A A deeper comparison with the current methods

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

A.1 The noiseless case

A.1.1 Nuclear norm minimization

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

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

Unfortunately, this problem is NP-hard, and all existing algorithms take doubly exponential time in terms of the dimensions of A [41]. To overcome this problem, Candes and Recht [1], motivated by an idea from the *sparse signal recovery* problem in the field of *compressed sensing* [42, 43], proposed to replace the rank with the nuclear norm of X, leading to

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

The paper [1] was shortly followed by Candes and Tao [2], with both improvements and trade-offs, and ultimately by Recht [3], who improved both previous results, proving that A is the unique solution to (9), given the sampling size bound

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

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

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 the optimal power of N while missing slightly from the optimal powers of r and $\log N$.

The key advantage of replacing the rank in Problem (8) with the nuclear norm is that Problem (9) is a convex program, which can be further translated into a semidefinite program [1, 2], solvable in polynomial time by a number of algorithms. However, convex optimization program usually runs slowly in practice. The survey [15] mentioned the interior point-based methods SDPT3 [44] and SeDuMi [45], which can take up to $O(|\Omega|^2 N^2)$ floating point operations (FLOPs) assuming (10), even if one takes advantage of the sparsity of A_Ω. Indeed, as Ω is at least $CN \log N$ (by coupon collector), the number of operations is $\Omega(N^4)$, which is too large even for moderate N. An *iterative singular value thresholding* method aiming to solve a regularized version of nuclear norm minimization, trading exactness for performance, has been proposed [46]. This motivates further research in the area to find faster methods. In what follows we discuss two other methods, which provide faster algorithms, but require extra assumptions to work.

A.1.2 Modified alternating projections

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

$$\text{minimize} \quad \|(A - XY^T)_\Omega\|_F^2 \quad \text{over } X \in \mathbb{R}^{m\times r}, Y \in \mathbb{R}^{r\times n}. \tag{11}$$

This, unfortunately, like (8), is NP-hard [16]. There have been many studies proposing variants of *alternating projections*, all of which involve the following basic idea: suppose one already obtains an approximator $X^{(l)}$ of X at iteration l, then $Y^{(l)}$ and $X^{(l+1)}$ are defined by

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

The survey [16] pointed out that these methods tend to outperform nuclear norm minimization in practice. Oh the other hand, there are few rigorous guarantees for recovery. The convergence and final output of the basic algorithm above also depends highly on the choice of $X^{(0)}$ [16].

Jain, Netrapalli and Sanghavi (2012) [24] developed one of the first alternating projections variants for matrix completion with rigorous recovery guarantees. They proved that, under the same setting in 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 gives us an exact recovery if we set $\varepsilon = \varepsilon_0/3$, where ε_0 is the precision level of A; see subsection **??**.

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 [5] and again by Hardt and Wooters [4] to logarithmic, at the cost of an increase in the powers of r, μ_0 and $\log N$.

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, [47]) 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 ouputs, 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.1.3 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 $\mathbf{E}\left[A_\Omega/p\right] = A$, 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 [7] 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 $\widetilde{A_\Omega}$.
2. *Low-rank approximation:* Compute the best rank-r approximation of $\widetilde{A_\Omega}$ via truncated SVD. Let $\mathsf{T}_r(\widetilde{A_\Omega}) = \tilde{U}_r\tilde{\Sigma}_r\tilde{V}_r^T$ be the output.
3. *Cleaning:* Solve for X, Y, S in the following optimization problem:

$$\text{minimize} \quad \left\|A_\Omega - (XSY^T)_\Omega\right\|_F^2 \quad \text{for} \quad X \in \mathbb{R}^{m\times r}, Y \in \mathbb{R}^{n\times r}, S \in \mathbb{R}^{r\times r}, \qquad (12)$$

using a gradient descent variant [7], starting with $X_0 = \tilde{U}_r$, $Y_0 = \tilde{V}_r$ 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 $X_*S_*Y_*^T$.

The last cleaning step resembles the optimization problem in alternating projections methods, but here the authors used gradient descent instead. They [7] showed that the algorithm returns an output arbitrarily close to A, given *enough* iterations in the cleaning step, provided the following sampling size condition:

$$|\Omega| \geq C \max\left\{\mu_0\sqrt{mn}\left(\frac{\sigma_1}{\sigma_r}\right)^2 r\log N, \quad \max\{\mu_0, \mu_1\}^2 r \min\{m,n\}\left(\frac{\sigma_1}{\sigma_r}\right)^6\right\}. \qquad (13)$$

It was pointed out [7] that the powers of r and $\log N$ are optimal by the coupon-collector limit, answering a question from [2]. On the other hand, the bound depends heavily on the condition number $\kappa := \sigma_1/\sigma_r$. Furthermore, similar to the situation in the previous subsection, one needs to know the rank r in advance; see Remark A.1.

In a later paper [47], Keshavan and Oh showed that one can compute r (with high probability) if the condition number satisfies $\kappa = O(1)$; see also Remark A.1. Thus, it seems that the critical extra assumption for this algorithm (apart from the three basic assumptions) to be efficient is that the singular values of A are of the same order of magnitude, i.e., $\kappa = O(1)$. This assumption is strong, and we do not know how often it holds in practice. In fact, Figure 1.3 shows that the centered data matrix made from the Yale face dataset has a large condition number.

From the complexity point of view, the first part (low rank approximation) of the algorithm is very fast, in both theory and practice, as it used truncated SVD only once. On the other hand, [7] did not provide a full convergence rate analysis of their (cleaning) gradient descent part. It only briefly mentioned that quadratic convergence is possible [7, Page 14].

A.1.4 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 inifnity norm. It has been shown, under different extra assumptions, that low rank approximation fullfils this purpose.

The first infinity norm result was obtained by Abbe, Fan, Wang, and Zhong [9]. They showed that the best rank-r approximation of $p^{-1}A_\Omega$ is close to A in the infinity norm [9, 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 [10] used different mathematical tools to analyzed 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**.)

As discussed, the main improvement of **AR** 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 [10] for more discussion). However, we do not know how often matrices in practice satisfy it. In fact, Figure 1.3 shows that the Yale face database matrix [26] 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 its suggests that the singular values decay fast, and the condition number is also large. So, from the mathematical view point, the situation is quite intriguing. We have two valid theorems with *constrasting extra assumptions* (beyond the three basic assumptions). The most logical explanation here should be that neither assumption is in fact needed. This conjecture, in the (more difficult) noisy setting, presented in the next section, is the motivation of our study.

A.2 Recovery with Noise

Candes and Plan, in their influential survey [8], pointed out that data is often noisy. Thus, a more realistic model for the recovery problem is to consider $A' = A + Z$, for A being the low rank ground truth matrix and Z the noise. We observe a sparse matrix A'_Ω, where each entry of A' is observed with probability p and set to 0 otherwise. In other words, we have access to a small random set of noisy entries. Notice that in this case, the truth matrix A is still low rank, but the noisy matrix A', whose entries we observe, can have full rank. In what follow, we denote our input by $A_{\Omega,Z}$, emphasizing the presence of the noise.

Recovery from noisy observation is clearly a harder problem, and most papers concerning noisy recovery aim for recovery in *the normalized Frobenius norm* (root mean square error; RMSE), rather than exact recovery.

Continuing the nuclear norm minimization approach, Candes and Plan [8] adapted to the noisy situation by relaxing the constraint on the observations, leading to the following problem:

$$\text{minimize} \quad \|X\|_* \quad \text{subject to} \quad \|X_\Omega - A_{\Omega,Z}\|_F \leq \delta, \tag{14}$$

where δ is a known upper bound on $\|Z_\Omega\|_F$. The authors showed that, under the same sample size condition in [3], with probability $1 - o(1)$, the optimal solution $\hat{A}$ satisfies

$$\frac{1}{\sqrt{mn}} \|\hat{A} - A\|_F \leq C \|Z_\Omega\|_F \sqrt{\frac{\min\{m,n\}}{|\Omega|}}. \tag{15}$$

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

$$|\Omega| \geq C \frac{\|Z_\Omega\|_F^2 \min\{m,n\}}{\varepsilon^2}, \tag{16}$$

which grows quadratically with $1/\varepsilon$.

For exact recovery, one needs to turn the approximation in the Frobenius norm into an approximation in the infinity norm; see subsection **??**. This is a major mathematical challange, and we do not know any efficient way to do this. The trivial bound that $\|M\|_\infty \leq \|M\|_F$ is too generous. If we use this and then use (15) to bound the RHS, then the corresponding bound on $|\Omega|$ in (16) becomes larger than mn, which is meaningless. This is the common situation with all Frobenius norm bounds discussed in this section.

Concerning the alternating method, a corollary of [4, Theorem 1] shows that we can obtain an approximation $\hat{A}$ of rank r, where

$$\|\hat{A} - A\| \leq (2 + o(1))\|Z\| + \varepsilon\sigma_1, \tag{17}$$

given that

$$p = \tilde{\Omega}\left(\frac{1}{n}\Big(1 + \frac{\|Z\|_F}{\varepsilon\sigma_1}\Big)\right)^2.$$

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

Continuing the spectral approach, Keshavan, Montanari and Oh [6] also extended their result from [7] to the noisy case, using the same algorithm. They proved that with the same sample size condition as (13), the output satisfies w.h.p.

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

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

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

where the dependence on ε is again quadratic.

Chatterjee [22] uses a spectral approach with a fixed truncation point independent from $\operatorname{rank} A$, and thus does not require knowing it. He required that $p \geq CN^{-1}\log^6 N$ and achieved the bound

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

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

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

The dependence on ε grows substantially faster than [6]. In fact, P. Tran's and Vu's recent result in random perturbation theory [27] can be used to prove a high-probability mean-squared-error bound, again without incoherence, requiring only a quadratic dependence on ε.

If we insist on exact recovery, the only approach which adapts well to the noisy situation is the infinity norm approach. As a matter of fact, the infinity norm bounds presented in Section 2.4 hold in both noiseless and noisy case (with some modification). The reason is this: even in the noiseless case, one already views the (rescalled) 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.

The result by Abbe et al. [9, Theorem 3.4] yields the same approximation as in the noiseless case, given

$$p \geq C^2\varepsilon^{-2}\mu_0^4\kappa^8(\|A\|_\infty + \sigma_Z)^2 N^{-1}\log N, \tag{20}$$

where σ_Z is the standard deviation of each entry of Z. If we set $\varepsilon < \varepsilon_0/3$ (see subsection **??**), then again rounding off would give as an exact recovery. Similarly, algorithm **AR** works in the noisy case; see [10] for the exact statement. The paper only proves this fact for the case where A has integer entries, but a careful examination of the proof shows that for a general absolute error tolerance ε, one requires $m, n = \Theta(N)$ and

$$p \geq C\varepsilon^{-5} \max\left\{C'(r, \|A\|_\infty, \mu), \log^{3.03} N\right\} N^{-1} \log N, \tag{21}$$

where $\mu = N^{1/2} \max\{\|U\|_\infty, \|V\|_\infty\}$ and C' is a term depending on r, $\|A\|_\infty$ and μ only.

Summary. To summarize, in the noisy case, the infinity norm approach is currently the only one that yields exact recovery. The latest results in this directions, [9] and [10], however, requires the extra assumptions that the condition number is small and the gaps are large, respectively. As disucssed at the end of the previous subsection, these conditions constrast each other, and we conjecture that both of them could be removed. This leads to the main question of this paper:

Question 1. *Can we use the infinity norm approach to obtain exact recovery in the noisy case with only the three basic assumptions (low rank, incoherence, density) ?*

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 [27], 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 theorey 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 $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_r$. 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** $\sin\Theta$ **theorem**, proven by Davis and Kahan [37], 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 [38], states that:

$$\|\tilde{U}_S\tilde{U}_S^T - U_S U_S^T\| \vee \|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\| \leq \frac{C\|E\|}{\Delta_S}. \tag{22}$$

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. [39], Tran and Vu [27] 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:

$$\|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\| \leq C\sqrt{|S|}\left(\frac{\|E\|}{\sigma_S} + \frac{\sqrt{r}\|U^T E V\|_\infty}{\Delta_S} + \frac{\|E\|^2}{\Delta_S \sigma_S}\right), \tag{23}$$

with high probability, effectively turning the *noise-to-gap* on the right-hand side of Eq. (22) 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

$$\frac{\|E\|}{\sigma_S} \vee \frac{2r\|U^T E V\|_\infty}{\Delta_S} \vee \frac{\sqrt{2r}\|E\|}{\sqrt{\Delta_S \sigma_S}} \leq \frac{1}{8}, \tag{24}$$

then

$$\|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\| \leq C r R_S, \text{ where } R_S := \frac{\|E\|}{\sigma_S} + \frac{2r\|U^T E V\|_\infty}{\Delta_S} + \frac{2ry}{\Delta_S \sigma_S}, \tag{25}$$

where

$$y := \frac{1}{2}\max_{i \neq j}(|u_i^T E E^T u_j| + |v_i^T E^T E v_j|) \tag{26}$$

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* [27].

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:*

$$\begin{aligned}
\tau_1 &:= \max_{0 \leq a \leq 10\log(m+n)} \frac{1}{\sqrt{r}} \max\left\{\frac{\left\|(EE^T)^a U\right\|_{2,\infty}}{\|E\|^{2a}}, \frac{\left\|(EE^T)^a E V\right\|_{2,\infty}}{\|E\|^{2a+1}}\right\}, \\
\tau_2 &:= \max_{0 \leq a \leq 10\log(m+n)} \frac{1}{\sqrt{r}} \max\left\{\frac{\left\|(E^T E)^a V\right\|_{2,\infty}}{\|E\|^{2a}}, \frac{\left\|(E^T E)^a E^T U\right\|_{2,\infty}}{\|E\|^{2a+1}}\right\}.
\end{aligned} \tag{27}$$

Suppose an arbitrary subset $S \subset [r]$ satisfies Eq. (24). *Then for a universal constant C,*

$$\left\|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\right\|_\infty \leq C\tau_1^2 r R_S, \tag{28}$$

$$\left\|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\right\|_{2,\infty} \leq C\tau_1 r R_S, \tag{29}$$

where R_S is defined in Eq. (25). *When $S = [s]$ for some $s \in [r]$, we also have*

$$\|\tilde{A}_s - A_s\|_\infty \leq C\tau_1\tau_2\sigma_s r R_s, \text{ where } R_s := R_{[s]}. \tag{30}$$

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

We use Eq. (30) to prove Theorem 3.1. While Eqs. (28) and (29) 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 [27] 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. (25).

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 [27] 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 (24). This bound was discovered by O'Rourke et al. [39] and was the best-known until [27]. 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 $j \in [n]$,

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

y can be much smaller than $\|E\|$ (see [27] 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:*

$$\mathbf{E}\left[E_{ij}\right] = 0, \quad \mathbf{E}\left[|E_{ij}|^2\right] \le \varsigma^2, \quad \mathbf{E}\big[E_{ij}|^l\big] \le M^{l-2}\varsigma^l \quad \textit{for all } \ l \in \mathbb{N}_{\ge 2}. \tag{31}$$

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

$$\tau_0(U,p) := \frac{p\|U\|_{2,\infty}}{\sqrt{r}}, \quad \tau_1(U,p) := \frac{Mp^3\|U\|_{2,\infty}}{\sqrt{rN}} + \frac{p^{3/2}}{\sqrt{N}}. \tag{32}$$

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

$$\max_{0\le a\le t\log N} \|e_{m,k}^T(EE^T)^a U\| \le \tau_0(U, \log\log N)\ \mathcal{H}^{2a}\sqrt{r}. \tag{33}$$

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

$$\max_{0\le a\le t\log N} \|e_{n,k}^T(E^TE)^a E^T U\| \le \tau_1(U, \log\log N)\ \mathcal{H}^{2a+1}\sqrt{r}. \tag{34}$$

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

$$\max_{0\le a\le t\log N}\max_{k\in[m]} \|e_{m,k}^T(EE^T)^a U\| \le \tau_0(U, \log N)\ \mathcal{H}^{2a}\sqrt{r}, \tag{35}$$

$$\max_{0\le a\le t\log N}\max_{k\in[n]} \|e_{n,k}^T(E^TE)^a E^T U\| \le \tau_1(U, \log N)\ \mathcal{H}^{2a+1}\sqrt{r} \tag{36}$$

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. (35) and (36) to prove Theorem 3.1, but for the sake of potential future applications, we still present Eqs. (33) and (34), whose bounds are better but non-uniform in k. More specifically, we use the following bounds, which directly results from the theorem.

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

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

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

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

$$\begin{aligned}
\left\|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\right\|_\infty &\leq C\frac{\mu(V)}{n}\left\|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\right\|, \\
\left\|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\right\|_{2,\infty} &\leq C\sqrt{\frac{\mu(V)}{n}}\left\|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\right\|, \\
\left\|\tilde{A}_s - A_s\right\|_\infty &\leq C\sqrt{\frac{\mu(U)\mu(V)}{mn}}\left\|\tilde{A}_s - A_s\right\|.
\end{aligned} \tag{37}$$

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. (28) asserts that

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

On the other hand, we have

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

Therefore, our bound says that in the best case scenario, the largest entry of the matrix is of the same magnitude as the average one, making Eq. (28) sharp. The sharpness (in the best case) of Eq. (29) and Eq. (30) can be argued similarly. Notice that this also fully justifies the optimality of the bound (**??**) in Theorem 3.1.

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 $\|U^T E V\|_\infty$, 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* (31) *with parameters M and ς. Let $N = m + n$. Replace τ_1 from Eq.* (27) *with*

$$\tau_1 := \frac{\|U\|_{2,\infty}\log N}{\sqrt{r}} + \frac{M\|V\|_{2,\infty}\log^3 N}{\sqrt{rN}} + \frac{\log^{3/2} N}{\sqrt{N}}, \tag{38}$$

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

$$\frac{\varsigma\sqrt{N}}{\sigma_S} \vee \frac{r\varsigma(\sqrt{\log N} + M\|U\|_\infty\|V\|_\infty \log N)}{\Delta_S} \vee \frac{\varsigma\sqrt{rN}}{\sqrt{\Delta_S\sigma_S}} \leq \frac{1}{16}. \tag{39}$$

Let us replace the term R_S in Eq. (25) *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 $1 - O(N^{-1})$,

$$\left\|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\right\|_\infty \leq C\tau_2^2 r R_S + \frac{1}{N}, \tag{40}$$

$$\left\|\tilde{V}_S\tilde{V}_S^T - V_S V_S^T\right\|_{2,\infty} \leq C\tau_2 r R_S + \frac{1}{N}. \tag{41}$$

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 $1 - O(N^{-1})$,

$$\|\tilde{A}_s - A_s\|_\infty \le C\tau_1\tau_2 r\sigma_s R_s + \frac{1}{N}. \tag{42}$$

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

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

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, $p \ge C(m^{-1} + n^{-1})\log^{10} N$ implies $M \le c\sqrt{N}\log^{-5} N$, meaning we can apply Theorem B.6, specifically Eq. (42) for this choice of ς and M if the condition (39) holds. We check it for $S = [s]$. Given that $\sigma_s \ge \delta_s \ge 40K\sqrt{rN/p}$, we have

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

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

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

It remains to transform the right-hand side of Eq. (42) to the right-hand side of Eq. (7). We have

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

Combining with the symmetric bound for τ_2, we get

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

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

Consider the term R_s. From the above, we have

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

Since $\delta_s \ge 40K\sqrt{rN/p}$, the fourth term is 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 $r_{\max} \le \log^2 N$ 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:* $\sigma_1 \geq 100 r\kappa\sqrt{r_{\max}N}$.
2. *Sampling density:* this is equivalent to the conjunction of three conditions:

$$p \geq \frac{Cr^4 r_{\max}\mu_0^2 K_{A,Z}^2}{\varepsilon^2}\left(\frac{1}{m}+\frac{1}{n}\right), \tag{43}$$

$$p \geq C\left(\frac{1}{m}+\frac{1}{n}\right)\log^{10} N, \tag{44}$$

$$p \geq \frac{Cr^3 K_{A,Z}^2}{\varepsilon^2}\left(1+\frac{\mu_0^2}{\log^2 N}\right)\left(1+\frac{r^3\log N}{N}\right)\left(\frac{1}{m}+\frac{1}{n}\right)\log^6 N. \tag{45}$$

Let $\rho := \hat{p}/p$. From the sampling density assumption, a standard application of concentration bounds [36, 48] guarantees that, with probability $1-O(N^{-2})$.

$$0.9 \leq 1-\frac{1}{\sqrt{N}} \leq 1-\frac{\log N}{\sqrt{pmn}} \leq \rho \leq 1+\frac{\log N}{\sqrt{pmn}} \leq 1+\frac{1}{\sqrt{N}} \leq 1.1. \tag{46}$$

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

$$\|E\| \leq 2\kappa\sqrt{N}, \tag{47}$$

with probability $1-O(N^{-1})$. See [35, 34], [40, Lemma A.7] or [35] for detailed proofs. Therefore we can assume both Eqs. (46) and (47) at the cost of an $O(N^{-1})$ exceptional probability.

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

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

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

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

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

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

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

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

   $$\sigma_{s+1} \leq r\delta_{s+1} \leq r(\tilde{\delta}_{s+1}+2\|E\|) \leq r(C_0\rho^{-1/2}\sqrt{r_{\max}}+4)\kappa\sqrt{N} \leq 2rC_0K_{A,Z}\sqrt{r_{\max}N/p}. \tag{48}$$

   For each fixed indices j,k, we have

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

   where the last inequality comes from the assumption (43) if C is large enough. Since this holds for all pairs (j,k), we have $\|A_s - A\|_\infty \leq \varepsilon/4$.

2. Secondly, we will show that $\|\tilde{A}_s - A_s\|_\infty \leq \varepsilon/4$ with probability $1-O(N^{-1})$. We aim to use Theorem 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:

   $$\sigma_s \geq \delta_s \geq \tilde{\delta}_s - 2\|E\| \geq C_0\rho^{-1/2}\kappa\sqrt{r_{\max}N} - 2\|E\| \geq .9C_0\kappa\sqrt{r_{\max}N}. \tag{49}$$

Consider the condition (39). If it holds, then we can apply Theorem B.6. We want

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

By Eq. (49), we can replace all three denominators above with $.9C_0\kappa\sqrt{r_{\max}N}$. Additionally, $\|U\|_\infty \leq \|U\|_{2,\infty} \leq \sqrt{\frac{r\mu_0}{m}}$ and $\|V\|_\infty \leq \|V\|_{2,\infty} \leq \sqrt{\frac{r\mu_0}{n}}$, so we can replace them with these upper bounds. We also replace K with $p^{-1/2}$ (its definition). We want

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

which is equivalent to

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

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

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

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

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

where the first inequality comes from (44) if C is large enough. Similarly,

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

Therefore,

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

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

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

Under the condition (45), 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,

$$\begin{aligned}\|\tilde{A}_s - A_s\|_\infty &\le \log^2 N \cdot \frac{\log N + 4\mu_0}{\sqrt{mn}} \cdot \frac{\sqrt{2}r^{3/2}K_{A,Z}}{\sqrt{p}} \left(\sqrt{N} + r^{3/2}\sqrt{\log N}\right) \\ &\le \sqrt{\frac{2r^3 K_{A,Z}^2 \log^6 N}{p}\left(1 + \frac{4\mu_0^2}{\log^2 N}\right)\left(1 + \frac{r^3 \log N}{N}\right)\left(\frac{1}{m} + \frac{1}{n}\right)} \le \varepsilon/4.\end{aligned} \tag{50}$$

where the last inequality comes from the condition (45) if C is large enough.

After the two steps above, we obtain $\|\tilde{A}_s - A\|_\infty \le \varepsilon/2$ with probablity $1 - O(N^{-1})$. Finally, we get, using Fact (46) and the triangle inequality,

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

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

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 [27], but with highly non-trivial adjustments to handle the inifnity norm, ínstead of spectral norm as in [27]. 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 (24). 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 (31), with the assumption $M \le (m+n)^{1/2}\log^{-5}(m+n)$, the moment argument in [34] can be used.
- $\|U^T EV\|_\infty = \max_{i,j} |u_i^T E v_j|$. These terms can be bounded with a simple Bernstein bound.
- $y = \frac{1}{2}\max_{i \ne j}(|u_i EE^T u_j| + |v_i E^T E v_j|)$. 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* (31) *with parameters M and ς. Suppose $M \le (m+n)^{1/2}\log^{-3}(m+n)$. Then with probability $1 - O((m+n)^{-2})$, all of the following hold:*

$$\|E\| \le 1.9\varsigma\sqrt{m+n} \le 2\varsigma\sqrt{m+n}, \tag{51}$$

$$\max_{i\neq j}(|u_i EE^T u_j| + |v_i E^T E v_j|) \le 2\|E\|^2 \le 8\varsigma^2(m+n). \tag{52}$$

$$\max_{i,j}|u_i^T E v_j| \le 2\varsigma(\sqrt{\log(m+n)} + M\|U\|_\infty\|V\|_\infty \log(m+n)). \tag{53}$$

Proof. Eq. (51) follows from the moment argument in [34]. Eq. (52) follows from Eq. (51). It remains to check Eq. (53). Fix $i, j \in [r]$. Write

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

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

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

We also have

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

By Bernstein's inequality [48], we have for all $t > 0$

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

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

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

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

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

thus

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

Combining everything above, we get

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

By a union bound over $(i,j) \in [r]\times[r]$, the proof of Eq. (53) and the lemma is complete. ☐

Now all that remains is computing τ_1 and τ_2. More precisely, since both are random, we compute a good choice of high-probability upper bounds for them. This, however, is likely intractable since the appearance of powers of $\|E\|$ in the denominator makes it hard to analyze the right-hand sides of Eq. (27). To overcome this, notice that the argument in Theorem B.2 works in the same way if, instead of being rigidly refined by Eq. (27), τ_1 and τ_2 are any real numbers satisfying

$$\begin{aligned}\tau_1 &\ge \max_{a\in[[0,10\log(m+n)]]}\frac{1}{\sqrt{r}}\max\left\{\frac{\left\|(EE^T)^aU\right\|_{2,\infty}}{\mathcal{H}^{2a}}, \frac{\left\|(EE^T)^aEV\right\|_{2,\infty}}{\mathcal{H}^{2a+1}}\right\},\\ \tau_2 &\ge \max_{a\in[[0,10\log(m+n)]]}\frac{1}{\sqrt{r}}\max\left\{\frac{\left\|(E^TE)^aV\right\|_{2,\infty}}{\mathcal{H}^{2a}}, \frac{\left\|(E^TE)^aE^TU\right\|_{2,\infty}}{\mathcal{H}^{2a+1}}\right\},\end{aligned} \tag{54}$$

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

From this point, we will discard Eq. (27) and treat $(\tau_1, \tau_2, \mathcal{H})$ as any tuple that satisfies Eq. (54). Specifically, we will choose $\tau_0(U)$, $\tau_1(U)$, $\tau_0(V)$, $\tau_1(V)$ such that

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

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

This is equivalent to bounding terms of the form

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

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 (24) 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 (39) 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 (39) holds, then (24) 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 (54) above, if we can show that τ_1, τ_2 and $\mathcal{H}$ satisfy (54), then the argument in Theorem B.2 still works. By Theorem B.4 for $t = 10$, (54) holds with probability $1 - O((m+n)^{-2})$, so the proof is complete. ☐

In the next section, we prove Theorem B.2. The proof is an adaptation of the main argument in [27] 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.

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

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

Given A and $\tilde{A} = A + E$, there are three terms we need to bound, correspoding to Eqs. (28), (29) and (30):

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

The strategy of bounding all three are almost identical, and is an extension to the SVD case of the strategy for the eigendecomposition in [27].

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

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

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

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

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

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

The authors of [27] came up with a clever idea. Imagine first, for simplicity, that both A and E are square symmetric matrices, and that V and $\tilde{V}$ contain the eigenvector, rather than singular vectors, of their respective matrices. In other words, $U = V$ and the numbers σ_i are temporarily viewed as eigenvalues. Instead of measuring the difference $g(\tilde{A}) - g(A)$ directly, they considered the difference of the *Stieltjes transforms*, and obtained the expansion:

$$(zI - \tilde{A})^{-1} - (zI - A)^{-1} = \sum_{\gamma=1}^{\infty} [(zI-A)^{-1}E]^{\gamma}(zI-A)^{-1}. \tag{55}$$

It is easy to show that this identity hold whenever the right-hand side converges. Conveniently, the convergence is also guaranteed by the condition (24) of Theorem B.2, as we will see later. To obtain $\tilde{V}_S\tilde{V}_S^T$ and $V_SV_S^T$, rewrite the left-hand side of Eq. (55) as

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

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

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

The precise details on how to choose this contour can be found in [27]. The final steps to bound the left-hand side will be:

1. Expand the right-hand side into sums involving products of E and $v_iv_i^T$ and $Q = I - VV^T$.
2. Bound each product by estimating the scalar contour integral and the norm of each factor.

Back to the context in this paper, where we handle the SVD instead of the eigendecomposition. In [27], the author used this expansion to obtain a bound on the spectral norm of the left-hand side by bounding each term in the series. We make appropriate adjustments to their argument to adapt it to the SVD, while also proving a novel *semi-isotropic* bound on powers of random matrices to extend the result to the infinity norm.

C.2.1 The power series expansion for the SVD case

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

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

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

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

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

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

as an eigendecomposition of A_{sym}. We have

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

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

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

We define $\tilde{\sigma}_i$, $\tilde{w}_i$ and $\tilde{W}_s$, $\tilde{W}_S$, $\tilde{A}_s$, $\tilde{A}_S$ similarly for $\tilde{A} = A + E$. From Eq. (55) for the symmetric case, we have the expansion

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

which is equivalent to

$$\sum_i \frac{\tilde{w}_i \tilde{w}_i^T}{z - \tilde{\sigma}_i} - \sum_{|i| \in [r]} \frac{w_i w_i^T}{z - \sigma_i} = \sum_{\gamma=1}^{\infty} \left[(zI - A_{\text{sym}})^{-1} E_{\text{sym}}\right]^{\gamma} (zI - A_{\text{sym}})^{-1}. \tag{56}$$

Let Γ_S denote a contour in $\mathbb{C}$ that encircles $\{\pm\sigma_i, \pm\tilde{\sigma}_i\}_{i \in S}$ and none of the other eigenvalues of $\tilde{W}$ and W, satisying that the right-hand side of Eq. (56) converges for every z on the contour. Integrating over Γ_S of both sides and dividing by $2\pi i$, we have

$$\begin{aligned} \begin{bmatrix} \tilde{U}_S \tilde{U}_S^T - U_S U_S^T & 0 \\ 0 & \tilde{V}_S \tilde{V}_S^T - V_S V_S^T \end{bmatrix} &= \tilde{W}_S \tilde{W}_S - W_S W_S^T \\ &= \sum_{\gamma=1}^{\infty} \oint_{\Gamma_S} \frac{\mathrm{d}z}{2\pi i} \left[(zI - A_{\text{sym}})^{-1} E_{\text{sym}}\right]^{\gamma} (zI - A_{\text{sym}})^{-1}. \end{aligned} \tag{57}$$

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

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

we have the expansion

$$(\tilde{V}_S \tilde{V}_S^T - V_S V_S^T)_{jk} = \sum_{\gamma=1}^{\infty} \oint_{\Gamma_S} \frac{\mathrm{d}z}{2\pi i} e_{m+n,m+j}^T \left[(zI - A_{\text{sym}})^{-1} E_{\text{sym}}\right]^{\gamma} (zI - A_{\text{sym}})^{-1} e_{m+n,m+k}, \tag{58}$$

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 [27]. The goal is still the same, but our expansion will be different from [27], 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. (58), the term with power γ becomes

$$\oint_{\Gamma_S} \frac{\mathrm{d}z}{2\pi i} e_{m+n,m+j}^T \left[\Big(\sum_{|i| \in [r]} \frac{\sigma_i P_i}{z(z - \sigma_i)} + \frac{I}{z} \Big) E_{\text{sym}} \right]^{\gamma} \Big(\sum_{|i| \in [r]} \frac{\sigma_i P_i}{z(z - \sigma_i)} + \frac{I}{z} \Big) e_{m+n,m+k}. \tag{59}$$

When expanding the above, we get monomials of the form

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

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

$$\begin{aligned} &\left[\oint_{\Gamma_S} \frac{\mathrm{d}z}{2\pi i} \frac{1}{z^{\alpha_0 + \beta_0 + \alpha_1 + \dots + \beta_{h-1} + \alpha_h}} \overbrace{\frac{\sigma_?}{z - \sigma_?} \frac{\sigma_?}{z - \sigma_?} \cdots \frac{\sigma_?}{z - \sigma_?}}^{\beta_1 + \beta_2 + \dots + \beta_h \text{ factors}} \right] \\ &e_{m+n,m+j}^T E_{\text{sym}}^{\alpha_0} \Bigg(\underbrace{P_? E_{\text{sym}} P_? E_{\text{sym}} \dots P_? E_{\text{sym}}}_{\beta_1 \text{ factors}} \Bigg) E_{\text{sym}}^{\alpha_1} \dots \Bigg(\underbrace{P_? E_{\text{sym}} P_? E_{\text{sym}} \dots P_?}_{(\beta_h - 1)\ E_{\text{sym}} \text{ factors}} \Bigg) E_{\text{sym}}^{\alpha_h} e_{m+n,m+k}, \end{aligned} \tag{60}$$

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

$$\begin{aligned} &\big(e_{m+n,m+j}^T E_{\text{sym}}^{\alpha_0} w_? \big) \underbrace{\big(w_?^T E_{\text{sym}} w_? \dots w_?^T E_{\text{sym}} w_? \big)}_{(\beta_1 - 1)\ E_{\text{sym}} \text{ factors}} \big(w_?^T E_{\text{sym}}^{\alpha_1 + 1} w_? \big) \\ &\dots \ \big(w_?^T E_{\text{sym}}^{\alpha_{h-1} + 1} w_? \big) \underbrace{\big(w_?^T E_{\text{sym}} w_? \dots w_?^T E_{\text{sym}} w_? \big)}_{(\beta_h - 1)\ E_{\text{sym}} \text{ factors}} \big(w_?^T E_{\text{sym}}^{\alpha_h} e_{m+n,m+k} \big) \end{aligned} \tag{61}$$

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 $w_?^T E_{\text{sym}} w_?$ give rise to the term $\|U^T E V\|_\infty$ in Eq. (24).
- The terms $w_?^T E_{\text{sym}}^{\alpha_i + 1} w_?$ mostly give rise to the term $\|E\|$, but in the special cases where $\alpha_i = 1$ for all i will be more strongly bounded with the term y in R_3.

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

C.2.2 Notation and roadmap

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

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

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

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

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

- Define $\tilde{w}_i$ similarly, with $\operatorname{rank} \tilde{A}$ instead of r.

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

- Γ_S is a contour encircling precisely the set $\{\sigma_i, \tilde{\sigma}_i : |i| \in S\}$ and no other eigenvalues, such that the right-hand side of Eq. (56) converges absolutely for all z on it.

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

$$\begin{aligned} &\bullet\quad \alpha_0, \alpha_h \geq 0, \quad \text{ and } \alpha_k \geq 1 \text{ for } 1 \leq k \leq h-1, \\ &\bullet\quad \beta_k \geq 1 \text{ for } 1 \leq k \leq h, \\ &\bullet\quad \alpha + \beta = \gamma + 1, \quad \text{ where } \alpha := \textstyle\sum_{k=0}^h \alpha_k, \quad \text{ and } \beta := \textstyle\sum_{k=1}^h \beta_k. \end{aligned} \tag{62}$$

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

- For each $\boldsymbol{\beta}$ above satisfying each $\beta_k \geq 1$, we use $\mathbf{I} = [i_1, i_2, \ldots, i_\beta]$ for an element of $[\pm r]^\beta$. Together, the triple $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{I})$ define uniquely a monomial of the form (60). Define $\mathbf{I}_{a:b}$ as the subsequence $[i_a, i_{a+1}, \ldots, i_b]$.

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

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

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

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

- Define the following terms:

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

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

From Eqs. (58), (59) and (60), we have

$$(\tilde{V}_S \tilde{V}_S^T - V_S V_S^T)_{jk} = e_{m+n,m+j}^T \mathcal{T} e_{m+n,m+k}. \tag{63}$$

At this point, we look at the larger context of Theorem B.2. Consider Eq. (29). To bound $\|\tilde{V}_S \tilde{V}_S^T - V_S V_S^T\|_{2,\infty}$, we can fix one index j and find a bound for its j^{th} row that holds with probability close enough to 1 to beat the n factor from the union bound. We have

$$(\tilde{V}_S \tilde{V}_S^T - V_S V_S^T)_{j,\cdot} = e_{m+n,m+j}^T \mathcal{T}. \tag{64}$$

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. (28) and (29).

Finally, consider Eq. (30). 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 (28) and (29). The details will be given later, when we prove (30).

C.2.3 Bounding the change in singular subspace expansions

Let us prove Eqs. (28) and (29) 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 (28) and (29).

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:*

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

We have,

$$|\mathcal{C}(\mathbf{I})| \le \frac{2^{\gamma+\beta-1}}{\sigma_S(\mathbf{I})^{\gamma+1-\beta}\Delta_S(\mathbf{I})^{\beta-1}} \le \frac{2^{\gamma+\beta-1}}{\sigma_S^{\gamma+1-\beta}\Delta_S^{\beta-1}}. \tag{65}$$

In the steps that follow, we will mainly use the second bound of Eq. (65), 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* $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \Pi_h(\gamma)$ *and* $\mathbf{I} = \{i_k : k \in \beta\} \in [\pm r]^\beta$*. Then*

$$\|M^T \mathcal{M}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{I}) M'\| \le \|E\|^{\alpha-\alpha_0-\alpha_h+h-1} \cdot \|W^T E_{\text{sym}} W\|_\infty^{\beta-h} \cdot \|w_{i_1}^T E_{\text{sym}}^{\alpha_0} M\| \cdot \|w_{i_\beta}^T E_{\text{sym}}^{\alpha_h} M'\|. \tag{66}$$

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

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

$$\tau(M) = \max_{0 \le \alpha \le 10\log(m+n)} \frac{1}{2r} \sum_{|i| \in [r]} \frac{\|w_i^T E_{\text{sym}}^\alpha M\|}{\|E_{\text{sym}}\|^\alpha}, \qquad \textit{and analogously for } \tau(M'). \tag{67}$$

$$R_1 := \frac{\|E\|}{\sigma_S} \vee \frac{2r\|W^T E_{\text{sym}} W\|_\infty}{\Delta_S}, \quad R_2 := \frac{\sqrt{2r}\|E\|}{\sqrt{\sigma_S \Delta_S}}, \quad R_3 := \frac{2r \max_{|i| \ne |j|} |w_i E_{\text{sym}}^2 w_j|}{\sigma_S \Delta_S} \tag{68}$$

and assume that

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

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

$$\|M^T \mathcal{T}(\boldsymbol{\alpha}, \boldsymbol{\beta}) M'\| \le \begin{cases} r\tau(M)\tau(M') 2^{\gamma+\beta} R_1 R^{\gamma-1} & \textit{if } 1 \le h < \gamma/2 + 1, \\ 16 r\tau(M)\tau(M')(4R)^{\gamma-2}(R_3 + R_1^2) & \textit{if } h = \gamma/2 + 1. \end{cases} \tag{69}$$

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

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

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

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

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

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

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

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

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

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

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

and

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

- Similarly, recall the definitions of τ_1 and τ_2 in Eq. (27). As a function of M, τ satisfies

$$\tau(e_{m+n,k}) = \tau_1 \text{ for } k \leq m, \quad \tau(e_{m+n,k}) = \tau_2 \text{ for } m+1 \leq k \leq m+n, \quad \text{ and } \tau(I) \leq 1, \tag{70}$$

To summarize, the logical structure is:

$$\text{Lemma C.5} \xrightarrow{\text{implies}} \text{Lemma C.6} \xrightarrow{\text{implies}} \text{Lemma C.7} \xrightarrow{\text{implies}} \text{Eqs. (28), (29) in Theorem B.2}$$

We will finish the last step, which is the proof of (28) and (29) here. The proofs of Lemmas C.5, C.6 and C.7 will be postpone to Section C.3.

Proof of Theorem B.2 part I. Consider the objects defined in Theorem B.2 and the additional objects in Setting C.2. By the remark above, the condition (24) in Theorem B.2 is equivalent to $R_1 \vee R_2 \leq 1/4$, so we can apply the lemmas in this section.

Let us prove Eq. (28). Consider arbitrary $j, k \in [n]$. From Eq. (63), $(\tilde{V}_S \tilde{V}_S^T - V_S V_S^T)_{jk}$ is $M^T \mathcal{T} M'$ for $M = e_{m+n,j+m}$ and $M' = e_{m+n,k+m}$. We apply the bound Lemma C.7, while replacing both $\tau(M)$ and $\tau(M')$ with τ_2 (permissible by Eq. (70)), to get

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

where the last inequality is due to the facts $\|M\| = \|M'\| = 1$ and $\tau_1, \tau_2 \geq (m+n)^{-1/2}$. This holds over all $j, k \in [n]$, so it extends to the infinity norm, proving Eq. (28).

Let us prove Eq. (29). Consider an arbitrary $j \in [n]$. By Eq. (64), $(\tilde{V}_S \tilde{V}_S^T - V_S V_S^T)_{j,\cdot} = M^T \mathcal{T} M'$ for the choices $M = e_{m+n,j+m}$ and $M' = I_{m+n}$. We repeat the previous calculations, but this time Eq. (70) tells us to replace $\tau(M)$ with τ_2 and $\tau(M')$ with 1, to get

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

which holds uniformly over $j \in [n]$, proving Eq. (29). $\square$

Next, we will finish proving Theorem B.2 by proving Eq. (30). 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. (56) 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

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

for every contour Γ encircling σ. Therefore, the analogy of Eq. (57) is

$$\begin{aligned}(\tilde{A}_s - A_s)_{\text{sym}} &= (\tilde{A}_{\text{sym}})_s - (A_{\text{sym}})_s = \sum_{|i|\in[s]} \left(\frac{z\tilde{w}_i\tilde{w}_i^T}{z-\tilde{\sigma}_i} - \frac{zw_iw_i^T}{z-\sigma_i} \right) \\ &= \sum_{\gamma=1}^{\infty} \oint_{\Gamma_s} \frac{z\mathrm{d}z}{2\pi i} \left[(zI - A_{\text{sym}})^{-1} E_{\text{sym}} \right]^{\gamma} (zI - A_{\text{sym}})^{-1}. \\ &= \sum_{\gamma=1}^{\infty} \oint_{\Gamma_s} \frac{z\mathrm{d}z}{2\pi i} \Big[\Big(\sum_{|i|\in[r]} \frac{\sigma_i P_i}{z(z-\sigma_i)} + \frac{I}{z} \Big) E_{\text{sym}} \Big]^{\gamma} \Big(\sum_{|i|\in[r]} \frac{\sigma_i P_i}{z(z-\sigma_i)} + \frac{I}{z} \Big).\end{aligned} \tag{71}$$

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

$$\mathcal{C}_1(\mathbf{I}) := \oint_{\Gamma_S} \frac{z\mathrm{d}z}{2\pi i} \frac{1}{z^{\gamma+1}} \prod_{j\in[\beta]} \frac{\sigma_{i_j}}{z-\sigma_{i_j}}. \tag{72}$$

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

The only piece we need to modify in the proofs of Eqs. (28) and (29) 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.* (72)*. We have,*

$$|\mathcal{C}_1(\mathbf{I})| \le \frac{2^{\gamma+\beta-2}}{\sigma_s(\mathbf{I})^{\gamma-\beta}\Delta_s(\mathbf{I})^{\beta-1}} \le \frac{2^{\gamma+\beta-2}}{\sigma_s^{\gamma-\beta}\Delta_s^{\beta-1}} = \frac{\sigma_s}{2} \cdot \frac{2^{\gamma+\beta-1}}{\sigma_s^{\gamma+1-\beta}\Delta_s^{\beta-1}}. \tag{73}$$

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. (72) is larger than 1, or when S does not contain exactly the first s singular values. Therefore, one can neither extend Eq. (30) 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. (30). Fix $j \in [m]$ and $k \in [n]$. By Eq. (71), $(\tilde{A}_s - A_s)_{jk} = M^T\mathcal{T}_1M'$ for $M = e_{m+n,j}$ and $M' = e_{m+n,m+k}$. The bound on $M^T\mathcal{T}_1M'$ will simply be the same bound for $M^T\mathcal{T}M'$ scaled up by $\sigma_s/2$. By Eq. (70), we can also replace $\tau(M)$ with τ_1 and $\tau(M')$ with τ_2. Therefore we obtain

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

where the last inequality holds due to $\tau_1, \tau_2 \ge (m+n)^{-1/2}$ and $\|M\| = \|M'\| = 1$. After passing to the inifinity norm, the proofs of Eq. (30) 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.

C.3 Bounding the generic series

Let us prove Lemma C.4.

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

$$\mathcal{M}(\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{I}) := E_{\text{sym}}^{\alpha_0}\Big(\prod_{j=1}^{\beta_1} P_{i_j}E_{\text{sym}}\Big)E_{\text{sym}}^{\alpha_1}\dots\Big(\prod_{j=\beta-\beta_h}^{\beta-1} P_{i_j}E_{\text{sym}}\Big)P_{i_{\beta-1}}E_{\text{sym}}^{\alpha_h}. \tag{74}$$

From Eq. (61), we can rearrange this to get

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

Let us break down this product into the following types:

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

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

We continue with proving Lemma C.5.

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

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

we obtain

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

Applying the second part of the bound (65) on $\mathcal{C}(\mathbf{I})$ in Lemma C.3, we get

$$\begin{aligned}
\|M^T\mathcal{T}(\boldsymbol{\alpha},\boldsymbol{\beta})M'\| &\le \|X\|_\infty^{\beta-h}\|E\|^{\alpha-\alpha_0-\alpha_h+h-1}\frac{2^{\gamma+\beta-1}}{\sigma_S^\alpha\Delta_S^{\beta-1}} \sum_{\mathbf{I}\in[\pm r]^\beta} \left\|M^T E_{\text{sym}}^{\alpha_0} w_{i_1}\right\| \left\|w_{i_{h\beta_h}}^T E_{\text{sym}}^{\alpha_h} M'\right\| \\
&= \|X\|_\infty^{\beta-h}\|E\|^{\alpha-\alpha_0-\alpha_h+h-1}\frac{2^{\gamma+\beta-1}(2r)^{\beta-2}}{\sigma_S^\alpha\Delta_S^{\beta-1}} \sum_{i\in[\pm r]} \left\|w_i^T E_{\text{sym}}^{\alpha_0} M\right\| \sum_{i\in[\pm r]} \left\|w_i^T E_{\text{sym}}^{\alpha_h} M'\right\| \\
&= \|X\|_\infty^{\beta-h}\|E\|^{\alpha+h-1}\frac{2^{\gamma+\beta-1}(2r)^{\beta}}{\sigma_S^\alpha\Delta_S^{\beta-1}} \sum_{i\in[\pm r]} \frac{\left\|w_i^T E_{\text{sym}}^{\alpha_0} M\right\|}{2r\|E\|^{\alpha_0}} \sum_{i\in[\pm r]} \frac{\left\|w_i^T E_{\text{sym}}^{\alpha_h} M'\right\|}{2r\|E\|^{\alpha_h}} \\
&\le \tau(M)\tau(M')\|X\|_\infty^{\beta-h}\|E\|^{\alpha+h-1}\frac{2^{\gamma+\beta-1}(2r)^{\beta}}{\sigma_S^\alpha\Delta_S^{\beta-1}}.
\end{aligned} \tag{75}$$

After rearrangements, we get

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

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

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

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

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

We get the first case of Eq. (69). Now consider the case $h = \gamma/2 + 1$, which only happens when γ is even. The previous bound becomes

$$\|M^T \mathcal{T}(\boldsymbol{\alpha}, \boldsymbol{\beta}) M'\| \leq r\tau(M)\tau(M')2^{\gamma+\beta-1}R_2^{2(h-1)} = r\tau(M)\tau(M')2^{\beta-1}(2R_2)^{\gamma}. \tag{76}$$

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 $\alpha_1 = \ldots = \alpha_{h-1} = \beta_1 = \ldots = \beta_h = 1$, thus $\beta = h$. Let $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ denote the corresponding tuple. Plugging into Eq. (74) and simplifying, we have

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

Consider the long product at the end of the right-hand side. For the purpose of this proof, let $y := \max_{|i| \neq |j|} |w_i^T E_{\text{sym}}^2 w_j|$ (the term in R_3's definition). Note that this is smaller than the term y in Theorem 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 $\left|w_{i_k}^T E_{\text{sym}}^2 w_{i_{k+1}}\right| \leq y$ and we are good. The rest of the product can be bounded by $\|E\|^2$. The total contribution of this subcase is at most

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

since we can simply replace a factor of R_2^2 in Eq. (76) 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

$$r\tau(M)\tau(M') \left[2^{3\gamma/2}R_3 R_2^{\gamma-2} + 2(4R_1)^{\gamma}\right] \leq 16r\tau(M)\tau(M')(4R)^{\gamma-2}\left(R_3 + R_1^2\right).$$

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. (75) again. We cannot use $\tau(M)$ and $\tau(M')$ anymore, but we can use the trivial upper bounds

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

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

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 $(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

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

$$M^T\mathcal{T}^{(\gamma)}M' = \sum_{h=1}^{\lfloor\gamma/2\rfloor+1}\sum_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\Pi_h(\gamma)} M^T\mathcal{T}(\boldsymbol{\alpha},\boldsymbol{\beta})M'. \tag{77}$$

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

$$\begin{aligned}\left\|M^T\mathcal{T}^{(\gamma)}M'\right\| &\le \sum_{h=1}^{\lfloor\gamma/2\rfloor+1}\sum_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\Pi_h(\gamma)} r\tau(M)\tau(M')2^{\beta}R_1(2R)^{\gamma-1}\\ &= r\tau(M)\tau(M')R_1(2R)^{\gamma-1}\sum_{h=1}^{\lfloor\gamma/2\rfloor+1}\sum_{\beta=h}^{\gamma+2-h}2^{\beta}\left|\{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\Pi_h(\gamma):\textstyle\sum_j\beta_j=\beta\}\right|\end{aligned} \tag{78}$$

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

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

The number of ways to choose such a tuple is $\binom{\beta-1}{h-1}\binom{\gamma+2-\beta}{h}$. Plugging into Eq. (78), we obtain

$$\begin{aligned}\left\|M^T\mathcal{T}_\nu^{(\gamma)}M'\right\| &\le r\tau(M)\tau(M')R_1(2R)^{\gamma-1}\sum_{h=1}^{\lfloor\gamma/2\rfloor+1}\sum_{\beta=h}^{\gamma+2-h}\binom{\beta-1}{h-1}\binom{\gamma+2-\beta}{h}2^{\beta}\\ &= r\tau(M)\tau(M')R_1(2R)^{\gamma-1}\sum_{\beta=1}^{\gamma+1}2^{\beta}\sum_{h=1}^{\beta\wedge(\gamma+2-\beta)}\binom{\beta-1}{h-1}\binom{\gamma+2-\beta}{h}\\ &\le r\tau(M)\tau(M')R_1(2R)^{\gamma-1}\sum_{\beta=1}^{\gamma+1}2^{\beta}\binom{\gamma+1}{\beta} = 9r\tau(M)\tau(M')R_1(6R)^{\gamma-1}.\end{aligned} \tag{79}$$

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

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

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

Now we finish the bound on the entire power series.

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

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

and

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

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

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

The proof is complete. □

D Proofs of technical lemmas

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

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

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

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

$$\mathcal{C}_\nu(\mathbf{I}) := \oint_{\Gamma_S} \frac{z^\nu \mathrm{d}z}{2\pi i} \frac{1}{z^{\gamma+1}} \prod_{k=1}^{\beta} \frac{\sigma_{i_k}}{z - \sigma_{i_k}}, \quad \text{where } \nu \in \{0, 1\} \text{ and } \beta \leq \gamma + 1. \tag{80}$$

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

$$\mathcal{C}_\nu(\mathbf{I}) = \prod_{i=1}^{l} a_i^{m_i} \prod_{j=1}^{k} b_j^{n_j} C(n_0; A, \mathbf{m}; B, \mathbf{n}), \tag{81}$$

where

$$C(n_0; A, \mathbf{m}; B, \mathbf{n}) := \oint_{\Gamma_A} \frac{\mathrm{d}z}{2\pi i} \frac{1}{z^{n_0}} \prod_{j=1}^{k} \frac{1}{(z - b_j)^{n_j}} \prod_{i=1}^{l} \frac{1}{(z - a_i)^{m_i}}, \tag{82}$$

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 S and simply denote the contour by Γ_A without affecting its meaning. The next three results will build up the argument to bound these sums and ultimately prove the target lemmas.

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

$$d \leq a, \qquad a \leq \min_i |a_i|, \qquad d \leq \min_{i,j} |a_i - b_j|. \tag{83}$$

Suppose that $0 \leq m_i' \leq m_i$ for each $i \in [l]$ and that $m' := \sum_{i=1}^{k} m_i' \leq n_0$. Then for $C(n_0; A, \mathbf{m}; B, \mathbf{n})$ defined Eq. (82)*, we have*

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

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

$$|C(n_0; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m + n + n_0 - 2}{m - 1} \frac{1}{d^{m+n+n_0-1}}. \tag{85}$$

We omit the details of the proof, which is a simple induction argument. We now use Eq. (85) to prove the following:

$$|C(n_0; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m + n + n_0 - 2}{m - 1} \frac{1}{a^{n_0} d^{m+n-1}}. \tag{86}$$

We proceed with induction. Let $P_1(N)$ be the following statement: "For any sets A and B, and the notations and conditions described in Lemma D.1, such that $m + n + n_0 = N$, Eq. (86) holds."

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

Consider $n \geq 3$ and assume $P_1(n-1)$. If $m = 0$, the integral is again 0. If $n_0 = 0$, Eq. (86) automatically holds by being the same as Eq. (85). Assume $m, n_0 \geq 1$. There must then be some $i \in [l]$ such that $m_i \geq 1$, without loss of generality let 1 be that i. We have

$$C(n_0; A, \mathbf{m}; B, \mathbf{n}) = \frac{1}{a_1}\Big[C(n_0 - 1; A, \mathbf{m}; B, \mathbf{n}) - C(n_0; A, \mathbf{m}^{(1)}; B, \mathbf{n})\Big] \tag{87}$$

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

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

$$|C(n_0 - 1; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m+n+n_0-3}{m-1} \frac{1}{a^{n_0-1} d^{m+n-1}}. \tag{88}$$

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

$$\left|C(n_0; A, \mathbf{m}^{(1)}; B, \mathbf{n})\right| \leq \binom{m+n+n_0-3}{m-2} \frac{1}{a^{n_0} d^{m+n-2}} \leq \binom{m+n+n_0-3}{m-2} \frac{1}{a^{n_0-1} d^{m+n-1}}. \tag{89}$$

Notice that the binomial coefficients in Eqs. (88) and (89) sum to the binomial coefficient in Eq. (86), we get $P_1(N)$, which proves Eq. (86) by induction.

Now we can prove Eq. (84). The logic is almost identical, with Eq. (86) playing the role of Eq. (85) in its own proof, handling an edge case in the inductive step. Let $P_2(n)$ be the statement: "For any sets A and B, and the notations and conditions described in Lemma D.1, such that $m + n + n_0 = N$, Eq. (84) holds."

The cases $N = 1$ and $N = 2$ are again trivially true. Consider $N \geq 3$ and assume $P_2(N-1)$. Fix any sequence $m'_1, m'_2, \ldots, m'_l$ satisfying $0 \leq m'_i \leq m_i$ for each $i \in [k]$ and $n_0 \geq m'_1 + \ldots + m'_k$. If $m'_1 = m'_2 = \ldots = m'_k = 0$, we are done by Eq. (86). By symmetry among the indices, assume $m'_1 \geq 1$. This also means $n_0 \geq 1$. Consider Eq. (87) again. For the first integral on the right-hand side, applying $P_2(N-1)$ for the parameters $n_0 - 1, n_1, \ldots, n_k, m_1, \ldots, m_l$ and $m'_1 - 1, m'_2, \ldots, m'_k$ yields the bound

$$|C(n_0 - 1; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{m+n+n_0-3}{m-1} \frac{1}{a^{n_0-m'} d^{m+n-1}} \frac{1}{|a_1|^{m'_1-1}} \prod_{i=2}^{l} \frac{1}{|a_i|^{m'_i}}. \tag{90}$$

Applying $P_2(N-1)$ for the parameters $n_0, n_1, \ldots, n_k, m_1 - 1, \ldots, m_l$ and $m'_1 - 1, m'_2, \ldots, m'_k$, we get the following bound for the second integral on the right-hand side of Eq. (87):

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

Summing up the bounds by summing the binomial coefficients, we get exactly $P_2(N)$, so Eq. (84) is proven by induction. □

Lemma D.2. *Let A, B, $\mathbf{m}$, $\mathbf{n}$, n_0, Γ_A and a, d be the same, with the same conditions as in Lemma 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*

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

Then for $C(n_0; A, \mathbf{m}; B, \mathbf{n})$ defined in Eq. (82), we have

$$|C(n_0; A, \mathbf{m}; B, \mathbf{n})| \leq \binom{n+n_0-n'+m-2}{m-1} \frac{(1+d/a)^{n'}}{a^{n_0-m'-n'} d^{m+n-1}} \prod_{i=1}^{l} \frac{1}{|a_i|^{m'_i}} \prod_{j=1}^{k} \frac{1}{|b_j|^{n'_j}}. \tag{91}$$

Proof. We have the expansion

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

Integrating both sides over Γ_A, we have

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

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

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

Summing up the bounds, we get

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

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

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

Proof of Lemmas C.3 and C.8. First rewrite the integral into the forms of (80), then (81) and (82). Let us consider two cases for $\mathcal{C}$:

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

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

   or equivalently,

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

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

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

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

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

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

which translates to

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

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

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

Since $\gamma \geq 1$, we have

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

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 $\sigma_t = \sigma_S(\mathbf{I})$, then $\Delta_S(\mathbf{I}) \geq \sigma_t - \sigma_{s+1}$, thus

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

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

$\square$

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

In this section, we prove TheoremB.4, which gives semi-itrosopic bounds for powers of E_{sym} in the second step of the main proof strategy.

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

Lemma D.3. *Let $m, r \in \mathbb{N}$ and $U \in \mathbb{R}^{m \times r}$ be a matrix whose columns $u_1, u_2, \ldots, u_r$ are unit vectors. Let E be a $m \times n$ random matrix following Model* (31) *with parameters M and $\varsigma = 1$, meaning E has independent entries and*

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

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

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

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

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

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

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

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

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

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

Proof of Theorem B.4. Consider the analogue of Eq. (33) for V (we wrote the proof for V before the final edit, and wanted to save the energy of changing to U) and Eq. (34), and assume $M \leq \log^{-2-\varepsilon}(m+n)\sqrt{m+n}$. Fix $k \in [n]$. It suffices to prove the following two bounds uniformly over all $a \in [\lfloor t\log(m+n)\rfloor]$:

$$\left\|e_{n,k}^T(E^TE)^aE^TU\right\| \leq C\tau_1(U, \log\log(m+n))(1.9\varsigma\sqrt{m+n})^{2a+1}\sqrt{r} \tag{92}$$

$$\left\|e_{n,k}^T(E^TE)^aV\right\| \leq C\tau_0(V, \log\log(m+n))(1.9\varsigma\sqrt{m+n})^{2a}\sqrt{r}. \tag{93}$$

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

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

so we can apply both Lemmas D.3 and D.4.

Let us prove Eq. (92) for a. Applying Lemma D.3 for the random matrix E/ς and $D = 2^{13}$ gives, with probability $1 - \log^{-4.04}(m+n)$,

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

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. (92), and thus Eq. (34) follows.

Let us prove Eq. (93). Applying Lemma D.3 for the random matrix E/ς and $D = 2^{10}$ gives, with probability $1 - \log^{-8}(m+n)$,

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

proving Eq. (93) and thus Eq. (33) after a union bound, similar to the previous case.

Let us now prove Eqs (35) and (36), focusing on the former first. Since the 2-to-∞ norm is the the largest norm among the rows, it suffices to prove Eq. (33) 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. (33), but with probability at least $1-(m+n)^{-4.04}$. Applying another union bound over $k \leq [n]$ gives Eq. (35) with probability at least $1-(m+n)^{-3}$. The proof of (36) is analogous. The proof of Theorem B.4 is complete. ☐

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

$$\begin{aligned}\mathbf{E}\left[\left\|e_{n,1}^T(E^TE)^aE^TU\right\|^{2p}\right] &= \mathbf{E}\left[\Big(\sum_{l=1}^r(e_{n,1}^T(E^TE)^aE^Tu_l)^2\Big)^p\right] \\ &= \sum_{l_1,\dots,l_p\in[r]}\mathbf{E}\left[\prod_{h=1}^p(e_{n,1}^T(E^TE)^aE^Tu_{l_h})^2\right].\end{aligned} \tag{94}$$

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

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

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

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

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

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

Define the following notation:

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

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

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

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

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

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

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

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

We can rewrite the previous sum as:

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

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

$$\sum_{P \in \mathcal{P}} |\mathbf{E}\left[E_P\right]| \prod_{h=1}^{p} \|U_{\cdot, i_{(2h-1)a}}\| \|U_{\cdot, i_{(2h)a}}\| = r^p \sum_{P \in \mathcal{P}} u_{P^{\text{end}}} |\mathbf{E}\left[E_P\right]| = r^p \sum_{S \in \mathcal{S}} \sum_{P \in \mathcal{P}(S)} u_{P^{\text{end}}} |\mathbf{E}\left[E_P\right]|, \tag{95}$$

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

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

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) is traversed in P. We have

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

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

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

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

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

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

where, we define $V_1(Q)$ as the set of vertices appearing in Q exactly once and $V_2(Q)$ as the set of vertices appearing at least twice, and to shorten the notation, we let $\nu(S) := |V_1(S^{\text{end}})| + 2|V_2(S^{\text{end}})|$. Combining the bounds, we get the upper bound below for (95):

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

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

$$M_1 \sum_{x,y,z,t\in\mathcal{A}} M^{-2(z+t)} m^{z-x/2-y} n^{t-1} \|u\|_\infty^{-x-2y} |\mathcal{S}(x,y,z,t)|, \tag{96}$$

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

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

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

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

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

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

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

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

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

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

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

$$\mathcal{A} = \left\{(x, y, z, t) \in \mathbb{N}^4_{\geq 0} : \ 1 \leq t;\ 1 \leq x + y \leq z;\ x + 2y \leq 2p;\ z - x/2 - y + t - 1 \leq 2ap\right\}. \tag{97}$$

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

Claim D.6. *Given a tuple $(x, y, z, t) \in \mathcal{A}$, where $\mathcal{A}$ is defined in Eq.* (97)*, we have*

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

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

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

This scheme only creates a *preliminary codeword* W, which does not yet uniquely determine the original S. To be able to trace back S, we need the scheme in [34] to add more details to the preliminary codewords. For completeness, we will describe this scheme later, but let us first bound the number of preliminary codewords.

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

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

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

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

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

Combining the bounds above, we have

$$|\mathcal{S}(x,y,z,t)| \leq \binom{2p(a+1)}{z}\binom{2pa}{t-1}\sum_{u+v=z+t-1}\binom{2p(a+1)-z}{u}\binom{2pa-t+1}{v}z^{f(z,u)}t^{g(t,v)}, \tag{98}$$

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

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

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

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

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

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

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

Back to the proof of Claim D.6, to uniquely determine the shape S, the general idea is the following. We first generated the preliminary codeword W from S, then attempt to decode it. If we encounter a plus or neutral edge, we immediately know the next vertex. If we see a minus edge that follows from a plus edge (u,v), we know that the next vertex is again u. Similarly, if there are chunks of the form $(++\ldots+--\ldots-)$ with the same number of each sign, the vertices are uniquely determined from the first vertex. Therefore, we can create a condensed codeword W^* repeatedly removing consecutive pairs of $(+-)$ until none remains. For example, the sections $(-+-+-)$ and $(-++--)$ both become $(-)$. Observe that the condensed codeword is always unique regardless of the order of removal, and has the form

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

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

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

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

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

It has been shown in [34] that the fully codeword $\overline{W}$ resulting from W by marking important edges uniquely determines S, and that when the shape of S *is that of a single walk*, the cost of these markings is at most a multiplicative factor of $2(4N+8)^N$, where N is the number of neutral edges in the preliminary W. To adapt this bound to our case, we treat the star shape S as a single walk, with a neutral edge marked by 1 after every $2a+1$ edges. There are $2p-1$ additional neutral edges from this perspective, making $N = 4p(a+1) - 2l - 1$ in total. Combining this with the bound on the number of preliminary codewords (Claim D.7) yields

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

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

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

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

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

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

Replacing the appropriate terms in the bound in Claim D.6 with these short forms, we get another series of upper bounds for the last double sum in Eq. (95):

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

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

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

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

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

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

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

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

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

For each d, there are at most $2(p-d)$ pairs (x, y) such that $d = p - (\lfloor \frac{x}{2} \rfloor + y)$, so overall we have the following series of upper bounds for the last double sum in Eq. (95):

$$\begin{aligned} & M_2(m+n)^{2pa} \sum_{d=0}^{p-1} 4(p-d) \|u\|_\infty^{-2(p-d)} \cdot \frac{(2M^{-2})^{N-d+1}(N-d+1)^{2d}(8p+8d-2)^{2p+2d-1}}{(2pa)! \cdot (2d)! \cdot (N-d)!} \\ & \leq M_3(m+n)^{2pa} \sum_{d=0}^{p-1} \|u\|_\infty^{2d} \cdot \frac{2^{-d} M^{2d}(N-d+1)^{2d}(8p+8d-2)^{2p+2d-1}}{(2d)! \cdot (N-d)!}, \end{aligned} \tag{99}$$

where

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

Let us bound the sum at the end of Eq. (99). Temporarily let A_d be the term corresponding to d and $x := 2^{-1/2} M \|u\|_\infty$. We have

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

Therefore

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

Plugging this into Eq. (99), we get another upper bound for (95):

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

where

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

To sum up, we have

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

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 moment bound above applies, so we have

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

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

D.2.2 Case 2: even powers

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where

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

We have

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

where $M_2 := M_1 [(2pa)!]^2 M^{-2} = M^{4ap} [(2pa)!]^2 \|v\|_\infty^{2p}$. Let C_l be the term corresponding to l in the last sum above. An analogous calculation from the proof of Lemma D.3 shows that under the assumption that $m + n \ge 256 M^2 p^6 (2a)^4$, $C_l \ge 2C_{l-1}$ for each l, so $\sum_{l=0}^{2pa} C_l \le 2C_{2pa}$, where

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

Therefore

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

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

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

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