leq \frac{1}{16}. \quad (37)$$

Let us replace the term R_S in Eq. (23) with

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

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

$$\|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\|_\infty \leq C\tau_2^2 r R_S + \frac{1}{N}, \quad (38)$$

$$\|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\|_{2,\infty} \leq C\tau_2 r R_S + \frac{1}{N}. \quad (39)$$

Analogous bounds for U and \tilde{U} hold, with τ_1 replacing τ_2 . When $S = [s]$ for some $s \in [r]$, we slightly abuse the notation to let $R_s := R_{[s]}$. Then with probability $1 - O(N^{-1})$,

$$\|\tilde{A}_s - A_s\|_\infty \leq C\tau_1\tau_2 r \sigma_s R_s + \frac{1}{N}. \quad (40)$$

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 $1 - \varepsilon - O(N^{-1})$.

B.4 Proof of Theorem 3.1 (using Theorem B.6)

We now move to the proof of Theorem 3.1. We treat Theorem B.6 as a black box. Its proof, along with the proofs of other main theorems, will be in Appendix C.

Proof of Theorem 3.1. 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 B.6, specifically Eq. (40) for this choice of ς and M if the condition (37) holds. We check it for $S = [s]$. Given that $\sigma_s \geq \delta_s \geq 40K\sqrt{rN/p}$, we have

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

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

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

It remains to transform the right-hand side of Eq. (40) to the right-hand side of Eq. (7). 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

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

which is the first factor on the right-hand side of Eq. (7).

Consider the term R_s . From the above, we have

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

Since $\delta_s \geq 40K\sqrt{rN/p}$, the fourth term is absorbed by the first term. Removing it recovers exactly the second factor on the right-hand side of Eq. (7). The proof is complete. \square

B.5 Proof of Theorem 2.3

In this section, we prove Theorem 2.3. We will first assume Theorem 3.1 as a black box, then prove Theorem 3.1 in the next subsection. It suffices to prove Theorem 2.3 in its full form, where the sampling condition (5) replaces (3) and the condition $r_{\max} \leq \log^2 N$ is removed.

Proof of the full Theorem 2.3. Let $C_2 = 1/c$ for the constant c in Theorem B.6. We rewrite the assumptions below:

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

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

$$p \geq \frac{Cr^4r_{\max}\mu_0^2K_{A,Z}^2}{\varepsilon^2} \left(\frac{1}{m} + \frac{1}{n} \right), \quad (41)$$

$$p \geq C \left(\frac{1}{m} + \frac{1}{n} \right) \log^{10} N, \quad (42)$$

$$p \geq \frac{Cr^3K_{A,Z}^2}{\varepsilon^2} \left(1 + \frac{\mu_0^2}{\log^2 N} \right) \left(1 + \frac{r^3 \log N}{N} \right) \left(\frac{1}{m} + \frac{1}{n} \right) \log^6 N. \quad (43)$$

Let $\rho := \hat{p}/p$. From the sampling density assumption, a standard application of concentration bounds [38, 39] guarantees that, with probability $1 - O(N^{-2})$,

$$0.9 \leq 1 - \frac{1}{\sqrt{N}} \leq 1 - \frac{\log N}{\sqrt{pmn}} \leq \rho \leq 1 + \frac{\log N}{\sqrt{pmn}} \leq 1 + \frac{1}{\sqrt{N}} \leq 1.1. \quad (44)$$

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

$$\|E\| \leq 2\kappa\sqrt{N}, \quad (45)$$

with probability $1 - O(N^{-1})$. See [8, 7], [40, Lemma A.7] or [8] for detailed proofs. Therefore we can assume both Eqs. (44) and (45) 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

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

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:

$$\sigma_{s+1} \leq r\delta_{s+1} \leq r(\tilde{\delta}_{s+1} + 2\|E\|) \leq r(C_0\rho^{-1/2}\sqrt{r_{\max}N} + 4)\kappa\sqrt{N} \leq 2rC_0K_{A,Z}\sqrt{r_{\max}N/p}. \quad (46)$$

For each fixed indices j, k , we have

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

where the last inequality comes from the assumption (41) if C is large enough. Since this holds for all pairs (j, k) , we have $\|A_s - A\|_\infty \leq \varepsilon/4$.

437
438
439

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 B.6, 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 :

$$\sigma_s \geq \delta_s \geq \tilde{\delta}_s - 2\|E\| \geq C_0 \rho^{-1/2} \kappa \sqrt{r_{\max} N} - 2\|E\| \geq .9C_0 \kappa \sqrt{r_{\max} N}. \quad (47)$$

440

Consider the condition (37). If it holds, then we can apply Theorem B.6. We want

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

441
442
443

By Eq. (47), 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

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

444

which is equivalent to

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

445

which easily holds. Therefore we can apply Theorem B.6. We get, for a constant C_1 ,

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

446

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

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

447

where the first inequality comes from (42) if C is large enough. Similarly,

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

448

Therefore,

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

449

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 (43), 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 B.6,

$$\begin{aligned} \|\tilde{A}_s - A_s\|_\infty &\leq \log^2 N \cdot \frac{\log N + 4\mu_0}{\sqrt{mn}} \cdot \frac{\sqrt{2}r^{3/2}K_{A,Z}}{\sqrt{p}} \left(\sqrt{N} + r^{3/2} \sqrt{\log N} \right) \\ &\leq \sqrt{\frac{2r^3 K_{A,Z}^2 \log^6 N}{p} \left(1 + \frac{4\mu_0^2}{\log^2 N} \right) \left(1 + \frac{r^3 \log N}{N} \right) \left(\frac{1}{m} + \frac{1}{n} \right)} \leq \varepsilon/4. \end{aligned} \quad (48)$$

where the last inequality comes from the condition (43) if C is large enough.

After the two steps above, we obtain $\|\tilde{A}_s - A\|_\infty \leq \varepsilon/2$ with probability $1 - O(N^{-1})$. Finally, we get, using Fact (44) 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. \square

C Proof of main results

As mentioned, Theorem B.6 is a corollary of B.2 when the noise matrix is random. In actuality, Theorem B.2 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 B.2 imply Theorem B.6 are trivial, and will be discussed when we prove the latter.

Proof structure. First, we will assume Theorem B.2 and use it to prove Theorem B.6, which directly implies Theorem 3.1. 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 B.2. The proof adapts the contour integral technique in [32], but with highly non-trivial adjustments to handle the infinity norm, instead of spectral norm as in [32]. The proof roughly has two steps:

1. Rewrite the quantities on the left-hand sides of the bounds in Theorem B.2 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 C.7, will be general enough to imply all three of bounds of Theorem B.2.

The structure for this section will be:

$$\text{Theorem B.6} \xleftarrow{\text{implied by}} \text{Theorem B.2} \xleftarrow{\text{implied by}} \text{Lemma C.7}.$$

C.1 The random version: Proof of Theorem B.6

In this section, we prove Theorem B.6, assuming Theorem B.2. 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 (22). 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 (29), with the assumption $M \leq (m+n)^{1/2} \log^{-5}(m+n)$, the moment argument in [7] 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 B.6, the naive bound $\|E\|^2$ suffices.

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

Lemma C.1. *Consider the objects in Setting B.1. Let $E \in \mathbb{R}^{m \times n}$ be a random matrix satisfying Model (29) with parameters M and ς . Suppose $M \leq (m+n)^{1/2} \log^{-3}(m+n)$. Then with probability $1 - O((m+n)^{-2})$, all of the following hold:*

$$\|E\| \leq 1.9\varsigma\sqrt{m+n} \leq 2\varsigma\sqrt{m+n}, \quad (49)$$

$$\max_{i \neq j} (|u_i E E^T u_j| + |v_i E^T E v_j|) \leq 2\|E\|^2 \leq 8\varsigma^2(m+n). \quad (50)$$

$$\max_{i,j} |u_i^T E v_j| \leq 2\varsigma(\sqrt{\log(m+n)} + M\|U\|_\infty\|V\|_\infty \log(m+n)). \quad (51)$$

Proof. Eq. (49) follows from the moment argument in [7]. Eq. (50) follows from Eq. (49). It remains to check Eq. (51). 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 $|Y_{kh}| \leq \|U\|_\infty \|V\|_\infty |E_{kh}|$. 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 $(k, h) \in [m] \times [n]$,

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

We also have

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

By Bernstein's inequality [39], we have for all $t > 0$

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

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

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

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

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

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. (51) and the lemma is complete. \square

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. (25). To overcome this, notice that the argument in Theorem B.2 works in the same way if, instead of being rigidly refined by Eq. (25), τ_1 and τ_2 are any real numbers satisfying

$$\begin{aligned}\tau_1 &\geq \max_{a \in [[0, 10 \log(m+n)]]} \frac{1}{\sqrt{r}} \max \left\{ \frac{\|(EE^T)^a U\|_{2,\infty}}{\mathcal{H}^{2a}}, \frac{\|(EE^T)^a EV\|_{2,\infty}}{\mathcal{H}^{2a+1}} \right\}, \\ \tau_2 &\geq \max_{a \in [[0, 10 \log(m+n)]]} \frac{1}{\sqrt{r}} \max \left\{ \frac{\|(E^T E)^a V\|_{2,\infty}}{\mathcal{H}^{2a}}, \frac{\|(E^T E)^a E^T U\|_{2,\infty}}{\mathcal{H}^{2a+1}} \right\},\end{aligned}\tag{52}$$

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

From this point, we will discard Eq. (25) and treat $(\tau_1, \tau_2, \mathcal{H})$ as any tuple that satisfies Eq. (52). Specifically, we will choose $\tau_0(U)$, $\tau_1(U)$, $\tau_0(V)$, $\tau_1(V)$ such that

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

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

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

uniformly over all choices for $k \in [m]$, $l \in [n]$ and $0 \leq a \leq 10 \log(m+n)$, motivating Theorem B.4. We will treat it as a black box the the sake of this proof. The proof of Theorem B.4 will be in the last subsection. Let us prove Theorem B.6 now using Theorem B.4 and Lemma C.1.

Proof of Theorem B.6. Consider the objects from Setting B.1. We aim to apply Theorem B.2. By Lemma C.1, with probability $1 - O((m+n)^{-1})$, we can replace condition (22) in Theorem B.2

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

with condition (37) in Theorem B.6

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

Assume (37) holds, then (22) also hold and we can now apply Theorem B.2. 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 Theorem B.4. These terms match exactly with τ_1 and τ_2 from the statement of Theorem B.6. If they also matched τ_1 and τ_2 in Theorem B.2, the proof would be complete. However, they do not.

Let $\mathcal{H} := 2\varsigma\sqrt{m+n}$, then $\mathcal{H} \geq \|E\|$ by Lemma C.1. Per the discussion around the condition (52) above, if we can show that τ_1 , τ_2 and \mathcal{H} satisfy (52), then the argument in Theorem B.2 still works. By Theorem B.4 for $t = 10$, (52) holds with probability $1 - O((m+n)^{-2})$, so the proof is complete. \square

In the next section, we prove Theorem B.2. The proof is an adaptation of the main argument in [32] for the SVD. While this adaptation is easy, it has several important adjustments, sufficient to make Theorem B.2 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 entire proof.

533 C.2 The deterministic version: Proof of Theorem B.2

534 In this section, we provide the proof of Theorem B.2.

535 Given A and $\tilde{A} = A + E$, there are three terms we need to bound, corresponding to Eqs. (26), (27)
536 and (28):

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

537 The strategy of bounding all three are almost identical, and is an extension to the SVD case of the
538 strategy for the eigendecomposition in [32].

539 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
540 example, consider the former. If one views A and thus U and V as fixed, the above can be viewed
541 as a function $f(E)$ satisfying $f(0) = 0$. The difficulty comes from the fact that we cannot (yet)
542 express this function as an arithmetic combination of basic functions, which is often what is needed
543 to analyze it in depth.

544 One basic idea to rewrite this function in a tractable form is to find a tractable form for the function
545 $g : A \mapsto V V^T$, and write

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

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

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

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

551 The authors of [32] came up with a clever idea. Imagine first, for simplicity, that both A and E are
552 square symmetric matrices, and that V and \tilde{V} contain the eigenvector, rather than singular vectors,
553 of their respective matrices. In other words, $U = V$ and the numbers σ_i are temporarily viewed as
554 eigenvalues. Instead of measuring the difference $g(\tilde{A}) - g(A)$ directly, they considered the difference
555 of the *Stieltjes transforms*, and obtained the expansion:

$$(zI - \tilde{A})^{-1} - (zI - A)^{-1} = \sum_{\gamma=1}^{\infty} [(zI - A)^{-1} E]^\gamma (zI - A)^{-1}. \quad (53)$$

556 It is easy to show that this identity hold whenever the right-hand side converges. Conveniently, the
557 convergence is also guaranteed by the condition (22) of Theorem B.2, as we will see later. To obtain
558 $\tilde{V}_S \tilde{V}_S^T$ and $V_S V_S^T$, rewrite the left-hand side of Eq. (53) as

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

559 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
560 right-hand side of the expansion converges for every point on that contour, one will be able to integrate
561 over Γ_S and obtain the power series expansion

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

562 The precise details on how to choose this contour can be found in [32]. The final steps to bound the
563 left-hand side will be:

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

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

566 Back to the context in this paper, where we handle the SVD instead of the eigendecomposition. In
 567 [32], the author used this expansion to obtain a bound on the spectral norm of the left-hand side by
 568 bounding each term in the series. We make appropriate adjustments to their argument to adapt it to
 569 the SVD, while also proving a novel *semi-isotropic* bound on powers of random matrices to extend
 570 the result to the infinity norm.

571 C.2.1 The power series expansion for the SVD case

572 Firstly, let us introduce the symmetrization trick, which translates the SVD into an eigendecomposition.
 573 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
 574 the *symmetrized version* of A :

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

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

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

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

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

577 as an eigendecomposition of A_{sym} . We have

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

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

- 580 • W_S is the matrix whose columns are $\{w_i : |i| \in S\}$. Note that the conventional notation
 581 would just be $\{w_i : i \in S\}$.
- 582 • $(A_{\text{sym}})_S = W_S \Lambda W_S^T = \sum_{|i| \in S} \sigma_i w_i w_i^T$. This way $(A_{\text{sym}})_S = (A_S)_{\text{sym}}$. The conventional
 583 notation would only involve half of the sum.
- 584 • 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
 585 best rank- $2s$ approximation of A_{sym} , as opposed to the conventional meaning of the notation.
- 586 • For convenience, let $\sigma_0 := 0$.

587 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. (53) for the symmetric
 588 case, we have the expansion

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

589 which is equivalent to

$$\sum_i \frac{\tilde{w}_i \tilde{w}_i^T}{z - \tilde{\sigma}_i} - \sum_{|i| \in [r]} \frac{w_i w_i^T}{z - \sigma_i} = \sum_{\gamma=1}^{\infty} [(zI - A_{\text{sym}})^{-1} E_{\text{sym}}]^{\gamma} (zI - A_{\text{sym}})^{-1}. \quad (54)$$

590 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}
 591 and W , satisfying that the right-hand side of Eq. (54) converges for every z on the contour. Integrating
 592 over Γ_S of both sides and dividing by $2\pi i$, we have

$$\begin{aligned} & \begin{bmatrix} \tilde{U}_S \tilde{U}_S^T - U_S U_S^T & 0 \\ 0 & \tilde{V}_S \tilde{V}_S^T - V_S V_S^T \end{bmatrix} = \tilde{W}_S \tilde{W}_S - W_S W_S^T \\ & = \sum_{\gamma=1}^{\infty} \oint_{\Gamma_S} \frac{dz}{2\pi i} [(zI - A_{\text{sym}})^{-1} E_{\text{sym}}]^{\gamma} (zI - A_{\text{sym}})^{-1}. \end{aligned} \quad (55)$$

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

$$(\tilde{V}_S \tilde{V}_S^T - V_S V_S^T)_{jk} = \sum_{\gamma=1}^{\infty} \oint_{\Gamma_S} \frac{dz}{2\pi i} e_{m+n,m+j}^T [(zI - A_{\text{sym}})^{-1} E_{\text{sym}}]^\gamma (zI - A_{\text{sym}})^{-1} e_{m+n,m+k}, \quad (56)$$

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 [32]. The goal is still the same, but our expansion will be different from [32], with the goal of creating powers of E_{sym} , rather than alternating products like $E_{\text{sym}} Q E_{\text{sym}} Q \dots E_{\text{sym}}$. To ease the notation, we denote

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

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. (56), the term with power γ becomes

$$\oint_{\Gamma_S} \frac{dz}{2\pi i} e_{m+n,m+j}^T \left[\left(\sum_{|i| \in [r]} \frac{\sigma_i P_i}{z(z - \sigma_i)} + \frac{I}{z} \right) E_{\text{sym}} \right]^\gamma \left(\sum_{|i| \in [r]} \frac{\sigma_i P_i}{z(z - \sigma_i)} + \frac{I}{z} \right) e_{m+n,m+k}. \quad (57)$$

When expanding the above, we get monomials of the form

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

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

$$\begin{aligned} & \left[\oint_{\Gamma_S} \frac{dz}{2\pi i} \frac{1}{z^{\alpha_0 + \beta_0 + \alpha_1 + \dots + \beta_{h-1} + \alpha_h}} \underbrace{\frac{\sigma_\gamma}{z - \sigma_\gamma} \frac{\sigma_\gamma}{z - \sigma_\gamma} \dots \frac{\sigma_\gamma}{z - \sigma_\gamma}}_{\beta_1 + \beta_2 + \dots + \beta_h \text{ factors}} \right] \\ & e_{m+n,m+j}^T E_{\text{sym}}^{\alpha_0} \underbrace{\left(P_\gamma E_{\text{sym}} P_\gamma E_{\text{sym}} \dots P_\gamma E_{\text{sym}} \right)}_{\beta_1 \text{ factors}} E_{\text{sym}}^{\alpha_1} \dots \underbrace{\left(P_\gamma E_{\text{sym}} P_\gamma E_{\text{sym}} \dots P_\gamma \right)}_{(\beta_h - 1) E_{\text{sym}} \text{ factors}} E_{\text{sym}}^{\alpha_h} e_{m+n,m+k}, \end{aligned} \quad (58)$$

At this point, one can see how several terms in Theorem B.2, especially the incoherence parameters τ and τ' , appear in the final bounds. The long matrix product can be rearranged as

$$\begin{aligned} & (e_{m+n,m+j}^T E_{\text{sym}}^{\alpha_0} w_\gamma) \underbrace{(w_\gamma^T E_{\text{sym}} w_\gamma \dots w_\gamma^T E_{\text{sym}} w_\gamma)}_{(\beta_1 - 1) E_{\text{sym}} \text{ factors}} (w_\gamma^T E_{\text{sym}}^{\alpha_1 + 1} w_\gamma) \\ & \dots (w_\gamma^T E_{\text{sym}}^{\alpha_{h-1} + 1} w_\gamma) \underbrace{(w_\gamma^T E_{\text{sym}} w_\gamma \dots w_\gamma^T E_{\text{sym}} w_\gamma)}_{(\beta_h - 1) E_{\text{sym}} \text{ factors}} (w_\gamma^T E_{\text{sym}}^{\alpha_h} e_{m+n,m+k}) \end{aligned} \quad (59)$$

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 $w_\gamma^T E_{\text{sym}} w_\gamma$ give rise to the term $\|U^T E V\|_\infty$ in Eq. (22).
- The terms $w_\gamma^T E_{\text{sym}}^{\alpha_i + 1} w_\gamma$ 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.

613 C.2.2 Notation and roadmap

614 **Setting C.2.** The following list also summarizes the notation used in the proof.

615 • For all matrices B , define $B_{\text{sym}} := \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$.

616 • Consider A . For each $i \in [r]$, let

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

617 Define $\Lambda := \{\sigma_i\}_{i \in [-r, r]}$ (which includes $\sigma_0 = 0$) and $W := [w_i]_{i \in [\pm r]}$, where $[\pm r] :=$
618 $\{i : |i| \in [r]\}$ (which does not include 0).

619 • Define \tilde{w}_i similarly, with rank \tilde{A} instead of r .

620 • Let $e_{N,k}$ be the k^{th} vector of the standard basis in \mathbb{R}^N .

621 • Γ_S is a contour encircling precisely the set $\{\sigma_i, \tilde{\sigma}_i : |i| \in S\}$ and no other eigenvalues, such
622 that the right-hand side of Eq. (54) converges absolutely for all z on it.

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

$$\begin{aligned} & \bullet \quad \alpha_0, \alpha_h \geq 0, \quad \text{and} \quad \alpha_k \geq 1 \text{ for } 1 \leq k \leq h-1, \\ & \bullet \quad \beta_k \geq 1 \text{ for } 1 \leq k \leq h, \\ & \bullet \quad \alpha + \beta = \gamma + 1, \quad \text{where } \alpha := \sum_{k=0}^h \alpha_k, \quad \text{and } \beta := \sum_{k=1}^h \beta_k. \end{aligned} \tag{60}$$

624 Note that the conditions above imply $2h - 1 \leq \gamma + 1$, so the maximum value for h is
625 $\lfloor \gamma/2 \rfloor + 1$.

626 • For each β above satisfying each $\beta_k \geq 1$, we use $\mathbf{I} = [i_1, i_2, \dots, i_\beta]$ for an element of
627 $[\pm r]^\beta$. Together, the triple $(\alpha, \beta, \mathbf{I})$ define uniquely a monomial of the form (58). Define
628 $\mathbf{I}_{a:b}$ as the subsequence $[i_a, i_{a+1}, \dots, i_b]$.

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

$$\begin{aligned} \mathcal{C}(\mathbf{I}) &:= \oint_{\Gamma_S} \frac{dz}{2\pi i} \frac{1}{z^{\gamma+1}} \prod_{j \in [\beta]} \frac{\sigma_{i_j}}{z - \sigma_{i_j}}, \\ \mathcal{M}(\alpha, \beta, \mathbf{I}) &:= E_{\text{sym}}^{\alpha_0} \left(\prod_{j=1}^{\beta_1} P_{i_j} E_{\text{sym}} \right) E_{\text{sym}}^{\alpha_1} \dots \left(\prod_{j=\beta_1+\dots+\beta_{h-1}}^{\beta-1} P_{i_j} E_{\text{sym}} \right) P_{i_{\beta-1}} E_{\text{sym}}^{\alpha_h}, \end{aligned}$$

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

632 • Define the following terms:

$$\begin{aligned} \mathcal{T}(\alpha, \beta) &= \sum_{\mathbf{I} \in [\pm r]^\beta} \mathcal{C}(\mathbf{I}) \mathcal{M}(\alpha, \beta, \mathbf{I}), & \mathcal{T}^{(\gamma, h)} &= \sum_{(\alpha, \beta) \in \Pi_h(\gamma)} \mathcal{T}(\alpha, \beta), \\ \mathcal{T}^{(\gamma)} &= \sum_{h=1}^{\lfloor \gamma/2 \rfloor + 1} \mathcal{T}^{(\gamma, h)}, & \mathcal{T} &= \sum_{\gamma \geq 1} \mathcal{T}^{(\gamma)}. \end{aligned}$$

633 From Eqs. (56), (57) and (58), we have

$$(\tilde{V}_S \tilde{V}_S^T - V_S V_S^T)_{jk} = e_{m+n, m+j}^T \mathcal{T} e_{m+n, m+k}. \tag{61}$$

634 At this point, we look at the larger context of Theorem B.2. Consider Eq. (27). To bound $\|\tilde{V}_S \tilde{V}_S^T -$
635 $V_S V_S^T\|_{2, \infty}$, we can fix one index j and find a bound for its j^{th} row that holds with probability close
636 enough to 1 to beat the n factor from the union bound. We have

$$(\tilde{V}_S \tilde{V}_S^T - V_S V_S^T)_{j, \cdot} = e_{m+n, m+j}^T \mathcal{T}. \tag{62}$$

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. (26) and (27).

Finally, consider Eq. (28). 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 C.2 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 (26) and (27). The details will be given later, when we prove (28).

C.2.3 Bounding the change in singular subspace expansions

Let us prove Eqs. (26) and (27) 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 (26) and (27).

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

Lemma C.3 (Bound on integral coefficients). *Consider the objects defined in Setting C.2. Let $\mathbf{I} = \{i_k : k \in \beta\} \in [\pm r]^\beta$ and denote the following:*

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

We have,

$$|\mathcal{C}(\mathbf{I})| \leq \frac{2^{\gamma+\beta-1}}{\sigma_S(\mathbf{I})^{\gamma+1-\beta} \Delta_S(\mathbf{I})^{\beta-1}} \leq \frac{2^{\gamma+\beta-1}}{\sigma_S^{\gamma+1-\beta} \Delta_S^{\beta-1}}. \quad (63)$$

In the steps that follow, we will mainly use the second bound of Eq. (63), with one exception where the first, more precise, bound is needed. It thus makes sense to keep both.

Lemma C.4 (Bound on monomial matrices). *Consider the objects defined in Setting C.2. Fix γ, h and $(\alpha, \beta) \in \Pi_h(\gamma)$ and $\mathbf{I} = \{i_k : k \in \beta\} \in [\pm r]^\beta$. Then*

$$\|M^T \mathcal{M}(\alpha, \beta, \mathbf{I}) M'\| \leq \|E\|^{\alpha-\alpha_0-\alpha_h+h-1} \cdot \|W^T E_{\text{sym}} W\|_\infty^{\beta-h} \cdot \|w_{i_1}^T E_{\text{sym}}^{\alpha_0} M\| \cdot \|w_{i_\beta}^T E_{\text{sym}}^{\alpha_h} M'\|. \quad (64)$$

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 $M^T \mathcal{T}_\nu(\alpha, \beta) M'$.

Lemma C.5 (Bound on $\mathcal{T}(\alpha, \beta)$). *Consider objects in Setting C.2. Fix γ and h such that $1 \leq h \leq \gamma/2 + 1$, and $\alpha, \beta \in \Pi_h(\gamma)$, and define the following terms*

$$\tau(M) = \max_{0 \leq \alpha \leq 10 \log(m+n)} \frac{1}{2r} \sum_{|i| \in [r]} \frac{\|w_i^T E_{\text{sym}}^\alpha M\|}{\|E_{\text{sym}}\|^\alpha}, \quad \text{and analogously for } \tau(M'). \quad (65)$$

$$R_1 := \frac{\|E\|}{\sigma_S} \vee \frac{2r \|W^T E_{\text{sym}} W\|_\infty}{\Delta_S}, \quad R_2 := \frac{\sqrt{2r} \|E\|}{\sqrt{\sigma_S} \Delta_S}, \quad R_3 := \frac{2r \max_{|i| \neq |j|} |w_i E_{\text{sym}}^2 w_j|}{\sigma_S \Delta_S} \quad (66)$$

and assume that

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

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

$$\|M^T \mathcal{T}(\alpha, \beta) M'\| \leq \begin{cases} r \tau(M) \tau(M') 2^{\gamma+\beta} R_1 R^{\gamma-1} & \text{if } 1 \leq h < \gamma/2 + 1, \\ 16r \tau(M) \tau(M') (4R)^{\gamma-2} (R_3 + R_1) & \text{if } h = \gamma/2 + 1. \end{cases} \quad (67)$$

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 $(\alpha, \beta) \in \Pi_h(\gamma)$ and all $1 \leq h \leq \gamma/2 + 1$, we get the following lemma.

668 **Lemma C.6** (Bound on each power term in \mathcal{T}). Consider the objects in Setting C.2 and R, R_1, R_2
 669 and R_3 from Lemma C.5. For each $1 \leq \gamma \leq 10 \log(m+n)$, we have

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

670 When $10 \log(m+n) < \gamma$, an analogous version of the above holds with $\|M\|$ and $\|M'\|$ replacing
 671 $\tau(M)$ and $\tau(M')$, respectively.

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

673 **Lemma C.7** (Bound on the whole \mathcal{T}). Consider the objects in Setting C.2 and R, R_1, R_2 and R_3
 674 from Lemma C.5. Suppose $R \leq 1/4$. Then the \mathcal{T} converges in the metric $\|\cdot\|$ and satisfies, for a
 675 universal constant C ,

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

676 Let us remark on the meanings of the new terms, which are simply translation of terms from Theorem
 677 B.2 into the language of Setting C.2.

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

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

682 and

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

- 683 • Similarly, recall the definitions of τ_1 and τ_2 in Eq. (25). As a function of M , τ satisfies

$$\tau(e_{m+n,k}) = \tau_1 \text{ for } k \leq m, \quad \tau(e_{m+n,k}) = \tau_2 \text{ for } m+1 \leq k \leq m+n, \quad \text{and } \tau(I) \leq 1, \quad (68)$$

684 To summarize, the logical structure is:

685 Lemma C.5 $\xrightarrow{\text{implies}}$ Lemma C.6 $\xrightarrow{\text{implies}}$ Lemma C.7 $\xrightarrow{\text{implies}}$ Eqs. (26), (27) in Theorem B.2

686 We will finish the last step, which is the proof of (26) and (27) here. The proofs of Lemmas C.5, C.6
 687 and C.7 will be postpone to Section C.3.

688 *Proof of Theorem B.2 part I.* Consider the objects defined in Theorem B.2 and the additional objects
 689 in Setting C.2. By the remark above, the condition (22) in Theorem B.2 is equivalent to $R_1 \vee R_2 \leq 1/4$,
 690 so we can apply the lemmas in this section.

691 Let us prove Eq. (26). Consider arbitrary $j, k \in [n]$. From Eq. (61), $(\tilde{V}_S \tilde{V}_S^T - V_S V_S^T)_{jk}$ is $M^T \mathcal{T} M'$
 692 for $M = e_{m+n,j+m}$ and $M' = e_{m+n,k+m}$. We apply the bound Lemma C.7, while replacing both
 693 $\tau(M)$ and $\tau(M')$ with τ_2 (permissible by Eq. (68)), to get

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

694 where the last inequality is due to the facts $\|M\| = \|M'\| = 1$ and $\tau_1, \tau_2 \geq (m+n)^{-1/2}$. This holds
 695 over all $j, k \in [n]$, so it extends to the infinity norm, proving Eq. (26).

696 Let us prove Eq. (27). Consider an arbitrary $j \in [n]$. By Eq. (62), $(\tilde{V}_S \tilde{V}_S^T - V_S V_S^T)_{j,\cdot} = M^T \mathcal{T} M'$
 697 for the choices $M = e_{m+n,j+m}$ and $M' = I_{m+n}$. We repeat the previous calculations, but this time
 698 Eq. (68) tells us to replace $\tau(M)$ with τ_2 and $\tau(M')$ with 1, to get

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

699 which holds uniformly over $j \in [n]$, proving Eq. (27). □

Next, we will finish proving Theorem B.2 by proving Eq. (28). 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.

C.2.4 Bounding the change in low rank approximations

Throughout this part, we assume $S = [s]$ for a fixed $s \in [r]$. Consider Eq. (54) again. We already know that integrating both sides gives $\tilde{W}_s \tilde{W}_s^T - W_s W_s^T$ on the left-hand side. Since we are aiming to bound $\tilde{A}_s - A_s$, we need $\tilde{W}_s \tilde{\Lambda}_s \tilde{W}_s^T - W_s \Lambda_s W_s^T$ on the left-hand side instead. This can be achieved by multiplying both sides with z before integrating, taking advantage of the fact

$$\oint_{\Gamma} \frac{z dz}{z - \sigma} = \sigma$$

for every contour Γ encircling σ . Therefore, the analogy of Eq. (55) is

$$\begin{aligned} (\tilde{A}_s - A_s)_{\text{sym}} &= (\tilde{A}_{\text{sym}})_s - (A_{\text{sym}})_s = \sum_{|i| \in [s]} \left(\frac{z \tilde{w}_i \tilde{w}_i^T}{z - \tilde{\sigma}_i} - \frac{z w_i w_i^T}{z - \sigma_i} \right) \\ &= \sum_{\gamma=1}^{\infty} \oint_{\Gamma_s} \frac{z dz}{2\pi i} [(zI - A_{\text{sym}})^{-1} E_{\text{sym}}]^{\gamma} (zI - A_{\text{sym}})^{-1}. \\ &= \sum_{\gamma=1}^{\infty} \oint_{\Gamma_s} \frac{z dz}{2\pi i} \left[\left(\sum_{|i| \in [r]} \frac{\sigma_i P_i}{z(z - \sigma_i)} + \frac{I}{z} \right) E_{\text{sym}} \right]^{\gamma} \left(\sum_{|i| \in [r]} \frac{\sigma_i P_i}{z(z - \sigma_i)} + \frac{I}{z} \right). \end{aligned} \quad (69)$$

Therefore, we can replace the integral coefficient $\mathcal{C}(\mathbf{I})$ from Setting C.2 with

$$\mathcal{C}_1(\mathbf{I}) := \oint_{\Gamma_s} \frac{z dz}{2\pi i} \frac{1}{z^{\gamma+1}} \prod_{j \in [\beta]} \frac{\sigma_{i_j}}{z - \sigma_{i_j}}. \quad (70)$$

Respectively define $\mathcal{M}_1(\alpha, \beta, \mathbf{I})$, $\mathcal{T}_1(\alpha, \beta)$, $\mathcal{T}_1^{(\gamma, h)}$, $\mathcal{T}_1^{(h)}$ and \mathcal{T}_1 analogously to $\mathcal{M}(\alpha, \beta, \mathbf{I})$, $\mathcal{T}(\alpha, \beta)$, $\mathcal{T}^{(\gamma, h)}$, $\mathcal{T}^{(h)}$ and \mathcal{T} from Setting C.2.

The only piece we need to modify in the proofs of Eqs. (26) and (27) is the integral coefficient bound, namely Lemma C.3. We have this bound for $\mathcal{C}_1(\mathbf{I})$:

Lemma C.8 (Bound on integral coefficients). *Consider the objects in Setting C.2 and Lemma C.3 and \mathcal{C}_1 defined in Eq. (70). We have,*

$$|\mathcal{C}_1(\mathbf{I})| \leq \frac{2^{\gamma+\beta-2}}{\sigma_s(\mathbf{I})^{\gamma-\beta} \Delta_s(\mathbf{I})^{\beta-1}} \leq \frac{2^{\gamma+\beta-2}}{\sigma_s^{\gamma-\beta} \Delta_s^{\beta-1}} = \frac{\sigma_s}{2} \cdot \frac{2^{\gamma+\beta-1}}{\sigma_s^{\gamma+1-\beta} \Delta_s^{\beta-1}}. \quad (71)$$

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. (70) is larger than 1, or when S does not contain exactly the first s singular values. Therefore, one can neither extend Eq. (28) to a general S nor to quantities like $\tilde{A}_s^2 - A_s^2$, at least not in a simple way.

Proof of Theorem B.2 part II. We prove Eq. (28). Fix $j \in [m]$ and $k \in [n]$. By Eq. (69), $(\tilde{A}_s - A_s)_{jk} = M^T \mathcal{T}_1 M'$ for $M = e_{m+n, j}$ and $M' = e_{m+n, m+k}$. The bound on $M^T \mathcal{T}_1 M'$ will simply be the same bound for $M^T \mathcal{T} M'$ scaled up by $\sigma_s/2$. By Eq. (68), we can also replace $\tau(M)$ with τ_1 and $\tau(M')$ with τ_2 . Therefore we obtain

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

where the last inequality holds due to $\tau_1, \tau_2 \geq (m+n)^{-1/2}$ and $\|M\| = \|M'\| = 1$. After passing to the infinity norm, the proofs of Eq. (28) and of Theorem B.2 are complete. \square

Now it remains to prove the lemmas in Sections C.2.3 and C.2.4. We will prove Lemmas C.4, C.5, C.6 and C.7. The proofs of the bounds on the integral coefficients (Lemmas C.3 and C.8) will be postponed to Section D due to their lengths.

730 C.3 Bounding the generic series

731 Let us prove Lemma C.4.

732 *Proof of Lemma C.4.* Consider a monomial matrix $\mathcal{M}(\alpha, \beta, \mathbf{I})$ has the form

$$\mathcal{M}(\alpha, \beta, \mathbf{I}) := E_{\text{sym}}^{\alpha_0} \left(\prod_{j=1}^{\beta_1} P_{i_j} E_{\text{sym}} \right) E_{\text{sym}}^{\alpha_1} \dots \left(\prod_{j=\beta-\beta_h}^{\beta-1} P_{i_j} E_{\text{sym}} \right) P_{i_{\beta-1}} E_{\text{sym}}^{\alpha_h}. \quad (72)$$

733 From Eq. (59), we can rearrange this to get

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

734 Let us break down this product into the following types:

- 735 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.
- 736 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
737 $(\beta_1-1) + (\beta_2-1) + \dots + (\beta_h-1) = \beta-h$.
- 738 3. $w_{i_j}^T E_{\text{sym}}^{\alpha_l+1} w_{i_{j+1}}$ for $j = \beta_1 + \dots + \beta_l$ and $\alpha = \alpha_l$ for some l : bounded by $\|E\|^{\alpha_l+1}$, and
739 their total power is $(\alpha_1+1) + (\alpha_2+1) + \dots + (\alpha_{h-1}+1) = \alpha - \alpha_0 - \alpha_h + h - 1$.

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

741 We continue with proving Lemma C.5.

742 *Proof of Lemma C.5.* For simplicity, let $X = W^T E_{\text{sym}} W$. Since

$$\mathcal{T}(\alpha, \beta) = \sum_{\mathbf{I} \in [2r]^\beta} \mathcal{C}(\mathbf{I}) \mathcal{M}(\alpha, \beta, \mathbf{I}),$$

743 we obtain

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

744 Applying the second part of the bound (63) on $\mathcal{C}(\mathbf{I})$ in Lemma C.3, we get

$$\begin{aligned} \|M^T \mathcal{T}(\alpha, \beta) M'\| &\leq \|X\|_\infty^{\beta-h} \|E\|^{\alpha-\alpha_0-\alpha_h+h-1} \frac{2^{\gamma+\beta-1}}{\sigma_S^\alpha \Delta_S^{\beta-1}} \sum_{\mathbf{I} \in [\pm r]^\beta} \|M^T E_{\text{sym}}^{\alpha_0} w_{i_1}\| \|w_{i_{\beta_h}}^T E_{\text{sym}}^{\alpha_h} M'\| \\ &= \|X\|_\infty^{\beta-h} \|E\|^{\alpha-\alpha_0-\alpha_h+h-1} \frac{2^{\gamma+\beta-1} (2r)^{\beta-2}}{\sigma_S^\alpha \Delta_S^{\beta-1}} \sum_{i \in [\pm r]} \|w_i^T E_{\text{sym}}^{\alpha_0} M\| \sum_{i \in [\pm r]} \|w_i^T E_{\text{sym}}^{\alpha_h} M'\| \\ &= \|X\|_\infty^{\beta-h} \|E\|^{\alpha+h-1} \frac{2^{\gamma+\beta-1} (2r)^\beta}{\sigma_S^\alpha \Delta_S^{\beta-1}} \sum_{i \in [\pm r]} \frac{\|w_i^T E_{\text{sym}}^{\alpha_0} M\|}{2r \|E\|^{\alpha_0}} \sum_{i \in [\pm r]} \frac{\|w_i^T E_{\text{sym}}^{\alpha_h} M'\|}{2r \|E\|^{\alpha_h}} \\ &\leq \tau(M) \tau(M') \|X\|_\infty^{\beta-h} \|E\|^{\alpha+h-1} \frac{2^{\gamma+\beta-1} (2r)^\beta}{\sigma_S^\alpha \Delta_S^{\beta-1}}. \end{aligned} \quad (73)$$

745 After rearrangements, we get

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

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

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

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

$$\|M^T \mathcal{T}(\alpha, \beta) M'\| \leq r\tau(M)\tau(M')2^{\gamma+\beta-1}R_1R^{\gamma-2h+1+2(h-1)} = r\tau(M)\tau(M')2^\beta R_1(2R)^{\gamma-1}.$$

We get the first case of Eq. (67). Now consider the case $h = \gamma/2 + 1$, which only happens when γ is even. The previous bound becomes

$$\|M^T \mathcal{T}(\alpha, \beta) M'\| \leq r\tau(M)\tau(M')2^{\gamma+\beta-1}R_2^{2(h-1)} = r\tau(M)\tau(M')2^{\beta-1}(2R_2)^\gamma. \quad (74)$$

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 B.3).

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 $\alpha_1 = \dots = \alpha_{h-1} = \beta_1 = \dots = \beta_h = 1$, thus $\beta = h$. Let (α^*, β^*) denote the corresponding tuple. Plugging into Eq. (72) and simplifying, we have

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

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 B.2. 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 $|w_{i_k}^T E_{\text{sym}}^2 w_{i_{k+1}}| \leq y$ and we are good. The rest of the product can be bounded by $\|E\|^2$. The total contribution of this subcase is at most

$$r\tau(M)\tau(M')2^{\gamma+\beta-1}R_3R_2^{\gamma-2} = r\tau(M)\tau(M')2^{3\gamma/2}R_3R_2^{\gamma-2},$$

since we can simply replace a factor of R_2^2 in Eq. (74) 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 (C.2) that $\mathcal{C}(\mathbf{I}) = 0$. Suppose $i \in S$, it is time for us to apply the first, stronger bound in Lemma C.3. 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

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

The proof is complete in the case $1 \leq \gamma \leq 10 \log(m+n)$. For the case $\gamma > 10 \log(m+n)$, consider Eq. (73) again. We cannot use $\tau(M)$ and $\tau(M')$ anymore, but we can use the trivial upper bounds

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

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

Let us proceed with the proof of Lemma C.6, which simply involve summing up the bounds in Lemma C.5 over all choices of (α, β) .

Proof of Lemma C.6. Let us consider the case $\gamma \leq 10 \log(m+n)$ first. Recall that

$$M^T \mathcal{T}^{(\gamma)} M' = \sum_{h=1}^{\lfloor \gamma/2 \rfloor + 1} \sum_{(\alpha, \beta) \in \Pi_h(\gamma)} M^T \mathcal{T}(\alpha, \beta) M'. \quad (75)$$

Consider the easy case where γ is odd. Then $h < \gamma/2 + 1$, and we have, by Lemma C.5,

$$\begin{aligned} \|M^T \mathcal{T}^{(\gamma)} M'\| &\leq \sum_{h=1}^{\lfloor \gamma/2 \rfloor + 1} \sum_{(\alpha, \beta) \in \Pi_h(\gamma)} r\tau(M)\tau(M') 2^\beta R_1(2R)^{\gamma-1} \\ &= r\tau(M)\tau(M') R_1(2R)^{\gamma-1} \sum_{h=1}^{\lfloor \gamma/2 \rfloor + 1} \sum_{\beta=h}^{\gamma+2-h} 2^\beta \left| \{(\alpha, \beta) \in \Pi_h(\gamma) : \sum_j \beta_j = \beta\} \right| \end{aligned} \quad (76)$$

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

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

The number of ways to choose such a tuple is $\binom{\beta-1}{h-1} \binom{\gamma+2-\beta}{h}$. Plugging into Eq. (76), we obtain

$$\begin{aligned} \|M^T \mathcal{T}_\nu^{(\gamma)} M'\| &\leq r\tau(M)\tau(M') R_1(2R)^{\gamma-1} \sum_{h=1}^{\lfloor \gamma/2 \rfloor + 1} \sum_{\beta=h}^{\gamma+2-h} \binom{\beta-1}{h-1} \binom{\gamma+2-\beta}{h} 2^\beta \\ &= r\tau(M)\tau(M') R_1(2R)^{\gamma-1} \sum_{\beta=1}^{\gamma+1} 2^\beta \sum_{h=1}^{\beta \wedge (\gamma+2-\beta)} \binom{\beta-1}{h-1} \binom{\gamma+2-\beta}{h} \\ &\leq r\tau(M)\tau(M') R_1(2R)^{\gamma-1} \sum_{\beta=1}^{\gamma+1} 2^\beta \binom{\gamma+1}{\beta} = 9r\tau(M)\tau(M') R_1(6R)^{\gamma-1}. \end{aligned} \quad (77)$$

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

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

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. \square

Now we finish the bound on the entire power series.

Proof of Lemma C.7. For convenience, let $k = \lfloor 10 \log(m+n) \rfloor$. Applying Lemma C.6, we have

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

and

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

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

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

789 The proof is complete. \square

790 D Proofs of technical lemmas

791 D.1 Proof of bound for contour integrals of polynomial reciprocals

792 In this section, we prove Lemmas C.3 and C.8, which provide the necessary bounds on the integral
 793 coefficients in the proof of Theorem B.2. Recall that the integrals we are interested in have the form

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

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

$$\mathcal{C}_\nu(\mathbf{I}) := \oint_{\Gamma_S} \frac{z^\nu dz}{2\pi i} \frac{1}{z^{\gamma+1}} \prod_{k=1}^{\beta} \frac{\sigma_{i_k}}{z - \sigma_{i_k}}, \quad \text{where } \nu \in \{0, 1\} \text{ and } \beta \leq \gamma + 1. \quad (78)$$

795 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 $a_i \in S$
 796 and each $b_j \notin S$, having multiplicities m_i and n_j respectively. We can rewrite the above into

$$\mathcal{C}_\nu(\mathbf{I}) = \prod_{i=1}^l a_i^{m_i} \prod_{j=1}^k b_j^{n_j} C(n_0; A, \mathbf{m}; B, \mathbf{n}), \quad (79)$$

797 where

$$C(n_0; A, \mathbf{m}; B, \mathbf{n}) := \oint_{\Gamma_A} \frac{dz}{2\pi i} \frac{1}{z^{n_0}} \prod_{j=1}^k \frac{1}{(z - b_j)^{n_j}} \prod_{i=1}^l \frac{1}{(z - a_i)^{m_i}}, \quad (80)$$

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

801 **Lemma D.1.** *Let $A = \{a_i\}_{i \in [l]}$ and $B = \{b_j\}_{j \in [k]}$ be disjoint set of complex non-zero numbers
 802 and $\mathbf{m} = \{m_i\}_{i \in [l]}$ and n_0 and $\mathbf{n} = \{n_j\}_{j \in [k]}$ be nonnegative integers such that $m + n + n_0 \geq 2$,
 803 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
 804 $B \cup \{0\}$. Let $a, d > 0$ be arbitrary such that:*

$$d \leq a, \quad a \leq \min_i |a_i|, \quad d \leq \min_{i,j} |a_i - b_j|. \quad (81)$$

805 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
 806 $C(n_0; A, \mathbf{m}; B, \mathbf{n})$ defined Eq. (80), we have

$$|C(n_0; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m+n+n_0-2}{m-1} \frac{1}{a^{n_0-m'} d^{m+n-1}} \prod_{i=1}^l \frac{1}{|a_i|^{m'_i}} \quad (82)$$

807 *Proof.* Firstly, given the sets A and B and the notations and conditions in Lemma D.1, the weak
 808 bound below holds

$$|C(n_0; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m+n+n_0-2}{m-1} \frac{1}{d^{m+n+n_0-1}}. \quad (83)$$

809 We omit the details of the proof, which is a simple induction argument. We now use Eq. (83) to prove
 810 the following:

$$|C(n_0; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m+n+n_0-2}{m-1} \frac{1}{a^{n_0} d^{m+n-1}}. \quad (84)$$

811 We proceed with induction. Let $P_1(N)$ be the following statement: “For any sets A and B , and the
812 notations and conditions described in Lemma D.1, such that $m + n + n_0 = N$, Eq. (84) holds.”

813 Since $m + n + n_0 \geq 2$, consider $N = 2$ for the base case. The only case where the integral is non-zero
814 is when $m = 1$ and $n + n_0 = 1$, meaning $A = \{a_1\}$, $m_1 = 1$ and either $B = \emptyset$ and $n_0 = 1$, or
815 $B = \{b_1\}$ 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
816 latter, confirming the inequality in both.

817 Consider $n \geq 3$ and assume $P_1(n - 1)$. If $m = 0$, the integral is again 0. If $n_0 = 0$, Eq. (84)
818 automatically holds by being the same as Eq. (83). Assume $m, n_0 \geq 1$. There must then be some
819 $i \in [l]$ such that $m_i \geq 1$, without loss of generality let 1 be that i . We have

$$C(n_0; A, \mathbf{m}; B, \mathbf{n}) = \frac{1}{a_1} \left[C(n_0 - 1; A, \mathbf{m}; B, \mathbf{n}) - C(n_0; A, \mathbf{m}^{(1)}; B, \mathbf{n}) \right] \quad (85)$$

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

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

$$|C(n_0 - 1; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m + n + n_0 - 3}{m - 1} \frac{1}{a^{n_0 - 1} d^{m + n - 1}}. \quad (86)$$

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

$$|C(n_0; A, \mathbf{m}^{(1)}; B, \mathbf{n})| \leq \binom{m + n + n_0 - 3}{m - 2} \frac{1}{a^{n_0} d^{m + n - 2}} \leq \binom{m + n + n_0 - 3}{m - 2} \frac{1}{a^{n_0 - 1} d^{m + n - 1}}. \quad (87)$$

823 Notice that the binomial coefficients in Eqs. (86) and (87) sum to the binomial coefficient in Eq. (84),
824 we get $P_1(N)$, which proves Eq. (84) by induction.

825 Now we can prove Eq. (82). The logic is almost identical, with Eq. (84) playing the role of Eq. (83)
826 in its own proof, handling an edge case in the inductive step. Let $P_2(n)$ be the statement: “For any
827 sets A and B , and the notations and conditions described in Lemma D.1, such that $m + n + n_0 = N$,
828 Eq. (82) holds.”

829 The cases $N = 1$ and $N = 2$ are again trivially true. Consider $N \geq 3$ and assume $P_2(N - 1)$. Fix
830 any sequence m'_1, m'_2, \dots, m'_l satisfying $0 \leq m'_i \leq m_i$ for each $i \in [l]$ and $n_0 \geq m'_1 + \dots + m'_l$.
831 If $m'_1 = m'_2 = \dots = m'_l = 0$, we are done by Eq. (84). By symmetry among the indices, assume
832 $m'_1 \geq 1$. This also means $n_0 \geq 1$. Consider Eq. (85) again. For the first integral on the right-hand
833 side, applying $P_2(N - 1)$ for the parameters $n_0 - 1, n_1, \dots, n_k, m_1, \dots, m_l$ and $m'_1 - 1, m'_2, \dots, m'_l$
834 yields the bound

$$|C(n_0 - 1; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m + n + n_0 - 3}{m - 1} \frac{1}{a^{n_0 - m'} d^{m + n - 1}} \frac{1}{|a_1|^{m'_1 - 1}} \prod_{i=2}^l \frac{1}{|a_i|^{m'_i}}. \quad (88)$$

835 Applying $P_2(N - 1)$ for the parameters $n_0, n_1, \dots, n_k, m_1 - 1, \dots, m_l$ and $m'_1 - 1, m'_2, \dots, m'_l$,
836 we get the following bound for the second integral on the right-hand side of Eq. (85):

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

837 Summing up the bounds by summing the binomial coefficients, we get exactly $P_2(N)$, so Eq. (82) is
838 proven by induction. \square

839 **Lemma D.2.** Let $A, B, \mathbf{m}, \mathbf{n}, n_0, \Gamma_A$ and a, d be the same, with the same conditions as in Lemma
840 D.1. Suppose that $0 \leq m'_i \leq m_i$ and $0 \leq n'_j \leq n_j$ for each $i, j \geq 1$ and

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

841 Then for $C(n_0; A, \mathbf{m}; B, \mathbf{n})$ defined in Eq. (80), we have

$$|C(n_0; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{n + n_0 - n' + m - 2}{m - 1} \frac{(1 + d/a)^{n'}}{a^{n_0 - m' - n'} d^{m + n - 1}} \prod_{i=1}^l \frac{1}{|a_i|^{m'_i}} \prod_{j=1}^k \frac{1}{|b_j|^{n'_j}}. \quad (89)$$

842 *Proof.* We have the expansion

$$\begin{aligned}
& \frac{1}{z^{n_0}} \prod_{j=1}^k \frac{b_j^{n'_j}}{(z-b_j)^{n_j}} \prod_{i=1}^l \frac{1}{(z-a_i)^{m_i}} = \frac{1}{z^{n_0-n'}} \prod_{j=1}^k \frac{1}{(z-b_j)^{n_j-n'_j}} \prod_{j=1}^k \left(\frac{1}{z} - \frac{1}{z-b_j} \right)^{n'_j} \prod_{i=1}^l \frac{1}{(z-a_i)^{m_i}} \\
& = \frac{1}{z^{n_0-n'}} \prod_{j=1}^k \frac{1}{(z-b_j)^{n_j-n'_j}} \sum_{0 \leq r_j \leq n'_j \forall j} \frac{(-1)^{r_1+\dots+r_k}}{z^{n'-r_1-\dots-r_k}} \prod_{j=1}^k \binom{n'_j}{r_j} \frac{1}{(z-b_j)^{r_j}} \prod_{i=1}^l \frac{1}{(z-a_i)^{m_i}} \\
& = \sum_{0 \leq r_j \leq n'_j \forall j} \frac{(-1)^{r_1+\dots+r_k}}{z^{n_0-r_1-\dots-r_k}} \prod_{j=1}^k \binom{n'_j}{r_j} \frac{1}{(z-b_j)^{r_j+n_j-n'_j}} \prod_{i=1}^l \frac{1}{(z-a_i)^{m_i}}.
\end{aligned}$$

843 Integrating both sides over Γ_A , we have

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

844 where the j -entry of $\mathbf{r} + \mathbf{n} - \mathbf{n}'$ is simply $r_j + n_j - n'_j$. Applying Lemma D.1 for each summand on
845 the right-hand side and rearranging the powers, we get

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

846 Summing up the bounds, we get

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

847 Rearranging the term, we get precisely the desired inequality. \square

848 With the lemma above, we are ready to prove both Lemmas C.3 and C.8.

849 *Proof of Lemmas C.3 and C.8.* First rewrite the integral into the forms of (78), then (79) and (80).
850 Let us consider two cases for \mathcal{C} :

851 1. $\nu = 0$, so $n_0 = \gamma + 1$. Let $a = \sigma_S(\mathbf{I})$, $d = \Delta_S(\mathbf{I})$, $m = \beta_S(\mathbf{I})$, $n = n' = \beta_{S^c}(\mathbf{I})$,
852 $m'_i = m_i$ and $n'_j = n_j$ for all i, j , then $m' + n' = \beta \leq \gamma + 1 = n_0$, so we can apply
853 Lemma D.2 to get

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

854 or equivalently,

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

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

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

857 where the last inequality holds due to $\sigma_S(\mathbf{I}) \geq \sigma_S$ and $\Delta_S(\mathbf{I}) \geq \Delta_S$. The proof of Lemma
858 C.3 is complete.

859 2. $\nu = 1$ and $S = [s]$ for some $s \in [r]$. This is the special case for Lemma C.8. Note that
 860 $n_0 = \gamma$ in this case. Without loss of generality, assume $|a_1| = \sigma_s(\mathbf{I})$, then we are guaranteed
 861 $m_1 \geq 1$. Applying Lemma D.2 for the same parameters as in the previous case, except that
 862 $m'_1 = m_1 - 1$, we get

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

863 which translates to

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

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

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

868 Since $\gamma \geq 1$, we have

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

869 It suffices to show $\sigma_s(\mathbf{I})/\Delta_s(\mathbf{I}) \leq \sigma_s/\delta_s$ to complete the last step. Choose $t \in [s]$ where
 870 $\sigma_t = \sigma_S(\mathbf{I})$, then $\Delta_S(\mathbf{I}) \geq \sigma_t - \sigma_{s+1}$, thus

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

871 This completes the final step, proving Lemma C.8. Note that the inequality above does not
 872 hold if S does not contain a contiguous chunk of the largest singular values.

873 □

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

875 In this section, we prove Theorem B.4, which gives semi-isotropic bounds for powers of E_{sym} in the
 876 second step of the main proof strategy.

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

879 **Lemma D.3.** *Let $m, r \in \mathbb{N}$ and $U \in \mathbb{R}^{m \times r}$ be a matrix whose columns u_1, u_2, \dots, u_r are unit
 880 vectors. Let E be a $m \times n$ random matrix following Model (29) with parameters M and $\varsigma = 1$,
 881 meaning E has independent entries and*

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

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

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

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

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

884 **Lemma D.4.** *Let E be a $m \times n$ random matrix following the model in Lemma D.3. For any matrix
 885 $V \in \mathbb{R}^{m \times l}$ with unit columns v_1, v_2, \dots, v_l , any $a \in \mathbb{N}$, $k \in [n]$, any $D > 0$, and any $p \in \mathbb{N}$ such
 886 that*

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

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

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

Let us prove the main objective of this section, Theorem B.4, before delving into the proof of the technical lemmas.

Proof of Theorem B.4. Consider the analogue of Eq. (31) for V (we wrote the proof for V before the final edit, and wanted to save the energy of changing to U) and Eq. (32), 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 $a \in [\lfloor t \log(m+n) \rfloor]$:

$$\|e_{n,k}^T (E^T E)^a E^T U\| \leq C\tau_1(U, \log \log(m+n))(1.9\zeta\sqrt{m+n})^{2a+1}\sqrt{r} \quad (90)$$

$$\|e_{n,k}^T (E^T E)^a V\| \leq C\tau_0(V, \log \log(m+n))(1.9\zeta\sqrt{m+n})^{2a}\sqrt{r}. \quad (91)$$

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 D.3 and D.4.

Let us prove Eq. (90) for a . Applying Lemma D.3 for the random matrix E/ζ and $D = 2^{13}$ gives, with probability $1 - \log^{-4.04}(m+n)$,

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

where the second inequality is due to $\alpha \leq (\sqrt{2}/1.9)^\alpha$. 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. (90), and thus Eq. (32) follows.

Let us prove Eq. (91). Applying Lemma D.3 for the random matrix E/ζ and $D = 2^{10}$ gives, with probability $1 - \log^{-8}(m+n)$,

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

proving Eq. (91) and thus Eq. (31) after a union bound, similar to the previous case.

Let us now prove Eqs (33) and (34), focusing on the former first. Since the 2-to- ∞ norm is the the largest norm among the rows, it suffices to prove Eq. (31) 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. (31), but with probability at least $1 - (m+n)^{-4.04}$. Applying another union bound over $k \leq [n]$ gives Eq. (33) with probability at least $1 - (m+n)^{-3}$. The proof of (34) is analogous. The proof of Theorem B.4 is complete. \square

Now let us handle the technical lemmas D.3 and D.4. The odd case (Lemma D.3) is more difficult, so we will handle it first to demonstrate our technique. The argument for the even case (Lemma D.4) is just a simpler version of the same technique.

D.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

$$\begin{aligned} \mathbf{E} \left[\|e_{n,1}^T (E^T E)^a E^T U\|^{2p} \right] &= \mathbf{E} \left[\left(\sum_{l=1}^r (e_{n,1}^T (E^T E)^a E^T u_l)^2 \right)^p \right] \\ &= \sum_{l_1, \dots, l_p \in [r]} \mathbf{E} \left[\prod_{h=1}^p (e_{n,1}^T (E^T E)^a E^T u_{l_h})^2 \right]. \end{aligned} \quad (92)$$

917 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
 918 bipartite graph $M_{m,n}$ such that $j_0 = 1$. Here the two parts of M are $I = \{1', 2', \dots, m'\}$ and
 919 $J = \{1, 2, \dots, n\}$, where the prime symbol serves to distinguish two vertices on different parts with
 920 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
 921 above as

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

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

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

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

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

925 Define the following notation:

- 926 1. \mathcal{P} is the set of all *star*, i.e. tuples of walks $P = (P_1, \dots, P_{2p})$ on the complete bipartite
 927 graph $M_{m,n}$, such that each walk $P_r \in \mathcal{W}$ and each edge appears at least twice.
 928 Rename each tuple $(W_{h1}, W_{h2})_{h=1}^p$ as a star P with $W_{hd} = P_{2h-2+d}$.
 929 For each P , let $V(P)$ and $E(P)$ respectively be the set of vertices and edges involved in P .
 930 Define the partition $V(P) = V_I(P) \cup V_J(P)$, where $V_I(P) := V(P) \cap I$ and $V_J(P) :=$
 931 $V(P) \cap J$.
 932 2. $E_P := E_{P_1} E_{P_2} \dots E_{P_{2p}}$.
 933 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
 934 any tuple $Q = (q_1, \dots, q_r)$.
 935 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
 936 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
 937 $\{i_{r's'} : r' \leq r, s' \leq s\}$, then it is strictly larger than all indices before it, and similarly for
 938 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
 939 transforms P into S . The notations $V(S)$, $V_I(S)$, $V_J(S)$, $E(S)$ are defined analogously.
 940 Observe that the shape of P is unique, and \mathcal{S} forms a set of equivalent classes on \mathcal{P} .
 941 5. Denote by $\mathcal{P}(S)$ the class associated with the shape S , namely the set of all stars P having
 942 shape S .

943 We can rewrite the previous sum as:

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

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

$$\sum_{P \in \mathcal{P}} |\mathbf{E}[E_P]| \prod_{h=1}^p \|U_{\cdot, i_{(2h-1)a}}\| \|U_{\cdot, i_{(2h)a}}\| = r^p \sum_{P \in \mathcal{P}} u_{P^{\text{end}}} |\mathbf{E}[E_P]| = r^p \sum_{S \in \mathcal{S}} \sum_{P \in \mathcal{P}(S)} u_{P^{\text{end}}} |\mathbf{E}[E_P]|, \quad (93)$$

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

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

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

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

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

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

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

952 Consider $\|u\|_l^l$ for an arbitrary $l \in \mathbb{N}$. When $l = 1$, $\|u\|_1^1 \leq \sqrt{m}$ by Cauchy-Schwarz. When $l \geq 2$,
 953 we have $\|u\|_l^l \leq \|u\|_\infty^{l-2} \|u\|_2^2 = \|u\|_\infty^{l-2}$. Thus

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

954 where, we define $V_1(Q)$ as the set of vertices appearing in Q exactly once and $V_2(Q)$ as the set of
 955 vertices appearing at least twice, and to shorten the notation, we let $\nu(S) := |V_1(S^{\text{end}})| + 2|V_2(S^{\text{end}})|$.
 956 Combining the bounds, we get the upper bound below for (93):

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

957 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
 958 subset of shapes having these quantities. To further shorten the notation, let $M_1 := M^{2p(2a+1)} \|u\|_\infty^{2p}$.
 959 Then we can rewrite the above as:

$$M_1 \sum_{x,y,z,t \in \mathcal{A}} M^{-2(z+t)} m^{z-x/2-y} n^{t-1} \|u\|_\infty^{-x-2y} |\mathcal{S}(x, y, z, t)|, \quad (94)$$

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

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

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

965 **Claim D.5.** *When each edge is repeated at least twice, we have $z - x/2 - y + t - 1 \leq 2ap$.*

966 *Proof of Claim D.5.* Let $S = (S_1, \dots, S_{2p})$, where $S_r = j_{r0} i_{r0} j_{r1} i_{r1} \dots j_{ra} i_{ra}$. We have $j_{r0} = 1$
 967 for all r . It is tempting to think (falsely) that when each edge is repeated at least twice, each vertex
 968 appears at least twice too. If this were to be the case, then each vertex in the set

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

969 appears at least twice. The sum of their repetitions is $4ap + x$, so the size of this set is at most
 970 $2ap + x/2$. Since this set covers every vertex, with the possible exceptions of $1 \in I$ and $V_2(S^{\text{end}})$, its
 971 size is at least $z - y + t - 1$, proving the claim. In general, there will be vertices appearing only once
 972 in S . However, we can still use the simple idea above. Temporarily let $A_1(S)$ be the set of vertices
 973 appearing once in S and $f(S)$ be the sum of all edges' repetitions in S . Let $S^{(0)} := S$. Suppose for

974 $k \geq 0$, $S^{(k)}$ is known and satisfies $|A(S^{(k)})| = |A(S)| - k$, $f(S^{(k)}) = 4pa + x - 2k$ and each edge
 975 appears at least twice in $S^{(k)}$. If $A_1(S^{(k)}) = \emptyset$, then by the previous argument, we have

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

976 proving the claim. If there is some vertex in $A_1(S^{(k)})$, assume it is some i_{rs} , then we must have
 977 $s \leq a - 1$ and $j_{rs} = j_{r(s+1)}$, otherwise the edge $j_{rs}i_{rs}$ appears only once. Create $S^{(k+1)}$ from
 978 $S^{(k)}$ by removing i_{rs} and identifying j_{rs} and $j_{r(s+1)}$, we have $|A(S^{(k+1)})| = |A(S)| - (k + 1)$
 979 and $|f(S^{(k+1)})| = 4pa + x - 2(k + 1)$. Further, since i_{rs} is unique, $j_{rs}i_{rs} \equiv i_{rs}j_{r(s+1)}$ are the only
 980 2 occurrences of this edge in $S^{(k)}$, thus the edges remaining in $S^{(k+1)}$ also appears at least twice.
 981 Now we only have $|A_1(S^{(k+1)})| \leq |A_1(S^{(k)})|$, with possible equality, since j_{rs} can be come unique
 982 after the removal, but since there is only a finite number of edges to remove, eventually we have
 983 $A_1(S^{(k)}) = \emptyset$, completing the proof of the claim. \square

984 Claim D.5 shows that we can define the set \mathcal{A} of *eligible sizes* as follows:

$$\mathcal{A} = \{(x, y, z, t) \in \mathbb{N}_{\geq 0}^4 : 1 \leq t; 1 \leq x + y \leq z; x + 2y \leq 2p; z - x/2 - y + t - 1 \leq 2ap\}. \quad (95)$$

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

986 **Claim D.6.** *Given a tuple $(x, y, z, t) \in \mathcal{A}$, where \mathcal{A} is defined in Eq. (95), we have*

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

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

- 990 1. Start with $V_J = \{1\}$ and $V_I = \emptyset$, $W = []$ and T being the tree with one vertex, 1.
- 991 2. For $r = 1, 2, \dots, 2p$:
 - 992 (a) Relabel S_r as $1k_1k_2 \dots k_{2a}$.
 - 993 (b) For $s = 1, 2, \dots, 2a$:
 - 994 • If $k_s \notin V(T)$ then add k_s to T and draw an edge connecting k_{s-1} and k_s , then
 995 mark that edge with a (+) in T , and append (+) to W . We call its instance in S_r a
 996 *plus edge*.
 - 997 • If $k_s \in V(T)$ and the edge $k_{s-1}k_s \in E(T)$ and is marked with (+): unmark it in
 998 T , and append (-) to W . We call its instance in S_r a *minus edge*.
 - 999 • 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
 1000 a *neutral edge*, and append the symbol k_s to W .

1001 This scheme only creates a *preliminary codeword* W , which does not yet uniquely determine the
 1002 original S . To be able to trace back S , we need the scheme in [7] to add more details to the preliminary
 1003 codewords. For completeness, we will describe this scheme later, but let us first bound the number of
 1004 preliminary codewords.

1005 **Claim D.7.** *Let $\mathcal{PC}(x, y, z, t)$ denote the set of preliminary codewords generable from shapes in*
 1006 *$\mathcal{S}(x, y, z, t)$. Then for $l := z + t - 1$ we have*

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

1007 Note that the bound above does not depend on x and y . In fact, for fixed z and t , the right-hand side
 1008 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
 1009 eligible. We believe there is plenty of room to improve this bound in the future.

1010 *Proof.* To begin, note that there are precisely z and $t - 1$ plus edges whose right endpoint is
 1011 respectively in I and J . Suppose we know u and v , the number of minus edges whose right endpoint
 1012 is in I and J , respectively. Then

- 1013 • The number of ways to place plus edges is at most $\binom{2p(a+1)}{z} \binom{2pa}{t-1}$.
- 1014 • The number of ways to place minus edges, given the position of plus edges, is at most
- 1015 $\binom{2p(a+1)-z}{u} \binom{2pa-t+1}{v}$.
- 1016 • The number of ways to choose the endpoint for each neutral edge is at most
- 1017 $z^{2p(a+1)-z-u} t^{2pa-t+1-v}$.

1018 Combining the bounds above, we have

$$|\mathcal{S}(x, y, z, t)| \leq \binom{2p(a+1)}{z} \binom{2pa}{t-1} \sum_{u+v=z+t-1} \binom{2p(a+1)-z}{u} \binom{2pa-t+1}{v} z^{f(z,u)} t^{g(t,v)}, \quad (96)$$

1019 where $f(z, u) = 2p(a+1) - z - u$ and $g(u, v) = 2pa - t + 1 - v$. Let us simplify this bound. The

1020 sum on the right-hand side has the form

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

1021 where $k = 2(p(2a+1) - (z+t-1))$, $N = 2p(a+1) - z$, $M = 2pa - t + 1$. We have

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

1022 Replacing M , N and k with their definitions, we get

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

1023 replacing $z+t-1$ with l , we prove the claim. □

1024 Back to the proof of Claim D.6, to uniquely determine the shape S , the general idea is the following.

1025 We first generated the preliminary codeword W from S , then attempt to decode it. If we encounter

1026 a plus or neutral edge, we immediately know the next vertex. If we see a minus edge that follows

1027 from a plus edge (u, v) , we know that the next vertex is again u . Similarly, if there are chunks

1028 of the form $(++ \dots + - - \dots -)$ with the same number of each sign, the vertices are uniquely

1029 determined from the first vertex. Therefore, we can create a condensed codeword W^* repeatedly

1030 removing consecutive pairs of $(+-)$ until none remains. For example, the sections $(-+-+)$ and

1031 $(-++-)$ both become $(-)$. Observe that the condensed codeword is always unique regardless of

1032 the order of removal, and has the form

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

1033 where we allow blocks of pure pluses and minuses to be empty. The minus blocks that remain in W^*

1034 are the only ones where we cannot decipher.

1035 Recall that during decoding, we also reconstruct the tree $T(S)$, and the partial result remains a tree

1036 at any step. If we encounter a block of minuses in W^* beginning with the vertex i , knowing the

1037 right endpoint j of the last minus edge is enough to determine the rest of the vertices, which is just

1038 the unique path between i and j in the current tree. We call the last minus edge of such a block an

1039 *important edge*. There are two cases for an important edge.

- 1040 1. If i and all vertices between i and j (excluding j) are only adjacent to at most two plus
- 1041 edges in the current tree (exactly for the interior vertices), we call this important edge *simple*
- 1042 and just mark the it with a direction (left or right, in addition to the existing minus). For
- 1043 example, $(-- \dots -)$ becomes $(-- \dots (-dir))$ where dir is the direction.

1044 2. If the edge is non-simple, we just mark it with the vertex j , so $(- \dots -)$ becomes
 1045 $(- \dots (-j))$.

1046 It has been shown in [7] that the fully codeword \overline{W} resulting from W by marking important edges
 1047 uniquely determines S , and that when the shape of S is *that of a single walk*, the cost of these
 1048 markings is at most a multiplicative factor of $2(4N + 8)^N$, where N is the number of neutral edges
 1049 in the preliminary W . To adapt this bound to our case, we treat the star shape S as a single walk,
 1050 with a neutral edge marked by 1 after every $2a + 1$ edges. There are $2p - 1$ additional neutral edges
 1051 from this perspective, making $N = 4p(a + 1) - 2l - 1$ in total. Combining this with the bound on
 1052 the number of preliminary codewords (Claim D.7) yields

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

1053 where $l = z + t - 1$. Claim D.6 is proven. \square

1054 Back to the proof of Lemma D.3. Temporarily let

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

1055 Note that $(2p(a+1))!(2pa)!F_l$ is precisely the upper bound on $|\mathcal{S}(x, y, z, t)|$ in Claim D.6. Also let

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

1056 Replacing the appropriate terms in the bound in Claim D.6 with these short forms, we get another
 1057 series of upper bounds for the last double sum in Eq. (93):

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

1058 Temporarily let C_l be the term corresponding to l in the sum above. For $l \geq x + y + 1$, we have

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

1059 The last power is approximately $e^{-2} \approx 0.135$, and for $p \geq 7$ a routine numerical check shows that it
 1060 is at least $1/8$. The second to last power is at least 1. The fraction be bounded as below.

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

1061 Therefore, under the assumption that $m+n \geq 256M^2 p^6 (2a+1)^4$, we have $C_l \geq 2C_{l-1}$ for all
 1062 $l \geq 1$, so $\sum_l C_l \leq 2C_{l^*}$, where $l^* = \lfloor 2pa + x/2 + y \rfloor$, the maximum in the range. We have

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

1063 Temporarily let $d = p - (\lfloor \frac{x}{2} \rfloor + y)$ and $N = p(2a+1)$, we have

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

1064 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
 1065 the following series of upper bounds for the last double sum in Eq. (93):

$$\begin{aligned} & M_2(m+n)^{2pa} \sum_{d=0}^{p-1} 4(p-d) \|u\|_{\infty}^{-2(p-d)} \cdot \frac{(2M^{-2})^{N-d+1} (N-d+1)^{2d} (8p+8d-2)^{2p+2d-1}}{(2pa)! \cdot (2d)! \cdot (N-d)!} \\ & \leq M_3(m+n)^{2pa} \sum_{d=0}^{p-1} \|u\|_{\infty}^{2d} \cdot \frac{2^{-d} M^{2d} (N-d+1)^{2d} (8p+8d-2)^{2p+2d-1}}{(2d)! \cdot (N-d)!}, \end{aligned} \quad (97)$$

1066 where

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

1067 Let us bound the sum at the end of Eq. (97). Temporarily let A_d be the term corresponding to d and
 1068 $x := 2^{-1/2} M \|u\|_{\infty}$. We have

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

1069 Therefore

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

1070 Plugging this into Eq. (97), we get another upper bound for (93):

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

1071 where

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

1072 To sum up, we have

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

1073 Let $D > 0$ be arbitrary. By Markov's inequality, for any p such that $m+n \geq 2^8 M^2 p^6 (2a+1)^4$, the
 1074 moment bound above applies, so we have

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

1075 with probability at least $1 - (2^5/D)^{2p}$. Replacing $\|u\|_{\infty}$ with $\frac{1}{\sqrt{r}} \|U\|_{2,\infty}$, we complete the proof. \square

1076 D.2.2 Case 2: even powers

1077 *Proof.* Without loss of generality, assume $k = 1$. We can reuse the first part and the notations from
 1078 the proof of Lemma D.3 to get the bound

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

1079 where $v_i = r^{-1/2} \|V_{\cdot,i}\|$. Again,

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

1080 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
 1081 shape S , and

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

1082 Note that a shape for a star now consists of walks of length $2a$:

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

1083 We have, for any shape S and $P \in \mathcal{P}(S)$,

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

1084 where the power of n in the last inequality is due to 1 having been fixed in $V_J(S)$. Therefore

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

1085 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
 1086 indices:

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

1087 Using the previous argument in the proof of Lemma D.3 for counting shapes, we have for $(z, t) \in \mathcal{A}$:

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

1088 where

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

1089 We have

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

1090 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
 1091 the last sum above. An analogous calculation from the proof of Lemma D.3 shows that under the
 1092 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

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

1093 Therefore

$$\begin{aligned} \mathbf{E} \left[\|e_{n,1}^T (E^T E)^a V\|^{2p} \right] &\leq r^p \sum_{S \in \mathcal{S}} \sum_{P \in \mathcal{P}(S)} v_{\text{pend}} |\mathbf{E}[E_P]| \\ &\leq 2r^p M_2 \frac{M^{-4ap} 2^{2ap+1} (8p-2)^{2p-1}}{[(2ap)!]^2} (m+n)^{2ap} = 4 \left(2^3 p r^{1/2} \|v\|_\infty [2(m+n)]^a \right)^{2p}. \end{aligned}$$

1094 Pick $D > 0$, by Markov's inequality, we have

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

1095 Replacing $\|v\|_\infty$ with $r^{-1/2} \|V\|_{2,\infty}$, we complete the proof. \square

References

- [1] 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>.
- [2] 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>.
- [3] 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>.
- [4] 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>.
- [5] 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>.
- [6] 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>.
- [7] 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>.
- [8] 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>.
- [9] Anonymous authors. Supplementary appendix to the paper “fast exact recovery of noisy matrix from few entries: the infinity norm approach”, 2025. URL https://github.com/thbl2016/Matrix-Completion-NeuRIPS2025/blob/main/Matrix_Completion_NeuRIPS2025_supplementary.pdf.
- [10] 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>.
- [11] 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>.
- [12] A. L. Chistov and D. Yu. Grigor’ev. Complexity of quantifier elimination in the theory of algebraically closed fields. In *Mathematical foundations of computer science, 1984 (Prague, 1984)*, volume 176 of *Lecture Notes in Comput. Sci.*, pages 17–31. Springer, Berlin, 1984. ISBN 3-540-13372-0. doi: 10.1007/BFb0030287. URL <https://doi.org/10.1007/BFb0030287>.
- [13] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009. ISSN 1615-3375,1615-3383. doi: 10.1007/s10208-009-9045-5. URL <https://doi.org/10.1007/s10208-009-9045-5>.

- [14] Emmanuel Candès and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007. ISSN 0266-5611,1361-6420. doi: 10.1088/0266-5611/23/3/008. URL <https://doi.org/10.1088/0266-5611/23/3/008>.
- [15] David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006. ISSN 0018-9448,1557-9654. doi: 10.1109/TIT.2006.871582. URL <https://doi.org/10.1109/TIT.2006.871582>.
- [16] 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>.
- [17] Benjamin Recht. A simpler approach to matrix completion. *J. Mach. Learn. Res.*, 12:3413–3430, 2011. ISSN 1532-4435,1533-7928.
- [18] 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.
- [19] K. C. Toh, M. J. Todd, and R. H. Tütüncü. SDPT3—a MATLAB software package for semidefinite programming, version 1.3. *Optim. Methods Softw.*, 11/12(1-4):545–581, 1999. ISSN 1055-6788,1029-4937. doi: 10.1080/10556789908805762. URL <https://doi.org/10.1080/10556789908805762>. Interior point methods.
- [20] Jos F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.*, 11/12(1-4):625–653, 1999. ISSN 1055-6788,1029-4937. doi: 10.1080/10556789908805766. URL <https://doi.org/10.1080/10556789908805766>. Interior point methods.
- [21] 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>.
- [22] Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.*, 20(4):1956–1982, 2010. ISSN 1052-6234,1095-7189. doi: 10.1137/080738970. URL <https://doi.org/10.1137/080738970>.
- [23] 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.
- [24] Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization (extended abstract). In *STOC’13—Proceedings of the 2013 ACM Symposium on Theory of Computing*, pages 665–674. ACM, New York, 2013. ISBN 978-1-4503-2029-0. doi: 10.1145/2488608.2488693. URL <https://doi.org/10.1145/2488608.2488693>.
- [25] 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>.
- [26] 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>.
- [27] 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>.
- [28] Raghunandan H. Keshavan and Sewoong Oh. A gradient descent algorithm on the grassman manifold for matrix completion, 2009. arXiv preprint <https://arxiv.org/abs/0910.5260>.

- 1192 [29] Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a
1193 few entries. *IEEE Trans. Inform. Theory*, 56(6):2980–2998, 2010. ISSN 0018-9448,1557-9654.
1194 doi: 10.1109/TIT.2010.2046205. URL <https://doi.org/10.1109/TIT.2010.2046205>.
- 1195 [30] Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from
1196 noisy entries. *J. Mach. Learn. Res.*, 11:2057–2078, 2010. ISSN 1532-4435,1533-7928.
- 1197 [31] Sourav Chatterjee. Matrix estimation by universal singular value thresholding. *Ann. Statist.*,
1198 43(1):177–214, 2015. ISSN 0090-5364,2168-8966. doi: 10.1214/14-AOS1272. URL <https://doi.org/10.1214/14-AOS1272>.
1199
- 1200 [32] Phuc Tran and Van Vu. New matrix perturbation bounds via combinatorial expansion i:
1201 Perturbation of eigenspaces, 2024. URL <https://arxiv.org/abs/2409.20207>. arXiv
1202 preprint <https://arxiv.org/abs/2409.20207>.
- 1203 [33] Sohom Bhattacharya and Sourav Chatterjee. Matrix completion with data-dependent miss-
1204 ingness probabilities. *IEEE Trans. Inform. Theory*, 68(10):6762–6773, 2022. ISSN 0018-
1205 9448,1557-9654. doi: 10.1109/tit.2022.3170244. URL <https://doi.org/10.1109/tit.2022.3170244>.
1206
- 1207 [34] Martin J. Wainwright. *Principal component analysis in high dimensions*, page 236–258.
1208 Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press,
1209 2019.
- 1210 [35] Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM*
1211 *J. Numer. Anal.*, 7:1–46, 1970. ISSN 0036-1429. doi: 10.1137/0707001. URL <https://doi.org/10.1137/0707001>.
1212
- 1213 [36] Per-Åke Wedin. Perturbation bounds in connection with singular value decomposition. *Nordisk*
1214 *Tidskr. Informationsbehandling (BIT)*, 12:99–111, 1972. ISSN 0901-246X. doi: 10.1007/
1215 bf01932678. URL <https://doi.org/10.1007/bf01932678>.
- 1216 [37] Sean O’Rourke, Van Vu, and Ke Wang. Random perturbation of low rank matrices: improving
1217 classical bounds. *Linear Algebra Appl.*, 540:26–59, 2018. ISSN 0024-3795,1873-1856. doi:
1218 10.1016/j.laa.2017.11.014. URL <https://doi.org/10.1016/j.laa.2017.11.014>.
- 1219 [38] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer.*
1220 *Statist. Assoc.*, 58:13–30, 1963. ISSN 0162-1459,1537-274X. URL [http://links.jstor.](http://links.jstor.org/sici?sici=0162-1459(196303)58:301<13:PIFSOB>2.0.CO;2-D&origin=MSN)
1221 [org/sici?sici=0162-1459\(196303\)58:301<13:PIFSOB>2.0.CO;2-D&origin=MSN](http://links.jstor.org/sici?sici=0162-1459(196303)58:301<13:PIFSOB>2.0.CO;2-D&origin=MSN).
- 1222 [39] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on
1223 the sum of observations. *Ann. Math. Statistics*, 23:493–507, 1952. ISSN 0003-4851. doi:
1224 10.1214/aoms/1177729330. URL <https://doi.org/10.1214/aoms/1177729330>.
- 1225 [40] BaoLinh Tran and Van Vu. The “power of few” phenomenon: the sparse case. *Random*
1226 *Structures Algorithms*, 66(1):Paper No. e21260, 16, 2025. ISSN 1042-9832,1098-2418. doi:
1227 10.1002/rsa.21260. URL <https://doi.org/10.1002/rsa.21260>.