Approximating Permanent of Random Matrices with Vanishing Mean: Made Better and Simpler*

Zhengfeng Ji† Zhihan Jin Pinyan Lu‡

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Abstract

The algorithm and complexity of approximating the permanent of a matrix is an extensively studied topic. Recently, its connection with quantum supremacy and more specifically BosonSampling draws a special attention to the average-case approximation problem of the permanent of random matrices with zero or small mean value for each entry. Eldar and Mehraban (FOCS 2018) gave a quasi-polynomial time algorithm for random matrices with mean at least $1/\operatorname{polyloglog}(n)$. In this paper, we improve the result by designing a deterministic quasi-polynomial time algorithm and a PTAS for random matrices whose module of mean is at least $1/\operatorname{polylog}(n)$. We note that if the algorithm can be further improved to work with a mean value that is a sufficiently small $1/\operatorname{poly}(n)$, it will disprove a central conjecture for quantum supremacy.

Our algorithm is also much simpler and has a better and flexible trade-off for running time. The running time can be quasi-polynomial in both n and $1/\varepsilon$, or PTAS (polynomial in n but exponential in $1/\varepsilon$), where ε is the approximation parameter.

1 Introduction

The computational complexity of computing the permanent of a matrix is of central importance to complexity theory and has been extensively studied ever since Valiant's seminal result [Val79b]. On one hand, the problem is algebraic in nature and plays an important role in the study of algebraic complexity [Val79a, BCS13]. In particular, its relation with the determinant is an important topic [MR04, CCL10]. On the other hand, it also exhibits rich combinatorial properties. The permanent can be viewed as counting the (weighted) number of perfect matchings of a bipartite graph and graph (perfect) matching is one of the most important graph problems in the study of algorithm and complexity [Edm65, Val79b, Val08].

Since the exact computation of the permanent is already #P-hard for matrices with non-negative entries, or even 0/1 entries [Val79b], more recent research focuses on either the *approximation* of the permanent or the *average-case* complexity of the problem where the matrices are sampled from certain distributions.

In the approximation approach, we require the algorithm to return a number Z' such that, if the actual value of the permanent of the input matrix is Z, the computed number Z' satisfies $|Z - Z'| \le \varepsilon |Z|$ within running time $\operatorname{poly}(n, \frac{1}{\varepsilon})$ where $\varepsilon > 0$ is the approximation parameter. This is called a fully polynomial-time approximation scheme (FPTAS). And its randomized relaxation is called a fully polynomial-time randomized approximation scheme (FPRAS), where we require that $|Z - Z'| \le \varepsilon |Z|$ holds with high probability. If the running time is quasi-polynomial in terms of n and $\frac{1}{\varepsilon}$, namely $2^{\operatorname{poly}(\log(n), \log \frac{1}{\varepsilon})}$, then it is called a quasi-polynomial time approximation scheme. If we only require the running time to be polynomial in n but not necessary in $\frac{1}{\varepsilon}$, we call it polynomial-time approximation scheme (PTAS). On the other hand, in the average-case approach, we allow the algorithm to be incorrect on a small fraction of instances with respect to some distributions over matrices. Usually, this distribution is over matrices with i.i.d. random entries and the algorithm is required to output either the exact value or an approximation of the permanent on at least $1 - o(1)$ fraction of the instances.

In fact, several worst-case approximation tractability and hardness results were known. For a matrix with non-negative entries, Jerrum, Sinclair and Vigoda gave a remarkable FPRAS to approximate its permanent [JSV04] via random sampling by Markov chain Monte Carlo (MCMC). How to derandomize this algorithm remains a long-standing open problem. However, it is impossible to extend this result to general matrices since it is already #P-hard to compute the sign of the permanent with possibly negative entries. Indeed, negative or complex values put this problem in GapP, a superset of #P [FFK94]. This difficulty is referred to as the "interference barrier". For example, random sampling based algorithms are no longer applicable since we cannot define negative or complex probability. For

*This work was done in partial when the authors were visiting ITCS, Shanghai University of Finance and Economics.

†Zhengfeng Ji is supported by Australian Research Council Discovery Project, Grant No. DP200100950.

‡Pinyan Lu is supported by Science and Technology Innovation 2030 – "New Generation of Artificial Intelligence" Major Project No.(2018AAA0100903), NSFC grant 61922052 and 61932002, Innovation Program of Shanghai Municipal Education Commission, Program for Innovative Research Team of Shanghai University of Finance and Economics, and the Fundamental Research Funds for the Central Universities.

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

specific families of complex matrices, there are quasi-polynomial time approximation schemes by Barvinok's interpolation [Bar16, Bar19].

The above algorithms and hardness results are all worst-case analysis. What do we know about the average-case complexity? It turns out that, for exact counting, the average-case problem remains #P-hard both for finite field entries and complex Gaussian entries [CPS99, AA13]. This leaves the complexity of the average-case approximation of the permanent an intriguing problem. Yet, very little was known when both average-case analysis and approximation are considered at the same time. More generally, while we have #P-hardness results for all other settings including worst-case approximation problems and average-case exact problems, essentially no hardness result is known for average-case approximate counting problems. On the tractability side, some recent algorithms and techniques show that random instances might be much easier than the worst-case for approximate counting [JKP19, LLLM19, MB19, EM18]. Our result also adds to this list.

An important motivation for studying the complexity of approximating the permanent of random matrices stems from the so-called BosonSampling program initiated by Aaronson and Arkhipov [AA13] in quantum computing. In [AA13], the conjectured #P-hardness of approximating the permanent of Gaussian matrices (Permanent-of-Gaussian's Conjecture) is connected to the sampling problem of linear optical networks so that the existence of any efficient classical simulation of this optical sampling process will imply $\mathsf{P}^{\#\mathsf{P}} = \mathsf{BPP}^{\mathsf{NP}}$, and hence the collapse of the polynomial hierarchy by Toda's theorem. This provides an explicitly defined problem which near-term quantum computing devices can efficiently solve while even today's most powerful classical supercomputer cannot. Such a dramatic contrast in computing powers, called quantum supremacy, poses the first serious challenge to the extended Church-Turing thesis and has been recently experimentally achieved by the Google team using a different model based on the random quantum circuit sampling problem [AAB+19] while the record of BosonSampling experiment is recently updated by [WQD+19].

The complexity of approximating the permanent of random matrices is obviously of vital importance to BosonSampling as it serves as one of the two conjectures on which the theory of BosonSampling bases. In particular, it is assumed in BosonSampling that approximating the permanent of random matrices whose entries are i.i.d. sampled from the normal distribution of zero mean value and unit variance is #P-hard. The result is strengthened in [EM18] showing that a biased distribution with mean at most $1/n^c$ for some large c is also good enough for BosonSampling. There is no clue yet on how one can prove such hardness results and it is not even clear whether they are true or not. On the other side, a surprising and interesting tractable result was obtained by Eldar and Mehraban [EM18]. They provided a quasi-polynomial time algorithm to approximate the permanent of random matrices with mean of $1/\operatorname{polyloglog}(n)$, which implies that the #P-hardness is unlikely to hold for these families of random matrices. This raises the interesting open question of whether the algorithm can be extended to the case with mean value $1/\operatorname{poly}(n)$ and disprove the hardness conjecture of BosonSampling or there is a "phase transition" in the complexity of approximating the permanent with respect to the mean of matrix entries.

1.1 Our results In this paper, we provide an exponential improvement in terms of the tractable region of the mean values to the problem of approximating the permanent of random matrices with vanishing mean value. We design a deterministic quasi-polynomial time algorithm and a PTAS that can compute the multiplicative approximation for $1 - o(1)$ fraction of random matrices whose module of mean is at least $1/\operatorname{polylog}(n)$. See Theorem 3.1, Corollary 3.1, Corollary 3.2 for more rigorous statements of our results. The strength of our results lies in the following four aspects.

Firstly and most importantly, the range of the tractable mean value parameters is exponentially better than that of [EM18]. In [EM18], their algorithm can only approximate the permanent of a random matrix with mean value of at least $1/\operatorname{polyloglog}(n)$. Our algorithm works for all mean value whose module is at least $1/\operatorname{polylog}(n)$. The exact range of mean values for which such approximation exists is extremely important due to its role in the "quantum supremacy". If one can further improve the mean to some $1/\operatorname{poly}(n)$ that is sufficiently small, it will disprove the conjecture in [AA13].

Secondly, the algorithm in [EM18] only works for some, but not all, mean values $\mu > 1/\operatorname{polyloglog}(n)$. This is a very strange situation due to their proof techniques and is rather counterintuitive as one would expect that the larger the mean value is, the easier it is to approximate the permanent. There is not even an algorithm for them to check whether a given mean is computable or not for their algorithm since they used a probabilistic argument while our algorithm does not suffer from such problems and works for all complex μ with $|\mu| > 1/\operatorname{polylog}(n)$.

Thirdly, our algorithm uses a completely different idea and is arguably simpler. The simplicity of our al-

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

gorithm also enables us to extend our result to a much larger family of entry distributions. While the technique of [EM18] is very interesting and uses Barvinok's interpolation method for approximate counting with several new developments of the technique in a few directions, the need to prove the zero-freeness of the polynomial in a segment-like region made the proof rather involved and caused the drawbacks of their result mentioned above in the previous two items. In our algorithm, we avoid all these complications and simply truncate a simple expansion of the permanent directly to $\mathcal{O}(\ln n + \ln \frac{1}{\epsilon})$ terms and compute them by brute-force. We note that in usual usage of Barvinok's method, a truncation of the polynomial directly rather than its logarithm will not succeed. To prove the correctness of our algorithm, we need a careful study of the distribution of the permanent of random matrices. To this end, we use several techniques inspired by the analysis of [RW04]. The result and analysis in [RW04] are asymptotic while we need much more careful quantitative bounds which we develop carefully in this paper.

Lastly, the running time of our algorithm is also better and flexible. It can be quasi-polynomial in both n and $1/\varepsilon$, where ε is the approximation parameter. We can also make it a PTAS, which is polynomial in n but exponential in $1/\varepsilon$. In particular, when $\varepsilon > n^{-\rho}$ for some fixed and universal constant $\rho > 0$, the algorithm is extremely simple and runs in only linear time. It is not clear how to make the previous algorithm in polynomial time rather than quasi-polynomial time even for a fixed constant ε.

This work leaves several interesting open problems. First, the most important problem left open is to either show the transition of complexity with respect to the mean value and prove that the corresponding problem is hard when the mean value is $1/\operatorname{poly}(n)$ or disprove the Permanent-of-Gaussian conjecture of BosonSampling. With our technique only, it is rather hard to go beyond the $1/\operatorname{polylog}(n)$ barrier and essential new ideas seem necessary if this is ever possible. Second, while we have been focusing exclusively on the problem of approximating the permanent and therefore it is only directly relevant to the BosonSampling scheme of quantum supremacy, we expect that our technique may find applications in understanding other average-case approximate counting problems and the hardness assumptions in other quantum supremacy schemes such as the instantaneous quantum computing model [BJS10] and the random circuit sampling model [AAB+19]. We believe that such generalizations are possible as the hardness conjectures behind different models of quantum supremacy are of the same flavor and it is usually possible to generalize results from one model to the other (see e.g. [BFNV18]).

The rest of the paper is organized as follows. In Section 2, we introduce the notations used in this paper. We state and present the proof outline in Section 3. The remaining sections contain the technical lemmas used in Section 3.

2 Preliminary

In this paper we use $[n]$ to denote the set $\{1, \cdots, n\}$. The set of natural numbers, real numbers, and complex numbers are denoted as $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$ respectively. We use $n^{\underline{k}} \triangleq n(n-1)\cdots(n-k+1)$ to denote the downward factorial and $C_{n,k}, P_{n,k}$ to denote all k-subsets of $[n]$ and all k-permutations of $[n]$ respectively. $\mathrm{M}_n(\mathbb{C})$ to denote the set of all $n \times n$ complex matrices. $\delta_{i,j}$ is the Kronecker function, i.e., $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ otherwise.

Definition 2.1. *Suppose $x_1, x_2, \cdots, x_n \in \mathbb{C}$ and $0 \le k \le n$, the power sum is defined by*

$$S_k(x_1, x_2, \cdots, x_n) \triangleq \sum_{i=1}^{n} x_i^k$$

and the kth elementary symmetric polynomial is defined by

$$e_k(x_1, x_2, \cdots, x_n) \triangleq \sum_{\{i_1, \cdots, i_k\} \in C_{n,k}} x_{i_1} \cdots x_{i_k}$$

with convention $e_0(x_1, x_2, \cdots, x_n) = 1$. We will write $S_k(n)$ and $e_k(n)$ if the variables x_i's are clear from context.

Definition 2.2. *The entry distribution $\mathcal{D}_\mu$ with mean value $\mu \in \mathbb{C}$ is a distribution over complex numbers such that*

$$\mathbb{E}_{x \sim \mathcal{D}_\mu}[x] = \mu, \quad \mathrm{Var}_{x \sim \mathcal{D}_\mu}[x] = 1,$$

and

$$\mathbb{E}_{x \sim \mathcal{D}_\mu} |x - \mu|^3 = \rho < \infty.$$

We use $\mathcal{D}$ to denote $\mathcal{D}_0$.

In this paper, we use ξ to denote the quasi-variance of $\mathcal{D}_\mu$,

$$\xi = \mathbb{E}_{x \sim \mathcal{D}_\mu}(x - \mu)^2.$$

The norm of the quasi-variance is upper bounded by the variance as

$$\begin{aligned} |\xi| &= \left|\mathbb{E}_{x \sim \mathcal{D}_\mu}(x - \mu)^2\right| \\ &\le \mathbb{E}_{x \sim \mathcal{D}_\mu} |x - \mu|^2 \\ &= \mathrm{Var}_{x \sim \mathcal{D}_\mu}(x) = 1. \end{aligned} \tag{2.1}$$

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

DEFINITION 2.3. *The matrix distribution $\mathcal{M}_{n,\mu}$ is the distribution over $\boldsymbol{R} \in \mathrm{M}_n(\mathbb{C})$ such that the entries of $\boldsymbol{R}$ are i.i.d. sampled from $\mathcal{D}_\mu$.*

Our aim is to design an average-case approximation algorithm for the permanent of a random matrix $\boldsymbol{R} \sim \mathcal{M}_{n,\mu}$ for $\mu = \mathrm{polylog}^{-1}(n)$. Following the notation used in [EM18], we introduce a matrix $\boldsymbol{X} = \boldsymbol{J} + z\boldsymbol{A}$ where z is a complex variable taken to be $1/\mu$ in the end, $\boldsymbol{J}$ is the all-ones matrix, and $\boldsymbol{A}$ is a random matrix with i.i.d. entries sampled from $\mathcal{D}$. We note that $\frac{\boldsymbol{X}}{z} \sim \mathcal{M}_{n,\mu}$ with $z = 1/\mu$ and thus it is equivalent to compute the permanent of matrix $\boldsymbol{X}$.

DEFINITION 2.4. *Suppose $\boldsymbol{A} \in \mathrm{M}_n(\mathbb{C}), \boldsymbol{B} \in \mathrm{M}_k(\mathbb{C})$. We write $\boldsymbol{B} \subseteq_k \boldsymbol{A}$ if $\boldsymbol{B}$ is a $k \times k$ submatrix of $\boldsymbol{A}$.*

LEMMA 2.1. *Suppose $\boldsymbol{A} \in \mathrm{M}_n(\mathbb{C})$ is any matrix and $\boldsymbol{J}$ is the all-ones matrix of size n. For $k = 0, 1, \ldots, n$, define*

$$a_k = \frac{1}{n^{\underline{k}}} \sum_{\boldsymbol{B} \subseteq_k \boldsymbol{A}} \mathrm{Per}(\boldsymbol{B}), \tag{2.2}$$

Then for all $z \in \mathbb{C}$ we can write

$$\frac{\mathrm{Per}(\boldsymbol{J} + z\boldsymbol{A})}{n!} = \sum_{k=0}^{n} a_k z^k.$$

Proof. This identity can be obtained by simple calculate and similar formula has appeared in [EM18, RW04]. Given any $n \times n$ matrix $\boldsymbol{M}$, define $G_{\boldsymbol{M}}$ to be the corresponding complete bipartite graph with n vertices on each side, both numbered from 1 to n, where the weight of edge $e = (i, j)$ is simply $\boldsymbol{M}_{i,j}$. Define the weight of any perfect matching in $G_{\boldsymbol{M}}$ to be the product of weights of all edges in it. By definition, permanent of $\boldsymbol{M}$ can be seen as weights of all perfect matchings of $G_{\boldsymbol{M}}$. Ideally, we can split any perfect matching of $G_{\boldsymbol{J}+z\boldsymbol{A}}$ into combinations of a k-matching of $G_{\boldsymbol{J}}$ and a $(n-k)$-matching of $G_{z\boldsymbol{A}}$ for some k such that the two matchings together form a perfect matching. Regarding $\mathrm{Per}(\boldsymbol{J} + z\boldsymbol{A})$ as a polynomial of z, the coefficient of z^k is simply the summation of weights over all k-matchings of $G_{\boldsymbol{A}}$ times the weights of all perfect matchings in the left graph of $G_{\boldsymbol{J}}$, which is $(n-k)!$. Thus, the lemma follows by rescaling. ∎

We record here some basic inequalities that we use extensively in the proofs

$$\begin{aligned} n! &\geq \left(\frac{n}{e}\right)^n \quad \forall n \in \mathbb{N}\,, \\ (1+x)^y &\leq e^{xy} \quad \forall x, y > 0\,. \end{aligned} \tag{2.3}$$

And we always assume $0^0 = 1$ in this paper.

3 Main Result

In this section, we state our main result and describe the overall proof structures. The key technical lemmas are proved in the later sections.

THEOREM 3.1. *For any constant $c \in (0, \frac{1}{8})$, there exists a deterministic quasi-polynomial time algorithm $\mathcal{P}$ such that, given both a matrix $\boldsymbol{R}$ sampled from $\mathcal{M}_{n,\mu}$ defined in Definition 2.3 for some complex $\mu : |\mu| \geq \ln^{-c}(n)$ and a real number $\varepsilon \in (0,1)$ as input, the algorithm computes in time $n^{\mathcal{O}(\ln n + \ln \frac{1}{\varepsilon})}$ a complex number $\mathcal{P}(\boldsymbol{R}, \varepsilon)$ that approximates the permanent $\mathrm{Per}(\boldsymbol{R})$ on average in the sense that*

$$\mathbb{P}\left(\left|1 - \frac{\mathcal{P}(\boldsymbol{R}, \varepsilon)}{\mathrm{Per}(\boldsymbol{R})}\right| \leq \varepsilon\right) \geq 1 - o(1),$$

where the probability is over the random matrix $\boldsymbol{R}$.

Proof. As discussed in Section 2, we will work with the permanent of matrix $\boldsymbol{X} = \boldsymbol{J} + z\boldsymbol{A}$ where $\boldsymbol{J}$ is the all-ones matrix. In the following, we design an algorithm that can approximate $\mathrm{Per}(\boldsymbol{X})$ on average for $\boldsymbol{A} \sim \mathcal{M}_n$. The algorithm $\mathcal{P}$ and its performance then follow by a simple scaling argument.

Since $\mathrm{Per}(\boldsymbol{X})$ is a summation of $n!$ products, it is convenient to focus on computing the normalized permanent $\frac{\mathrm{Per}(\boldsymbol{X})}{n!}$, which can be written as $\sum_{k=0}^{n} a_k z^k$ by Lemma 2.1 for

$$a_k = \frac{1}{n^{\underline{k}}} \sum_{\boldsymbol{B} \subseteq_k \boldsymbol{A}} \mathrm{Per}(\boldsymbol{B}).$$

In the technical proof of the paper, we will frequently choose parameters that are either absolute constants or quantities depending on n and ε. These parameters need to satisfy several constraints for the claims in the proof to hold. It is therefore convenient and clear to explicitly enumerate them and their constraints in one place. In the rest of the proof, we will use a set of parameters that form any solution of the following constraints.

$$\begin{cases} 0 < c < \nu < \frac{1}{8}, \\ 0 < \gamma < \beta < \frac{1}{2}, \\ 0 < \gamma < \nu - c, \\ |z| \leq (\ln n)^c, \\ t = \ln n + \ln \frac{1}{\varepsilon}, \\ \theta(n) = \ln \ln n. \end{cases} \tag{3.4}$$

Note that z can be a complex number.

To approximate the normalized permanent of $\boldsymbol{X}$, our algorithm computes the first $t + 1$ coefficients $a_0, a_1, \ldots, a_t$ and outputs the number $\sum_{k=0}^{t} a_k z^k$. For

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

such choice of t, the algorithm has time complexity $t \cdot \binom{n}{t}\binom{n}{t}t! = \mathcal{O}(n^{2t})$, which is quasi-polynomial. The rest of the proof is to show that the first $t+1$ terms in the summation is actually an $\Theta(\varepsilon)$ approximation of $\frac{\operatorname{Per}(\boldsymbol{X})}{n!}$ with high probability. Note that this probability could depend on parameters defined in Eq. (3.4) since they are all fixed constants.

The easy part is to prove that the remaining terms are indeed small with high probability. Namely, with probability $1 - o(1)$,

$$\left|\sum_{k=t+1}^{n} a_k z^k\right| \le n^{-\gamma}\,\varepsilon \tag{3.5}$$

This is small in absolute sense. To show that it is small relatively, we need to give a lower bound of $\sum_{k=0}^{t} a_k z^k$. Namely, with probability $1 - o(1)$,

$$\left|\sum_{k=0}^{t} a_k z^k\right| = \Omega(n^{-\gamma}). \tag{3.6}$$

As the constant in $\Theta(n^{-\gamma})$ does not depend on ε, these two facts together give a proof of the main theorem. At first glance, we might think that γ, as in Eq. (3.4), could be arbitrarily close to 0, which means $n^{-\gamma}$ could be arbitrarily close to 1. However, we would notice from the detailed proofs that the probability for Eq. (3.6) to hold depends highly on γ. In particular, we need $n = \exp\left[\Omega(\gamma^{-\frac{1}{1-c}})\right]$ as in Lemma 6.6.

Eq. (3.5), the easy part, is proven in Lemma 4.3 while Eq. (3.6) is more difficult. To overcome this, we consider symmetric polynomials of the column sum of matrix $\boldsymbol{A}$ to approximate the permanent.

For all $j = 1, 2, \ldots, n$, define

$$C_j \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_{i,j}, \tag{3.7}$$

where $a_{i,j}$ is the (i,j)-the entry of matrix $\boldsymbol{A}$ and for all $k = 0, 1, 2, \ldots, n$,

$$V_k \triangleq \frac{1}{n^{k/2}}\, e_k(C_1, C_2, \ldots, C_n), \tag{3.8}$$

$$D_k \triangleq \frac{1}{n^{k/2}}\, S_k(C_1, C_2, \ldots, C_n), \tag{3.9}$$

where polynomials S_k and e_k are defined in Definition 2.1.

Consider a_k and V_k as multivariate polynomials of $a_{i,j}$'s. Note that, as $n \to \infty, V_k$ and a_k share almost the same monomials and similar weights, i.e. $\frac{1}{n^{k/2}}$ and $\frac{1}{n^{\underline{k}}}$, thus nearly of the same value. Formally, we prove in Lemma 4.4 that with probability $1 - o(1)$,

$$\left|\sum_{k=0}^{t} a_k z^k - \sum_{k=0}^{t} V_k z^k\right| \le n^{-\beta} = o(n^{-\gamma}), \tag{3.10}$$

which is negligible compared to the target $n^{-\gamma}$.

As proven in Lemma 5.1, V_k satisfy the so-called Newton's identities, i.e. for $k \ge 2$,

$$V_k = \frac{V_{k-1}V_1 - V_{k-2}D_2 + \sum_{i=2}^{k-1}(-1)^i V_{k-1-i}D_{i+1}}{k}. \tag{3.11}$$

Furthermore, we prove in Lemma 5.3 that D_2 is concentrated at ξ, the quasi-variance of $\mathcal{D}$, and in Lemma 5.4 that D_k is inverse-polynomially small for $k \ge 3$, both with high probability. This motivates us to consider V'_k, an asymptotic approximation of V_k, as follows

$$V'_k = \begin{cases} 1, & k = 0, \\ V_1, & k = 1, \\ \frac{V'_{k-1}V'_1 - V'_{k-2}\xi}{k}, & k \ge 2. \end{cases} \tag{3.12}$$

Note that k can be larger than n for notation convenience when analyzing $\sum_{k=0}^{\infty} V'_k z^k$. And we prove in Lemma 7.1 that V_k and V'_k are close and in Lemma 7.2 that with probability $1 - o(1)$,

$$\left|\sum_{k=0}^{t} V_k\, z^k - \sum_{k=0}^{t} V'_k\, z^k\right| = \mathcal{O}(n^{c-\nu}) = o(n^{-\gamma}). \tag{3.13}$$

Comparing the two recursions, we use the "probabilists' Hermite polynomials" to explicitly express (V'_k)'s in Eq. (6.27). Due to Lemma 6.5, the summation $\sum_{k=0}^{t} V'_k z^k$ can be estimated by $\sum_{k=0}^{\infty} V'_k z^k$ with a negligible $n^{-\omega(1)}$ additive error by an upper-bound of "probabilists' Hermite polynomials" in Lemma 6.1. This, together with Eqs. (3.10) and (3.13), implies that it is enough to give a $\Omega(n^{-\gamma})$ lower-bound of $\left|\sum_{k=0}^{\infty} V'_k z^k\right|$.

On the other hand, from Eq. (6.28), $\sum_{k=0}^{\infty} V'_k z^k$ is simply $e^{V_1 z - \frac{\xi z^2}{2}}$, where V_1 is the normalized average of all entries in $\boldsymbol{A}$. By Chebyshev's inequality, we know that $|V_1|$ is small with high probability. This can be used to prove the fact (see Lemma 6.6) that with probability $1 - o(1)$,

$$\left|e^{V_1 z - \frac{\xi z^2}{2}}\right| \ge n^{-\gamma},$$

which completes the proof. ∎

If we relax the approximation requirement a bit, we can simply compute V_1 and return $n! e^{V_1 z - \frac{\xi z^2}{2}}$ as an approximation of $\operatorname{Per}(\boldsymbol{X})$. This is a truly polynomial time algorithm and extremely simple. By the above argument, we can get the following approximation guarantee.

Corollary 3.1. *For any constant $c \in (0, \frac{1}{8})$ and $0 < \rho < \frac{1}{8} - c$, there exists a deterministic polynomial*

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

time algorithm $\mathcal{P}$ such that, given a matrix $\boldsymbol{R}$ sampled from $\mathcal{M}_{n,\mu}$ defined in Definition 2.3 for $|\mu| \geq \ln^{-c}(n)$, the algorithm computes a complex number $\mathcal{P}(\boldsymbol{R})$ that approximates the permanent $\mathrm{Per}(\boldsymbol{R})$ on average in the sense that

$$\mathbb{P}\left(\left|1-\frac{\mathcal{P}(\boldsymbol{R})}{\mathrm{Per}(\boldsymbol{R})}\right| \leq n^{-\rho}\right) = 1-o(1),$$

where the probability is over the random matrix $\boldsymbol{R}$.

The drawback of this simple algorithm is that we do not have a parameter ε to control the approximation precision. However, for large n this is already a very good approximation algorithm while for small n we can just compute $\mathrm{Per}(\boldsymbol{R})$ directly by Ryser formula in time 2^n. By this idea, we can convert the above algorithm into a PTAS but not FPTAS, whose running time is polynomial in n but possibly exponential in $\frac{1}{\varepsilon}$. Let us fix a constant $0 < \rho < \frac{1}{8} - c$. For a given approximation parameter ε, if $\varepsilon > n^{-\rho}$ we use the above polynomial time algorithm and otherwise simply compute it directly. The running time is bounded by $\max\left\{\mathrm{poly}(n), 2^{\varepsilon^{-\frac{1}{\rho}}}\right\}$, which shows that the modified algorithm is a PTAS.

COROLLARY 3.2. *For any constant $c \in (0, \frac{1}{8})$, there exists a deterministic PTAS to approximate $\mathrm{Per}(\boldsymbol{R})$ for $1-o(1)$ fraction of random matrices $\boldsymbol{R}$ sampled from $\mathcal{M}_{n,\mu}$ defined in Definition 2.3 for $|\mu| \geq \ln^{-c}(n)$.*

4 Estimation with Summation of Columns

In this section, we prove that the two summations $\sum_{k=t+1}^{n} a_k z^k$ and $\sum_{k=0}^{t}(a_k - V_k)z^k$ are both small with high probability. Their proofs are similar and simply follow from the fact that they have zero mean and exponentially decaying variance.

Recall from Eqs. (2.2) and (3.8) that

$$\begin{cases} a_k = \dfrac{1}{n^{\underline{k}}} \displaystyle\sum_{\boldsymbol{B}\subseteq_k \boldsymbol{A}} \mathrm{Per}(\boldsymbol{B}), \\ V_k = \dfrac{1}{n^{k/2}} \displaystyle\sum_{\{j_1,\cdots,j_k\}\in C_{n,k}} C_{j_1}\cdots C_{j_k}. \end{cases}$$

LEMMA 4.1. *For any $k, \ell \in \{0, 1, \ldots, n\}$,*

$$\mathbb{E}[a_k] = \delta_{k,0}\,, \quad \mathbb{E}[a_k\,\overline{a_\ell}] = \frac{\delta_{k,\ell}}{k!}.$$

Proof. As $a_0 \equiv 1$, it holds that $\mathbb{E}[a_0] = 1$. For $k > 0$,

$$\begin{aligned} \mathbb{E}[a_k] &= \frac{1}{n^{\underline{k}}} \sum_{\boldsymbol{B}\subseteq_k \boldsymbol{A}} \mathbb{E}[\mathrm{Per}(\boldsymbol{B})] \\ &= \frac{1}{n^{\underline{k}}} \sum_{\substack{\{i_1,\cdots,i_k\}\in C_{n,k}\\ \{j_1,\cdots,j_k\}\in C_{n,k}}} \mathbb{E}\left[\sum_{\sigma\in P_{k,k}} \prod_{t=1}^{k} a_{i_t,j_{\sigma_t}}\right] \qquad (4.14)\\ &= \frac{1}{n^{\underline{k}}} \sum_{\substack{\{i_1,\cdots,i_k\}\in C_{n,k}\\ (j_1,\cdots,j_k)\in P_{n,k}}} \mathbb{E}\left[\prod_{t=1}^{k} a_{i_t,j_t}\right]. \end{aligned}$$

By the fact that the entries of $\boldsymbol{A}$ are i.i.d. and have 0 mean value, the above expectation is 0. This proves $\mathbb{E}[a_k] = 0$ for $k > 0$.

For the second part, we first define the following notation for convenience.

$$\Lambda(J,I,J',I') \triangleq \sum_{\substack{\{j_1,\cdots,j_k\}\in J\\ (i_1,\cdots,i_k)\in I}} \sum_{\substack{\{j'_1,\cdots,j'_\ell\}\in J'\\ (i'_1,\cdots,i'_\ell)\in I'}} \mathbb{E}\left[\prod_{t=1}^{k} a_{i_t,j_t} \prod_{t=1}^{\ell} \overline{a_{i'_t,j'_t}}\right], \tag{4.15}$$

where $J \subseteq C_{n,k}, I \subseteq P_{n,k}, J' \subseteq C_{n,\ell}, I' \subseteq P_{n,\ell}$ and $k, l \in [n]$. Since all the entries of $\boldsymbol{A}$ are i.i.d. and of zero mean, the expectation of $\prod_{t=1}^{k} a_{i_t,j_t} \prod_{t=1}^{\ell} \overline{a_{i'_t,j'_t}}$ is non-zero only if

$$\{(i_1,j_1),\cdots,(i_k,j_k)\} = \{(i'_1,j'_1),\cdots,(i'_l,j'_l)\}.$$

Thus, we could simplify $\Lambda(J,I,J',I')$ into

$$\Lambda(J,I,J',I') = \sum_{\substack{\{j_1,\cdots,j_k\}\in J\cap J'\\ (i_1,\cdots,i_k)\in I\cap I'}} \prod_{t=1}^{k} \mathbb{E}\left|a_{i_t,j_t}\right|^2. \tag{4.16}$$

In particular, $\Lambda(J,I,J',I') = 0$ for $k \neq \ell$. Then we could express $\mathbb{E}[a_k\overline{a_\ell}]$ as follows.

$$\begin{aligned} \mathbb{E}[a_k\overline{a_\ell}] =& \frac{1}{n^{\underline{k}}\,n^{\underline{\ell}}} \sum_{\boldsymbol{B}\subseteq_k \boldsymbol{A}} \sum_{\boldsymbol{B}'\subseteq_\ell \boldsymbol{A}} \mathbb{E}\left[\mathrm{Per}(\boldsymbol{B})\overline{\mathrm{Per}(\boldsymbol{B}')}\right]. \\ =& \frac{1}{n^{\underline{k}}\,n^{\underline{\ell}}} \Lambda(C_{n,k}, P_{n,k}, C_{n,\ell}, P_{n,\ell}) \\ =& \frac{\delta_{k,\ell}}{(n^{\underline{k}})^2} \sum_{\substack{\{j_1,\cdots,j_k\}\in C_{n,k}\\ (i_1,\cdots,i_k)\in P_{n,k}}} \prod_{t=1}^{k} \mathbb{E}\left|a_{i_t,j_t}\right|^2 \\ =& \frac{\delta_{k,\ell}}{(n^{\underline{k}})^2} \binom{n}{k} \frac{n!}{k!} = \frac{\delta_{k,\ell}}{k!}. \end{aligned}$$

Here, we use the fact that the variance of each entry is 1. ∎

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

LEMMA 4.2. *For any $m \in \{0,1,\cdots,n\}$ and $k,\ell \in \{1,2,\cdots,n\}$,*

$$\mathbb{E}[V_m] = \delta_{m,0}, \quad \mathbb{E}\Big[(V_k - a_k)\overline{(V_\ell - a_\ell)}\Big] \le \delta_{k,\ell}\frac{k(k-1)}{2n\cdot k!}.$$

Proof. By definition, $V_0 \equiv 1$. Thus, $\mathbb{E}[V_0] = 1$. For $m > 0$, similar to Lemma 4.1,

$$\begin{aligned}\mathbb{E}[V_m] =& \frac{1}{n^{m/2}} \sum_{\{j_1,\cdots,j_m\}\in C_{n,m}} \mathbb{E}[C_{j_1}\cdots C_{j_m}] \\ =& \frac{1}{n^m} \sum_{\substack{\{j_1,\cdots,j_m\}\in C_{n,m}\\ i_1,\cdots,i_m\in[n]}} \mathbb{E}\left[\prod_{t=1}^m a_{i_t,j_t}\right] = 0.\end{aligned}$$

For $\mathbb{E}\Big[(V_k - a_k)\overline{(V_l - a_l)}\Big]$, note that

$$\begin{aligned}V_k - a_k =& \sum_{\substack{\{j_1,\cdots,j_k\}\in C_{n,k}\\ (i_1,\cdots,i_k)\in P_{n,k}}} \prod_{t=1}^k a_{i_t,j_t}\left(\frac{1}{n^k} - \frac{1}{n^{\underline{k}}}\right) \\ &+ \sum_{\substack{\{j_1,\cdots,j_k\}\in C_{n,k}\\ (i_1,\cdots,i_k)\in[n]^k - P_{n,k}}} \prod_{t=1}^k a_{i_t,j_t}\cdot\frac{1}{n^k}.\end{aligned}$$

We can then expand the expectation with $\Lambda(\cdot,\cdot,\cdot,\cdot)$ (see Eq. (4.15)).

$$\begin{aligned}&\mathbb{E}\Big[(V_k - a_k)\overline{(V_\ell - a_\ell)}\Big] \\ =& \left(\frac{1}{n^k} - \frac{1}{n^{\underline{k}}}\right)\left(\frac{1}{n^\ell} - \frac{1}{n^{\underline{\ell}}}\right)\cdot\Lambda(C_{n,k},P_{n,k},C_{n,\ell},P_{n,\ell}) \\ &+ \left(\frac{1}{n^k} - \frac{1}{n^{\underline{k}}}\right)\frac{1}{n^\ell}\cdot\Lambda(C_{n,k},P_{n,k},C_{n,\ell},[n]^\ell - P_{n,\ell}) \\ &+ \frac{1}{n^k}\left(\frac{1}{n^\ell} - \frac{1}{n^{\underline{\ell}}}\right)\cdot\Lambda(C_{n,k},[n]^k - P_{n,k},C_{n,\ell},P_{n,\ell}) \\ &+ \frac{1}{n^{k+\ell}}\cdot\Lambda(C_{n,k},[n]^k - P_{n,k},C_{n,\ell},[n]^\ell - P_{n,\ell}).\end{aligned}$$

By Eq. (4.16), $\mathbb{E}\Big[(V_k - a_k)\overline{(V_\ell - a_\ell)}\Big]$ is non-zero only if $k=\ell$, which proves for $k \ne \ell$. When $k=\ell$,

$$\begin{aligned}&\mathbb{E}\Big[(V_k - a_k)\overline{(V_\ell - a_\ell)}\Big] \\ =& \left(\frac{1}{n^k} - \frac{1}{n^{\underline{k}}}\right)^2\cdot\Lambda(C_{n,k},P_{n,k},C_{n,k},P_{n,k}) \\ &+ \frac{1}{n^{2k}}\cdot\Lambda(C_{n,k},[n]^k - P_{n,k},C_{n,k},[n]^k - P_{n,k}) \\ =& \left(\frac{1}{n^k} - \frac{1}{n^{\underline{k}}}\right)^2 \sum_{\substack{\{j_1,\cdots,j_k\}\in C_{n,k}\\ (i_1,\cdots,i_k)\in P_{n,k}}} \prod_{t=1}^k \mathbb{E}|a_{i_t,j_t}|^2 \\ &+ \frac{1}{n^{2k}} \sum_{\substack{\{j_1,\cdots,j_k\}\in C_{n,k}\\ (i_1,\cdots,i_k)\in[n]^k - P_{n,k}}} \prod_{t=1}^k \mathbb{E}|a_{i_t,j_t}|^2 \\ =& \left(\frac{1}{n^k} - \frac{1}{n^{\underline{k}}}\right)^2 \binom{n}{k} n^{\underline{k}} + \frac{1}{n^{2k}}\binom{n}{k}(n^k - n^{\underline{k}}) \\ =& \frac{n^k - n^{\underline{k}}}{k!n^k} = \frac{1}{k!}\left[1 - \prod_{t=0}^{k-1}\left(1 - \frac{t}{n}\right)\right] \\ \le& \frac{1}{k!}\sum_{t=0}^{k-1}\frac{t}{n} = \frac{k(k-1)}{2n\cdot k!}.\end{aligned}$$

Here, the last inequality comes from the union bound if we consider the probability that none of the k independent bad events, each of which happens with probability $\frac{t}{n}$, happen. $\square$

Then we prove Eqs. (3.5) and (3.10) as follows.

LEMMA 4.3. *With parameters satisfying Eq. (3.4), it holds that*

$$\mathbb{P}\left(\left|\sum_{k=t+1}^n a_k z^k\right| \le n^{-\gamma}\varepsilon\right) \ge 1 - o(1).$$

Proof. To apply Chebyshev's inequality, we first calculate the variance of $\sum_{k=t+1}^n a_k z^k$. By Lemma 4.1, we have

$$\mathbb{E}\left[\sum_{k=t+1}^n a_k z^k\right] = \sum_{k=t+1}^n \mathbb{E}[a_k]\, z^k = 0,$$

and

$$\begin{aligned}\operatorname{Var}\left(\sum_{k=t+1}^n a_k z^k\right) &= \mathbb{E}\left|\sum_{k=t+1}^n a_k z^k\right|^2 \\ &= \sum_{k,l=t+1}^n \mathbb{E}(a_k\overline{a_\ell})\, z^k\overline{z}^\ell = \sum_{k=t+1}^n \frac{|z|^{2k}}{k!}.\end{aligned}$$

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Applying Chebyshev's inequality, we have

$$\mathbb{P}\left[\left|\sum_{k=t+1}^{n} a_k z^k\right| \geq n^{-\gamma}\,\varepsilon\right] \leq \frac{n^{2\gamma}}{\varepsilon^2}\operatorname{Var}\left(\sum_{k=t+1}^{n} a_k z^k\right) \leq \frac{n^{2\gamma}}{\varepsilon^2}\sum_{k=t+1}^{n}\frac{|z|^{2k}}{k!}. \tag{4.17}$$

As chosen in Eq. (3.4),

$$\begin{cases} |z| \leq (\ln n)^{\frac{1}{8}} \\ t = \ln n + \ln \frac{1}{\varepsilon}. \end{cases} \tag{4.18}$$

In this case, for any $k \geq t$ and large n,

$$\frac{|z|^{2(k+1)}}{(k+1)!} \div \frac{|z|^{2k}}{k!} = \frac{|z|^2}{k+1} \leq \frac{|z|^2}{t+1} < 1/2.$$

We can continue Eq. (4.17) as

$$\mathbb{P}\left[\left|\sum_{k=t+1}^{n} a_k z^k\right| \geq n^{-\gamma}\,\varepsilon\right] \leq \frac{n^{2\gamma}}{\epsilon^2}\sum_{k=t+1}^{\infty}\frac{|z|^{2k}}{t!2^{k-t}} \leq \frac{n^{2\gamma}\,|z|^{2t}}{\varepsilon^2\,t!}.$$

Since $t = \ln n + \ln \frac{1}{\varepsilon}$, it is clear that $\frac{t!}{|z|^{2t}}$ is super-polynomial for large n, which means

$$\mathbb{P}\left[\left|\sum_{k=t+1}^{n} a_k z^k\right| \geq n^{-\gamma}\,\varepsilon\right] = o(1). \tag{4.19}$$

∎

Lemma 4.4. *With all parameters satisfying Eq. (3.4),*

$$\mathbb{P}\left(\left|\sum_{k=0}^{t} a_k z^k - \sum_{k=0}^{t} V_k z^k\right| \leq n^{-\beta}\right) = 1 - o(1).$$

Proof. It follows from Lemmas 4.1 and 4.2 that

$$\mathbb{E}[V_k] = \mathbb{E}[a_k] = \delta_{k,0}, \quad V_0 = a_0 \equiv 1.$$

This implies that

$$\operatorname{Var}\left(\sum_{k=0}^{t} a_k z^k - \sum_{k=0}^{t} V_k z^k\right) = \mathbb{E}\left[\sum_{k=1}^{t}(V_k - a_k)z^k\right]\left[\sum_{k=1}^{t}\left(\overline{V_k - a_k}\right)\overline{z}^k\right].$$

With Lemma 4.2, we can then simplify the expansion.

$$\begin{aligned}
&\operatorname{Var}\left(\sum_{k=0}^{t} a_k z^k - \sum_{k=0}^{t} V_k z^k\right) \\
&= \sum_{k=0}^{t} |z|^{2k}\,\mathbb{E}\left[(a_k - V_k)\overline{(a_k - V_k)}\right] \\
&\leq \sum_{k=0}^{t}\frac{k(k-1)}{2n\cdot k!}\,|z|^{2k} = \frac{|z|^4}{2n}\sum_{k=0}^{t-2}\frac{|z|^{2k}}{k!} \leq \frac{|z|^4\,e^{|z|^2}}{2n},
\end{aligned}$$

where the last step comes from the Taylor expansion of $e^{|z|^2}$.

Applying Chebyshev's inequality, we acquire that

$$\mathbb{P}\left(\left|\sum_{k=0}^{t} a_k z^k - \sum_{k=0}^{t} V_k z^k\right| \geq n^{-\beta}\right) \leq \frac{|z|^4\,e^{|z|^2}}{2\,n^{1-2\beta}}.$$

Since $\beta < \frac{1}{2}$ and $|z| \leq (\ln n)^{\frac{1}{8}}$, this probability is in fact $o(1)$. ∎

5 Upper Bounds of the Power-Sum of Columns

In this section, we establish the recursion for V_k's and a concentration bound of D_k. This uses the elementary symmetric polynomial and moment inequalities.

5.1 Newton's Identities in Terms of Symmetric Polynomials The elementary symmetric polynomials and power sums, as defined in Definition 2.1, follow the so-called Newton's identities. We give the following elementary derivation for reader's convenience.

Lemma 5.1. *Given variables $x_1, x_2, \ldots, x_n$ and any $m \in [n]$, we have*

$$e_m(n) = \frac{1}{m}\sum_{k=0}^{m-1}(-1)^k e_{m-k-1}(n)S_{k+1}(n).$$

Proof. Let us introduce auxiliary variables $Q_{m,k}$ defined by

$$Q_{m,k} \triangleq \sum_{\{j_1,\cdots,j_m\}\in C_{n,m}} x_{j_1}x_{j_2}\cdots x_{j_m}\sum_{i=1}^{m} x_{j_i}^{k-1}$$

for any $0 \leq m \leq n, k \geq 1$ with the convention $Q_{0,k} \equiv 0$ for $k \geq 1$. By definition, $Q_{m,1} = me_m(n), Q_{1,k} = S_k$.

Then we consider a counting problem: choose a $(m-1)$-subset A of $[n]$ together with an $i \in [n]$, and the contribution of this choice is $x_i^k \prod_{j\in A} x_j$. On the other hand, we can partition all choices by the criterion whether $k \in A$. Thus,

$$Q_{m,k} + Q_{m-1,k+1} = e_{m-1}(n)S_k$$

holds for all $m \geq 1$. Solving $Q_{m,1}$, the lemma immediately follows. ∎

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

5.2 A Third Moment Inequality By some calculation, we can derive the following upper-bound of the absolute third moment of a sequence of i.i.d. complex random variables.

LEMMA 5.2. *Suppose $X_1, X_2, \cdots, X_n$ is a sequence of i.i.d. random variables following distribution $\mathcal{D}$, then there exists an absolute constant $\eta > 0$ such that*

$$\mathbb{E}\left|\frac{\sum_{i=1}^n X_i}{\sqrt{n}}\right|^3 \le \eta\left(1+\frac{\rho}{\sqrt{n}}\right).$$

Proof. Let

$$\begin{cases}\sigma_1 = \sqrt{\mathbb{E}_{X\sim\mathcal{D}}[\mathrm{Re}(X)^2]}, \quad \sigma_2 = \sqrt{\mathbb{E}_{X\sim\mathcal{D}}[\mathrm{Im}(X)^2]},\\ \rho_1 \triangleq \mathbb{E}_{X\sim\mathcal{D}}\left|\mathrm{Re}(X)\right|^3, \quad \rho_2 \triangleq \mathbb{E}_{X\sim\mathcal{D}}\left|\mathrm{Im}(X)\right|^3,\end{cases}$$

and

$$x_i \triangleq \mathrm{Re}(X_i), \quad y_i \triangleq \mathrm{Im}(X_i).$$

Since $\rho < \infty$, ρ_1 and ρ_2 exists. For $m \in [n]$, define

$$R_m \triangleq \sum_{j=1}^m x_j, \quad T_m \triangleq \sum_{j=1}^m y_j$$

Then we could derive the following recursions.

$$\begin{aligned}\mathbb{E}\left|R_{2k}\right|^3 =& \mathbb{E}\left|R_k + (R_{2k}-R_k)\right|^3\\ \le& \mathbb{E}\left[|R_k| + |R_{2k}-R_k|\right]^3\\ =& \mathbb{E}\left|R_k\right|^3 + \mathbb{E}\left|R_{2k}-R_k\right|^3\\ &+ 3\mathbb{E}\left|R_k\right|^2\left|R_{2k}-R_k\right| + 3\mathbb{E}\left|R_k\right|\left|R_{2k}-R_k\right|^2\\ =& 2\mathbb{E}\left|R_k\right|^3 + 6\mathbb{E}R_k^2\mathbb{E}\left|R_k\right|\\ \le& 2\mathbb{E}\left|R_k\right|^3 + 6\mathbb{E}R_k^2\sqrt{\mathbb{E}R_k^2}\\ =& 2\mathbb{E}\left|R_k\right|^3 + 6k\sigma_1^2\cdot\sqrt{k}\sigma_1.\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left|R_{2k+1}\right|^3 \le& \mathbb{E}\left[|R_{2k}| + |x_{2k+1}|\right]^3\\ \le& \mathbb{E}\left|R_{2k}\right|^3 + \mathbb{E}\left|x_{2k+1}\right|^3\\ &+ 3\mathbb{E}\left|R_{2k}\right|^2\left|x_{2k+1}\right| + 3\mathbb{E}\left|R_{2k}\right|\left|x_{2k+1}\right|^2\\ \le& \mathbb{E}\left|R_{2k}\right|^3 + \mathbb{E}\left|x_{2k+1}\right|^3\\ &+ 3\mathbb{E}R_{2k}^2\sqrt{\mathbb{E}x_{2k+1}^2} + 3\sqrt{\mathbb{E}R_{2k}^2}\mathbb{E}x_{2k+1}^2\\ =& \mathbb{E}\left|R_{2k}\right|^3 + \rho_1 + 6k\sigma_1^3 + 3\sqrt{2k}\sigma_1^3.\end{aligned}$$

Applying induction with the above rules, it is easy to see that

$$\begin{aligned}\mathbb{E}\left|R_n\right|^3 &\le n\rho_1 + \sigma_1^3\sum_{i\ge 1}\left[6\left(\frac{n}{2^i}\right)^{3/2} + \frac{6n}{2^i} + 3\sqrt{2}\sqrt{\frac{n}{2^i}}\right]\\ &\le C'\left(n\rho_1 + n^{3/2}\sigma_1^3\right)\end{aligned}$$

for some constant C'. A similar reasoning for the imaginary part gives

$$\mathbb{E}\left|T_n\right|^3 \le C'\left(n\rho_2 + n^{3/2}\sigma_2^3\right).$$

For $0 \le k \le n$, we also have

$$\begin{aligned}&\mathbb{E}\left|R_n\right|^2\left|T_n\right|\\ =&\mathbb{E}\left|R_k + (R_n - R_k)\right|^2\left|T_k + (T_n - T_k)\right|\\ \le&\mathbb{E}\Big[\left|R_k\right|^2\left|T_k\right| + \left|R_k\right|^2\left|T_n - T_k\right|\\ &+ 2\left|R_k\right|\left|R_n - R_k\right|\left|T_k\right| + 2\left|R_k\right|\left|R_n - R_k\right|\left|T_n - T_k\right|\\ &+ \left|R_n - R_k\right|^2\left|T_k\right| + \left|R_n - R_k\right|^2\left|T_n - T_k\right|\Big]\\ \le&\mathbb{E}R_k^2\left|T_k\right| + \mathbb{E}R_k^2\sqrt{\mathbb{E}T_{n-k}^2}\\ &+ 2\sqrt{\mathbb{E}R_k^2\,\mathbb{E}T_k^2}\sqrt{\mathbb{E}R_{n-k}^2} + 2\sqrt{\mathbb{E}R_k^2}\sqrt{\mathbb{E}R_{n-k}^2\,\mathbb{E}T_{n-k}^2}\\ &+ \mathbb{E}R_{n-k}^2\sqrt{\mathbb{E}T_k^2} + \mathbb{E}R_{n-k}^2\left|T_{n-k}\right|\\ =&\mathbb{E}\Big[\left|R_k\right|^2\left|T_k\right| + \left|R_{n-k}\right|^2\left|T_{n-k}\right|\Big]\\ &+ 3\sigma_1^2\sigma_2\sqrt{k(n-k)}(\sqrt{k} + \sqrt{n-k}).\end{aligned}$$

This establishes that

$$\mathbb{E}\left|R_n\right|^2\left|T_n\right| \le n\mathbb{E}\left|x_1^2 y_1\right| + C''n^{3/2}\sigma_1^2\sigma_2,$$

for some constant C''. Symmetrically, we have

$$\mathbb{E}\left|R_n\right|\left|T_n\right|^2 \le n\mathbb{E}\left|x_1 y_1^2\right| + C''n^{3/2}\sigma_1\sigma_2^2.$$

Therefore,

$$\begin{aligned}&\mathbb{E}\left|\frac{\sum_{j=1}^n X_j}{\sqrt{n}}\right|^3\\ =&n^{-\frac{3}{2}}\mathbb{E}\left|R_n + iT_n\right|^3\\ \le&n^{-\frac{3}{2}}\mathbb{E}\big[\left|R_n\right|^3 + \left|T_n\right|^3 + 3\left|R_n\right|^2\left|T_n\right| + 3\left|R_n\right|\left|T_n\right|^2\big]\\ \le&n^{-\frac{3}{2}}\Big[C'n(\rho_1+\rho_2) + C'n^{\frac{3}{2}}\left(\sigma_1^3 + \sigma_2^3\right)\\ &+ 3C''n^{\frac{3}{2}}\sigma_1\sigma_2(\sigma_1+\sigma_2) + 3n\mathbb{E}\big(|x_1|^2|y_1| + |x_1||y_1|^2\big)\Big]\end{aligned}$$

Using the basic inequality

$$x^2y + xy^2 \le x^3 + y^3, \quad \forall x, y > 0,$$

and the fact that

$$\begin{cases}\rho_1, \rho_2 \le \rho,\\ \sigma_1, \sigma_2 \le 1,\end{cases}$$

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

we have for some constant $\eta > 0$ that

$$\begin{aligned}&\mathbb{E}\left|\frac{\sum_{j=1}^n X_j}{\sqrt{n}}\right|^3\\ \le\,& n^{-\frac{3}{2}}\left[2C'n\rho + 2C'n^{\frac{3}{2}} + 6C''n^{\frac{3}{2}} + 3n\mathbb{E}\big(|x_1|^3 + |y_1|^3\big)\right]\\ \le\,& \eta(1+\frac{\rho}{\sqrt{n}}).\end{aligned}$$

□

5.3 Bounds for D_k

LEMMA 5.3. *For any $0 < \phi < \frac{1}{2}$,*

$$\mathbb{P}\big(|D_2 - \xi| \le n^{-\phi}\big) = 1 - o(1).$$

Proof. Define

$$\begin{cases} X_{i,j} \triangleq a_{i,j}\mathbb{1}_{|a_{i,j}|\le n},\\ \mu_k \triangleq \mathbb{E}X_{i,j}^k,\\ \mu_k^* \triangleq \mathbb{E}\left|X_{i,j}\right|^k,\\ \mu^\dagger \triangleq \mathbb{E}\Big[\left|X_{i,j}\right|^2 X_{i,j}\Big].\end{cases}$$

Since all elements in $\boldsymbol{A}$ are i.i.d. and $X_{i,j}$'s are bounded, these values are well-defined. Note that we only care about the asymptotic behavior, we assume $n \ge \rho$ in the following proof.

Observe that

$$\mathbb{P}(|a_{i,j}| > n) \le \frac{\mathbb{E}\left|a_{i,j}\right|^3}{n^3} \le \frac{\rho}{n^3},$$

$\boldsymbol{A}$ satisfies

$$\mathbb{P}\big(\exists i,j \in [n] : |a_{i,j}| > n\big) \le \frac{\rho}{n}.$$

Therefore,

$$\begin{aligned}&\mathbb{P}(|D_2 - \xi| > \varepsilon)\\ \le\,&\mathbb{P}\left(\left|\frac{\sum_{j=1}^n\left(\sum_{i=1}^n X_{i,j}\right)^2}{n^2} - \xi\right| > \varepsilon\right) + \frac{\rho}{n}.\end{aligned} \tag{5.20}$$

Next, we bound some moments. For μ_1,

$$\begin{aligned}|\mu_1| =& \left|-\mathbb{E}\left[a_{i,j}\mathbb{1}_{\{|a_{i,j}|>n\}}\right]\right| \le \mathbb{E}\left[\left|a_{i,j}\right|\mathbb{1}_{|a_{i,j}|>n}\right]\\ \le&\,\mathbb{E}\left[\left|a_{i,j}\right|\left(\left|a_{i,j}\right|/n\right)^2\right] \le \frac{\rho}{n^2}.\end{aligned}$$

For μ_2, we first notice that

$$\begin{aligned}|\xi - \mu_2| =& \left|\mathbb{E}\left[a_{1,1}^2 - a_{1,1}^2\mathbb{1}_{|a_{1,1}|\le n}\right]\right|\\ =& \left|\mathbb{E}\big[a_{1,1}^2\mathbb{1}_{|a_{1,1}|>n}\big]\right| \le \mathbb{E}\Big[\left|a_{1,1}\right|^3/n\Big] = \rho/n.\end{aligned}$$

Plus, $n \ge \rho$ by assumption. We could then derive

$$|\mu_2| \le |\xi| + \rho/n \le 1 + \rho/n \le 2.$$

Also, for μ_2^*, μ_4^* and $\mu^\dagger$,

$$\begin{cases}\mu_2^* =\mathbb{E}\Big[\left|a_{i,j}\right|^2\mathbb{1}_{|a_{i,j}|\le n}\Big] \le \mathbb{E}\Big[\left|a_{i,j}\right|^2\Big] = 1\\ \mu_4^* =\mathbb{E}\Big[\left|a_{i,j}\right|^4\mathbb{1}_{|a_{i,j}|\le n}\Big] \le n\mathbb{E}\Big[\left|a_{i,j}\right|^3\Big] \le n\rho\\ \left|\mu^\dagger\right| = \left|\mathbb{E}\Big[\left|a_{i,j}\right|^2 a_{i,j}\mathbb{1}_{|a_{i,j}|\le n}\Big]\right| \le \mathbb{E}\left|a_{i,j}\right|^3 = \rho.\end{cases}$$

Let $S_j \triangleq \sum_{i=1}^n X_{i,j}, S \triangleq \sum_{j=1}^n (S_j^2 - n\xi)$. Since

$$\begin{aligned}\mathbb{E}\left|S\right|^2 =&\, \mathrm{Var}[S] + \left|\mathbb{E}[S]\right|^2\\ \le&\, n\,\mathrm{Var}\big[S_1^2 - n\xi\big] + n^2\left|\mathbb{E}\big[S_1^2 - n\xi\big]\right|^2\\ =&\, n\,\mathrm{Var}\big[S_1^2\big] + n^2\left|\mathbb{E}\big[S_1^2 - n\xi\big]\right|^2\\ \le&\, n\,\mathbb{E}\big[S_1^2\bar{S}_1^2\big] + n^2\left|\mathbb{E}\big[S_1^2 - n\xi\big]\right|^2,\end{aligned} \tag{5.21}$$

we only need to bound $\left|\mathbb{E}\big[S_1^2 - n\xi\big]\right|$ and $\mathbb{E}\big[S_1^2\bar{S}_1^2\big]$ separately.

For the first part, since $n \ge \rho$,

$$\begin{aligned}\left|\mathbb{E}\big[S_1^2 - n\xi\big]\right| =& \left|n\mu_2 + n(n-1)\mu_1^2 - n\xi\right|\\ \le&\, n\left|\mu_2 - \xi\right| + n(n-1)\left|\mu_1\right|^2 \le \rho + 1.\end{aligned}$$

For the second, consider all five kinds of monomials in

$$\mathbb{E}\Big[S_1^2\,\overline{S}_1^2\Big] = \sum_{i,j,k,l\in[n]}\mathbb{E}\big[X_{i,1}X_{j,1}\overline{X_{k,1}X_{l,1}}\big],$$

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

we could expand it as follows.

$$
\begin{aligned}
&\mathbb{E}\Big[S_1^2\,\overline{S}_1^2\Big]\\
&= \sum_{i=1}^{n}\mathbb{E}\Big[X_i^2\overline{X}_i^2\Big]\\
&\quad+\sum_{i\neq j}\mathbb{E}\Big[X_iX_j\overline{X}_j\overline{X}_j+X_jX_i\overline{X}_j\overline{X}_j\\
&\qquad\qquad+X_jX_j\overline{X}_i\overline{X}_j+X_jX_j\overline{X}_j\overline{X}_i\Big]\\
&\quad+\sum_{i<j}\ \mathbb{E}\big[X_iX_i\overline{X}_j\overline{X}_j+X_jX_j\overline{X}_i\overline{X}_i+4X_iX_j\overline{X}_i\overline{X}_j\big]\\
&\quad+\sum_{(i,j,k)\in P_{n,3},\,j<k}\mathbb{E}\Big[2X_iX_i\overline{X}_j\overline{X}_k+2X_jX_k\overline{X}_i\overline{X}_i\\
&\qquad\qquad+4X_iX_j\overline{X}_i\overline{X}_k+4X_iX_k\overline{X}_i\overline{X}_j\Big]\\
&\quad+\sum_{(i,j,k,l)\in P_{n,4}}\mathbb{E}\big[X_iX_j\overline{X}_k\overline{X}_l\big]\\
&= n\mu_4^*+n(n-1)\left(2\mu_1\overline{\mu^\dagger}+2\overline{\mu_1}\mu^\dagger\right)\\
&\quad+\binom{n}{2}\left[2\mu_2\overline{\mu_2}+4(\mu_2^*)^2\right]\\
&\quad+n\binom{n-1}{2}\left(2\mu_2\overline{\mu_1}^2+2\mu_1^2\overline{\mu_2}+8\mu_2^*\mu_1\overline{\mu_1}\right)+n^4\mu_1^2\overline{\mu_1^2}\\
&\le n\mu_4^*+2n^2\left(\mu_1\overline{\mu^\dagger}+\overline{\mu_1}\mu^\dagger\right)+n^2\left[\mu_2\overline{\mu_2}+2(\mu_2^*)^2\right]\\
&\quad+n^3\left(\mu_2\overline{\mu_1}^2+\mu_1^2\overline{\mu_2}+4\mu_2^*\mu_1\overline{\mu_1}\right)+n^4\left|\mu_1\right|^4\\
&\le n^2\rho+4n^2\frac{\rho}{n^2}\cdot\rho+n^2\left(2^2+2\cdot1^2\right)\\
&\quad+n^3\left(2\frac{\rho^2}{n^4}+2\frac{\rho^2}{n^4}+4\cdot1\cdot\frac{\rho^2}{n^4}\right)+n^4\frac{\rho^4}{n^8}\\
&\le 20n^2\rho,
\end{aligned}
$$

where we slightly abuse the notation to use X_i to denote $X_{i,1}$ and use the assumption $n\ge\rho$ in the last step. Therefore, by Eq. (5.21), $n\ge\rho\ge(\sigma_1^2+\sigma_2^2)^{3/2}=1$, and Chebyshev's inequality, we have

$$
\begin{aligned}
\mathbb{P}\left(\left|\frac{S}{n^2}\right|>\varepsilon\right)&\le\frac{\mathbb{E}\big[S\overline{S}\big]}{n^4\varepsilon^2}\le\frac{n\times20n^2\rho+n^2(\rho+1)^2}{n^4\varepsilon^2}\\
&\le\frac{20n^3\rho+n^24\rho^2}{n^4\varepsilon^2}\le\frac{24\rho}{n\varepsilon^2}.
\end{aligned}
$$

Taking $\varepsilon=n^{-\phi}$ with $0<\phi<\frac{1}{2}$ and applying Eq. (5.20), the lemma then follows from

$$
\mathbb{P}\big(|D_2-\xi|\le n^{-\phi}\big)\ge1-\frac{\rho}{n}-24\rho n^{2\phi-1}=1-o(1). \tag{5.22}
$$

∎

Lemma 5.4. *Fix any positive constant $\Delta<\frac{1}{6}$, it holds that*

$$
\mathbb{P}\big(\forall\,k\ge3,\ |D_k|\le n^{-\Delta k}\big)=1-o(1).
$$

Proof. The statement is equivalent to the following bound

$$
\mathbb{P}\Big(\exists\,k\ge3,\ |D_k|>n^{-\Delta k}\Big)=o(1).
$$

The left-hand side can be bounded as

$$
\begin{aligned}
&\mathbb{P}\Big(\exists\,k\ge3,\ |D_k|>n^{-\Delta k}\Big)\\
&=\mathbb{P}\left(\exists\,k\ge3,\ \frac{\sum_{j=1}^n|C_j|^k}{n^{k/2}}>n^{-\Delta k}\right)\\
&=\mathbb{P}\left(\exists\,k\ge3,\ \left(\sum_{j=1}^n|C_j|^k\right)^{1/k}>n^{1/2-\Delta}\right)\\
&\le\mathbb{P}\left(\left(\sum_{j=1}^n|C_j|^3\right)^{1/3}>n^{1/2-\Delta}\right),
\end{aligned}\tag{5.23}
$$

where the last step follows from the well-known decreasing property of the L^p norm,

$$
\left(\sum_{j=1}^n|C_j|^k\right)^{1/k}\le\left(\sum_{j=1}^n|C_j|^3\right)^{1/3}\quad\forall\,k\ge3.
$$

Recall that by Lemma 5.2, there is a constant $\eta>0$, such that

$$
\mathbb{E}|C_j|^3\le\eta\left(1+\frac{\rho}{\sqrt{n}}\right)\quad\forall\,j\in[n].
$$

We can continue the bound by Markov's inequality in Eq. (5.23) as

$$
\begin{aligned}
\mathbb{P}\Big(\exists\,k\ge3:|D_k|>n^{-\Delta k}\Big)&\le\frac{n\,\mathbb{E}\,|C_j|^3}{n^{3/2-3\Delta}}\\
&\le\frac{\eta(1+\rho n^{-1/2})}{n^{1/2-3\Delta}}.
\end{aligned}
$$

The right-hand side is $o(1)$ for $\Delta<1/6$ and this proves the lemma. ∎

6 Explicit Expression and Upper-bounds of V_k'

In this section, we solve the recursion of V_k' utilizing the well-known "probabilists' Hermite polynomials" and establish some bounds of V_k' which will be used to bound the difference between V_k and V_k' in the next section.

6.1 Probabilists' Hermite Polynomials The "probabilists' Hermite polynomials" are given by

$$
H_{e_n}(x)=(-1)^ne^{\frac{x^2}{2}}\frac{\mathrm{d}^n}{\mathrm{d}x^n}e^{-\frac{x^2}{2}}=\left(x-\frac{\mathrm{d}}{\mathrm{d}x}\right)^n\cdot1
$$

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited
Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

for $n \in \mathbb{N}$. The following explicit expression could then be derived by solving this equation.

$$H_{e_n}(x) = n! \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^k x^{n-2k}}{k!\,(n-2k)!\,2^k}, \tag{6.24}$$

Note that $H_{e_n}(x)$ satisfies

$$H_{e_n}(x) = \begin{cases} 1, & n = 0, \\ x, & n = 1, \\ xH_{e_{n-1}}(x) - (n-1)H_{e_{n-2}}(x), & n \geq 2. \end{cases}$$

We can derive a similar recursion for $h_n(x) \triangleq \frac{1}{n!} H_{e_n}(x)$.

$$h_n(x) = \begin{cases} 1, & n = 0, \\ x, & n = 1, \\ \frac{xh_{n-1}(x) - h_{n-2}(x)}{n}, & n \geq 2. \end{cases} \tag{6.25}$$

The following upper bound on the Hermite polynomials will be useful in later proofs.

Lemma 6.1. *For any $n \in \mathbb{N}$ and any $x \in \mathbb{C}$, it holds that*

$$|h_n(x)| \leq \max(1, |x|)^n \left(\frac{n}{e^2}\right)^{-\frac{n}{2}}.$$

Proof. By the definition of $h_n(x)$, we have

$$\begin{aligned} |h_n(x)| &\leq \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \frac{|x|^{n-2k}}{k!\,(n-2k)!\,2^k} \\ &\leq \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} |x|^{n-2k} \left(\frac{2k}{e}\right)^{-k} \left(\frac{n-2k}{e}\right)^{-n+2k}. \end{aligned}$$

Use $\phi(k)$ to denote the inverse of the coefficient of the k-th term of the previous equation,

$$\phi(k) \triangleq \left(\frac{2k}{e}\right)^k \left(\frac{n-2k}{e}\right)^{n-2k} > 0$$

for $k \in \left[0, \lfloor \frac{n}{2} \rfloor\right]$. The derivative of $\ln \phi(k)$ is

$$\frac{\mathrm{d}}{\mathrm{d}k} \ln \phi(k) = \ln(2k) - 2\ln(n-2k),$$

showing that the minimum value of $\phi(k)$ is achieved at

$$k_0 = \frac{n}{2} + \frac{1}{4} - \frac{\sqrt{4n+1}}{4}.$$

For $n \geq 1$, we have

$$\begin{cases} k_0 \geq \frac{n-\sqrt{n}}{2} \geq 0, \\ n - 2k_0 \geq \sqrt{n} - \frac{1}{2} \geq 0, \end{cases}$$

and furthermore

$$\begin{aligned} \phi(k) &\geq \phi(k_0) \\ &\geq \left(\frac{n-\sqrt{n}}{e}\right)^{\frac{n-\sqrt{n}}{2}} \left(\frac{\sqrt{n}-\frac{1}{2}}{e}\right)^{\sqrt{n}-\frac{1}{2}} \\ &= e^{-\frac{n+\sqrt{n}-1}{2}} \left(n-\sqrt{n}\right)^{\frac{n-\sqrt{n}}{2}} \left(n - \sqrt{n} + \frac{1}{4}\right)^{\frac{\sqrt{n}}{2}-\frac{1}{4}} \\ &> e^{-\frac{n+\sqrt{n}-1}{2}} \left(n-\sqrt{n}\right)^{\frac{n}{2}-\frac{1}{4}} \\ &= e^{-\frac{n+\sqrt{n}-1}{2}} n^{\frac{n}{2}} \left(n-\sqrt{n}\right)^{-\frac{1}{4}} \left(1 - n^{-\frac{1}{2}}\right)^{\frac{n}{2}} \\ &= \left(\frac{n}{e}\right)^{\frac{n}{2}} e^{-\frac{\sqrt{n}-1}{2}} \left(n-\sqrt{n}\right)^{-\frac{1}{4}} \left(1 - n^{-\frac{1}{2}}\right)^{\frac{n}{2}}. \end{aligned}$$

Since $\left(1 - \frac{1}{x}\right)^x$ is increasing in $(1, \infty)$, it holds that for all $n \geq 4$,

$$\begin{aligned} \left(1 - n^{-\frac{1}{2}}\right)^{\frac{n}{2}} &= \left[\left(1 - n^{-\frac{1}{2}}\right)^{\sqrt{n}}\right]^{\frac{\sqrt{n}}{2}} \\ &\geq \left(1 - \frac{1}{\sqrt{4}}\right)^{\sqrt{4}\cdot\frac{\sqrt{n}}{2}} = 2^{-\sqrt{n}}. \end{aligned}$$

Then we can continue the bound on $\phi(k)$ as

$$\begin{aligned} \phi(k) &\geq \left(\frac{n}{e}\right)^{\frac{n}{2}} e^{-\frac{\sqrt{n}-1}{2}} \times n^{-\frac{1}{4}} \times 2^{-\sqrt{n}} \\ &= \left(\frac{n}{e}\right)^{\frac{n}{2}} \exp\left(-\frac{\sqrt{n}-1}{2} - \frac{1}{4}\ln n - \sqrt{n}\ln 2\right). \end{aligned}$$

When $n \geq 25$, it holds that

$$-\frac{\sqrt{n}-1}{2} - \frac{1}{4}\ln n - \sqrt{n}\ln 2 \geq -\frac{n}{2} + \ln n,$$

which implies

$$\begin{aligned} \phi(k) &\geq \left(\frac{n}{e}\right)^{\frac{n}{2}} \exp\left(-\frac{n}{2} + \ln n\right) \\ &= n\left(\frac{n}{e^2}\right)^{\frac{n}{2}} \geq \frac{n+2}{2}\left(\frac{n}{e^2}\right)^{\frac{n}{2}}. \end{aligned} \tag{6.26}$$

Checking the remaining cases by hand, we conclude that

$$\phi(k) \geq \frac{n+2}{2}\left(\frac{n}{e^2}\right)^{\frac{n}{2}}$$

holds for $n \in \mathbb{N}, 0 \leq k \leq \lfloor \frac{n}{2} \rfloor$ with convention that $0^0 = 1$. Thus,

$$\begin{aligned} |h_n(x)| &\leq \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} |x|^{n-2k} / \phi(k) \\ &\leq \frac{n+2}{2} \cdot \max(1, |x|)^n \cdot \frac{2}{n+2}\left(\frac{n}{e^2}\right)^{-\frac{n}{2}} \\ &= \max(1, |x|^n)\left(\frac{n}{e^2}\right)^{-\frac{n}{2}}. \end{aligned}$$

∎

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

6.2 Upper-bound of V'_k Comparing the recursion of V'_k in Eq. (3.12) to that of $h_k(x)$ from Eq. (6.25), we have

$$\begin{cases} V'_k = \frac{V_1^k}{k!} & \text{if } \xi = 0, \\ V'_k = \xi^{\frac{k}{2}} h_k\left(\frac{V_1}{\sqrt{\xi}}\right) & \text{otherwise.} \end{cases} \tag{6.27}$$

Lemma 6.1 can be used to establish an upper bound of V'_k by using Eq. (6.27).

LEMMA 6.2. *For all function $\theta(n) = \omega(1)$, it holds that*

$$\mathbb{P}(|V_1| \le \theta) = 1 - o(1).$$

Proof. By the Chebyshev's inequality, we have

$$\mathbb{P}(|V_1| > \theta) \le \frac{\mathrm{Var}(V_1)}{\theta^2} = o(1).$$

□

LEMMA 6.3. *For any $k \in \mathbb{N}$, it holds that*

$$|V'_k| \le \max\left(1, |V_1|^k\right) \left(\frac{k}{e^2}\right)^{-\frac{k}{2}}.$$

Note that k might be larger than n for notation convenience in Eq. (6.28).

Proof. Consider the following two cases depending on whether $\xi = 0$ or not.

1. $\xi = 0$. By definition, we have $V'_k = \frac{V_1^k}{k!}$ and
$$\begin{aligned} |V'_k| = \frac{|V_1|^k}{k!} &\le |V_1|^k \left(\frac{k}{e}\right)^{-k} \\ &\le \max\left(1, |V_1|^k\right) \left(\frac{k}{e^2}\right)^{-\frac{k}{2}}. \end{aligned}$$

2. $\xi \ne 0$. Recall that $V'_k = \xi^{\frac{k}{2}} h_k\left(\frac{V'_1}{\sqrt{\xi}}\right)$. We can apply Lemma 6.1 as follows.
$$\begin{aligned} |V'_k| &= |\xi|^{\frac{k}{2}} \left| h_k\left(\frac{V_1}{\sqrt{\xi}}\right) \right| \\ &\le |\xi|^{\frac{k}{2}} \max\left(1, \left|\frac{V_1}{\sqrt{\xi}}\right|\right)^k \left(\frac{k}{e^2}\right)^{-\frac{k}{2}} \\ &\le \max\left(1, |V_1|^k\right) \left(\frac{k}{e^2}\right)^{-\frac{k}{2}}, \end{aligned}$$
where in the final step we used the fact that $|\xi| \le 1$.

□

LEMMA 6.4. *Let $\theta \triangleq \theta(n) = o(\sqrt[4]{\ln n})$ be a function of n such that $\theta \ge 1$ and $|V_1| \le \theta$. Fixing any constant $\tau > 0$, for sufficiently large n and any $k \in \mathbb{N}$, it holds that*

$$|V'_k| \le n^\tau k^{-\frac{k}{4}}.$$

Additionally, we have the uniform bound

$$|V'_k| \le e^{2\theta^2}.$$

Proof. By Lemma 6.3, we have

$$|V'_k| \le \max\left(1, |V_1|^k\right) \left(\frac{k}{e^2}\right)^{-\frac{k}{2}}.$$

This together with $|V_1| \le \theta$ implies that, for all $k \ge 0$,

$$|V'_k| \le \theta^k e^k k^{-\frac{k}{2}} = \exp\left(k \ln \theta + k - \frac{k \ln k}{2}\right) = (*)$$

Define function

$$\phi(x) = \theta^x e^x x^{-\frac{x}{4}},$$

for $x \ge 0$. Calculating the derivative of $\ln \phi(x)$, we see that the maximum value of $\phi(x)$ is achieved at $x = e^3 \theta^4$ and

$$\phi(x) \le \phi(e^3 \theta^4) = \exp\left(\frac{e^3 \theta^4}{4}\right) = n^{o(1)},$$

where in the last step we use the condition $\theta = o(\sqrt[4]{\ln n})$. Then for sufficiently large n, $\phi(x)$ is bounded by n^τ, which means

$$|V'_k| \le \phi(k) k^{-\frac{k}{4}} \le n^\tau k^{-\frac{k}{4}}.$$

For the uniform bound, by calculating the derivative of $(*)$ it follows that

$$\begin{aligned} |V'_k| &\le \exp\left(k \ln \theta + k - \frac{k \ln k}{2}\right) \Bigg|_{k = e\theta^2} \\ &= \exp\left(\frac{e\theta^2}{2}\right) < e^{2\theta^2}. \end{aligned}$$

□

6.3 Summation of V'_k In view of the following two well-known expansion formulas, for any $z, t \in \mathbb{C}$,

$$\begin{cases} \sum_{k=0}^{\infty} \frac{z^k}{k!} = e^z, \\ \sum_{k=0}^{\infty} \frac{H_{e_k}(z) t^k}{k!} = e^{zt - \frac{t^2}{2}}. \end{cases}$$

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

Thus, Eq. (6.27) implies

$$\sum_{k=0}^{\infty} V_k' z^k = \begin{cases} \sum_{k=0}^{\infty} \frac{V_1^k z^k}{k!} = e^{V_1 z}, & \xi = 0 \\ \sum_{k=0}^{\infty} \frac{\sqrt{\xi}^k H_{e_k}\left(\frac{V_1}{\sqrt{\xi}}\right) z^k}{k!} = e^{V_1 z - \frac{\xi z^2}{2}}, & \xi \neq 0 \end{cases}$$

Therefore, in both cases,

$$\sum_{k=0}^{\infty} V_k' z^k = e^{V_1 z - \frac{\xi z^2}{2}}. \tag{6.28}$$

With help of Lemma 6.4, we prove the following tail bound.

LEMMA 6.5. *With all parameters satisfying Eq. (3.4), with probability* $1 - o(1)$,

$$\left| \sum_{k=t+1}^{\infty} V_k' z^k \right| = n^{-\omega(1)}.$$

Proof. Applying Lemma 6.2 with $\theta(n) = \ln \ln n$ as in Eq. (3.4), with probability $1 - o(1)$,

$$|V_1| \leq \theta.$$

In this case, it follows from Lemma 6.4 that

$$\left| \sum_{k=t+1}^{\infty} V_k' z^k \right| \leq \sum_{k=t+1}^{\infty} |V_k'| \, |z|^k \leq n^{\tau} \sum_{k=t+1}^{\infty} k^{-\frac{k}{4}} \, |z|^k.$$

As in Eq. (3.4), $|z|^8 \leq \ln n < t$, which means for sufficiently large n,

$$\frac{(k+1)^{-\frac{k+1}{4}} \, |z|^{k+1}}{k^{-\frac{k}{4}} \, |z|^k} = \frac{|z|}{\sqrt[4]{k+1}} \left(1 + \frac{1}{k}\right)^{-\frac{k}{4}} < \frac{|z|}{\sqrt[4]{t}} \leq \frac{1}{2}.$$

Thus,

$$\left| \sum_{k=t+1}^{\infty} V_k' z^k \right| \leq n^{\tau} t^{-\frac{t}{4}} \, |z|^t = n^{-\omega(1)}.$$

□

Since $|e^z| = e^{\mathrm{Re}(z)}$ holds for $z \in \mathbb{C}$. Eq. (6.28) says that the summation is small only if the $\mathrm{Re}(V_1')$ is small, which has small probability by concentration of V_1. Formally,

LEMMA 6.6. *With all parameters satisfying Eq. (3.4),*

$$\mathbb{P}\left[\left| e^{V_1 z - \frac{\xi z^2}{2}} \right| \geq n^{-\gamma}\right] = 1 - o(1).$$

Proof. We upper-bound the probability

$$\begin{aligned} & \mathbb{P}\left[\left| e^{V_1 z - \frac{\xi z^2}{2}} \right| < n^{-\gamma}\right] \\ & = \mathbb{P}\left[\mathrm{Re}\left(V_1 z - \frac{\xi z^2}{2}\right) < -\gamma \ln n\right] \\ & = \mathbb{P}\left[\mathrm{Re}\left(V_1 \frac{z}{|z|}\right) < -\frac{\gamma \ln n}{|z|} + \frac{\mathrm{Re}(\xi z^2)}{2\,|z|}\right]. \end{aligned} \tag{6.29}$$

Since $|z|^8 \leq \ln n$, for large n, it holds that

$$\frac{\mathrm{Re}(\xi z^2)}{2\,|z|} \leq \left| \frac{\xi z^2}{2z} \right| \leq \frac{\gamma \ln n}{2\,|z|}.$$

Therefore, we can continue Eq. (6.29) as

$$\mathbb{P}\left[\left| e^{V_1 z - \frac{\xi z^2}{2}} \right| < n^{-\gamma}\right] \leq \mathbb{P}\left[|V_1| > \frac{\gamma \ln n}{2\,|z|}\right],$$

which is easily shown to be $o(1)$ applying Lemma 6.2. □

7 Difference of V_k and V_k'

In this section, we bound the difference between V_k and V_k'. To this end, we simply apply triangle inequality of absolute values and induction repeatedly.

LEMMA 7.1. *With* $\theta(n) = \ln \ln n$ *as in Eq. (3.4), fixing any positive constant* $\nu < \frac{1}{8}$, *there exists a constant* $n_k = n_k(\sigma_1, \sigma_2, \delta, \rho)$ *such that for any* $n \geq n_k$, *with probability* $1 - o(1)$, *the difference* $\varepsilon_k \triangleq |V_k' - V_k|$ *is bounded by*

$$\varepsilon_k \leq n^{-\nu} k^{-\nu k}$$

for any $0 \leq k \leq n$.

Proof. Note that $V_0 = V_0' \equiv 1$ and $V_1 \equiv V_1'$ by definition. This gives $\varepsilon_0 = \varepsilon_1 = 0$.

For $k \geq 2$, recall that

$$\begin{cases} V_k = \frac{V_{k-1} V_1 - V_{k-2} D_2 + \sum_{i=2}^{k-1} (-1)^i V_{k-1-i} D_{i+1}}{k}, \\ V_k' = \frac{V_{k-1}' V_1' - V_{k-2}' \xi}{k}. \end{cases} \tag{7.30}$$

The triangle inequality and the bound $|\xi| \leq 1$ as proved

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

in Eq. (2.1) then establish the following upper bound

$$\begin{aligned}
k\,\varepsilon_k = &\Big|\left(V'_{k-1}V'_1 - V'_{k-2}\xi\right) \\
&\quad - \Big(V_{k-1}V_1 - V_{k-2}D_2 + \sum_{i=2}^{k-1}(-1)^i V_{k-1-i}D_{i+1}\Big)\Big| \\
\le &\Big|V'_{k-1}V_1 - V_{k-1}V_1 - V'_{k-2}\xi + V_{k-2}\xi \\
&\quad - V_{k-2}\xi + V_{k-2}D_2\Big| + \sum_{i=2}^{k-1}|V_{k-1-i}D_{i+1}| \\
\le &\,|V_1|\,\varepsilon_{k-1} + \varepsilon_{k-2} + |V_{k-2}|\,|D_2-\xi| \\
&\quad + \sum_{i=2}^{k-1}|V_{k-1-i}|\,|D_{i+1}|\,.
\end{aligned}$$

Therefore, we can bound ε_k as

(7.31)

$$\begin{aligned}
\varepsilon_k \le \frac{1}{k}\Big(&|V_1|\,\varepsilon_{k-1} + \varepsilon_{k-2} \\
&+ \varepsilon_{k-2}\,|D_2-\xi| + \left|V'_{k-2}\right|\,|D_2-\xi| \\
&+ \sum_{i=2}^{k-1}\left|V'_{k-1-i}\right|\,|D_{i+1}| + \sum_{i=2}^{k-1}\varepsilon_{k-1-i}\,|D_{i+1}|\Big).
\end{aligned}$$

Choose τ and Δ such that

$$(7.32)\qquad \begin{cases} \tau > 0, \\ \Delta > \nu, \\ \frac{1}{8} < \Delta < \frac{1}{6}, \\ 2\tau + \nu < 2\Delta. \end{cases}$$

Applying Lemma 6.2 with $\theta(n) = \ln\ln n$ as assumed,

$$\mathbb{P}\left(|V_1| \le \theta\right) = 1 - o(1).$$

Plus Lemmas 5.3 and 5.4, with probability $1-o(1)$, it holds that

$$(7.33)\qquad \begin{cases} |V'_1| \le \theta, \\ |D_2 - \xi| \le n^{-2\Delta}, \\ |D_k| \le n^{-k\Delta}. \end{cases}$$

For the rest of this proof, we assume that Eq. (7.33) holds. In this case,

(7.34)

$$\begin{aligned}
\varepsilon_k &\le \frac{1}{k}\Big[|V'_1|\,\varepsilon_{k-1} + \varepsilon_{k-2} + \sum_{i=2}^{k}\left(\varepsilon_{k-i} + \left|V'_{k-i}\right|\right)n^{-\Delta i}\Big] \\
&\le \frac{1}{k}\Big[\theta(\varepsilon_{k-1}+\varepsilon_{k-2}) + \sum_{i=0}^{k-2}\left(\varepsilon_i + |V'_i|\right)n^{-\Delta(k-i)}\Big].
\end{aligned}$$

We prove the claim by considering two cases $k \le \frac{\ln n}{\ln\ln n}$ and $k > \frac{\ln n}{\ln\ln n}$.

We first apply induction for $k \le \frac{\ln n}{\ln\ln n}$. The base cases for $k = 0, 1$ holds simply because $\varepsilon_0 = \varepsilon_1 \equiv 0$. Assume $\varepsilon_j < n^{-\nu}j^{-\nu j} < 1$ holds for any $j < k$, by the uniform upper bound on V'_k proven in Lemma 6.4 and Eq. (7.34), we have

$$\begin{aligned}
\varepsilon_k \le &\frac{1}{k}\Big[\theta(\varepsilon_{k-1}+\varepsilon_{k-2}) + \sum_{i=2}^{k}\left(\varepsilon_{k-i} + e^{2\theta^2}\right)n^{-\Delta i}\Big] \\
\le &\frac{1}{k}\Big[\theta(\varepsilon_{k-1}+\varepsilon_{k-2}) + 2\sum_{i=2}^{k} e^{2\theta^2}n^{-\Delta i}\Big] \\
\le &\theta(\varepsilon_{k-1}+\varepsilon_{k-2}) + 2\,e^{2\theta^2}n^{-2\Delta}.
\end{aligned}$$

Define $\theta' = 3\,e^{2\theta^2}n^{-2\Delta}$. The above equation can be relaxed as

$$\varepsilon_k \le \theta(\varepsilon_{k-1}+\varepsilon_{k-2}) + \theta'.$$

Using an induction on k, it is easy to see that $\varepsilon_k \le \theta^k\theta' 3^k$ since $\theta(n) > 1$ for large n. That is, for large n,

$$\varepsilon_k \le 3^{k+1}\theta^k e^{2\theta^2}n^{-2\Delta} < n^{-\nu}k^{-\nu k}$$

holds since $\nu < \Delta$ as in Eq. (7.32).

Now consider the case when $k \ge \frac{\ln n}{\ln\ln n}$, and we will prove by another induction on k. The base case for $k = \frac{\ln n}{\ln\ln n}$ is proven in the previous case. Assume $\varepsilon_j \le n^{-\nu}j^{-\nu j}$ holds for all $j < k$, we then prove the bound for $j = k$ as follows.

First, we bound the summation

$$A \triangleq \sum_{j=0}^{k-2}\left|V'_j\right|n^{-\Delta(k-j)}$$

as

$$A \le n^{\tau}\sum_{j=0}^{k-2} j^{-\frac{j}{4}}n^{-\Delta(k-j)} \le n^{\tau-\Delta k}\sum_{j=0}^{k-2} j^{-\frac{j}{4}}n^{\Delta j}.$$

Define function $\psi(x) = x^{-\frac{x}{4}}n^{\Delta x}$. The above equation can be written as

$$(7.35)\qquad A \le n^{\tau-\Delta k}\sum_{j=0}^{k-2}\psi(j).$$

Computing the derivative of $\ln\psi(x)$, it is easy to see that $\psi(x)$ is increasing for $x \in \left[0, \frac{n^{4\Delta}}{e}\right]$ and decreasing for $x \in \left[\frac{n^{4\Delta}}{e}, \infty\right)$. If $k - 2 \le \frac{n^{4\Delta}}{e}$, Eq. (7.35) can be bounded as

$$\begin{aligned}
A \le &n^{\tau-\Delta k}\,(k-1)\,\psi(k-2) \\
\le &n^{\tau-k\Delta}\,(k-1)(k-2)^{-\frac{k-2}{4}}n^{(k-2)\Delta} \\
\le &n^{\tau-2\Delta}\,k^{-\frac{k}{8}} \\
\le &n^{2\tau-2\Delta}\,k^{-\frac{k}{8}},
\end{aligned}$$

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

where the third step holds for sufficiently large n since $k \geq \frac{\ln n}{\ln \ln n}$. Similarly, if $k-2 > \frac{n^{4\Delta}}{e}$, Eq. (7.35) can be bounded as

$$\begin{aligned} A \leq & n^{\tau-\Delta k}(k-1)\psi\left(\frac{n^{4\Delta}}{e}\right) \\ = & n^{\tau-k\Delta}(k-1)e^{\frac{n^{4\Delta}}{4e}} \\ \leq & n^{\tau-2\Delta} n^{-(k-2)\Delta}(k-1)e^{\frac{k-2}{4}} \\ \leq & n^{\tau-2\Delta} k^{-(k-2)\Delta}(k-1)e^{\frac{k-2}{4}} \\ \leq & n^{2\tau-2\Delta} k^{-\frac{k}{8}}, \end{aligned}$$

where the third step holds for the condition $k-2 > \frac{n^{4\Delta}}{e}$, and the last step holds for sufficiently large n since $\Delta > \frac{1}{8}$ as in Eq. (7.32). Combining the above two cases, the following inequality holds

$$A \leq n^{2\tau-2\Delta}k^{-\frac{k}{8}}.$$

Equation (7.34) then implies that

(7.36)
$$\varepsilon_k \leq \frac{1}{k}\left[\theta(\varepsilon_{k-1}+\varepsilon_{k-2})+\sum_{j=0}^{k-2}\varepsilon_j n^{-\Delta(k-j)}+n^{2\tau-2\Delta}k^{-\frac{k}{8}}\right].$$

We bound each term of the summation as follows.

- First,

(7.37)
$$\begin{aligned} & \theta(\varepsilon_{k-1}+\varepsilon_{k-2}) \\ \leq & 2\theta\, n^{-\nu}(k-2)^{-\nu(k-2)} \\ = & \frac{1}{2}k\, n^{-\nu}k^{-\nu k}\cdot 4\theta k^{2\nu-1}\left(1+\frac{2}{k-2}\right)^{\nu(k-2)} \\ \leq & \frac{1}{2}k\, n^{-\nu}k^{-\nu k}\cdot 4\theta k^{2\nu-1}e^{2\nu} \\ \leq & \frac{1}{2}k\, n^{-\nu}k^{-\nu k}, \end{aligned}$$

  where the final step holds for sufficiently large n since $\nu < \frac{1}{8} < \frac{1}{2}$ as in assumption.

- Second, since

$$\begin{aligned} \frac{n^{-\nu-\Delta(k-j)}j^{-\nu j}}{n^{-\nu-\Delta(k-j-1)}(j+1)^{-\nu(j+1)}} &= (j+1)^{\nu}\frac{\left(1+\frac{1}{j}\right)^{\nu j}}{n^{\Delta}} \\ &\leq \frac{(j+1)^{\nu}e^{\nu}}{n^{\Delta}} < \frac{1}{2}, \end{aligned}$$

  for sufficiently large n and the choice $\nu < \Delta$ in Eq. (7.32), we have

$$\begin{aligned} \sum_{j=0}^{k-2}\varepsilon_j n^{-\Delta(k-j)} &\leq \sum_{j=0}^{k-2} n^{-\nu-\Delta(k-j)}j^{-\nu j} \\ &\leq 2n^{-\nu-2\Delta}(k-2)^{-\nu(k-2)}. \end{aligned}$$

  For sufficiently large n, we have

$$\sum_{j=0}^{k-2}\varepsilon_j n^{-\Delta(k-j)} \leq \frac{1}{4}k\, n^{-\nu}k^{-\nu k}.$$

  Here we use the fact that $\nu, \Delta > 0$.

- Lastly, for large n,

$$n^{2\tau-2\Delta}k^{-\frac{k}{8}} \leq \frac{1}{4}k\, n^{-\nu}k^{-\nu k}$$

  holds because $2\Delta > 2\tau+\nu$ in Eq. (7.32) and $\nu < \frac{1}{8}$ in the statement.

Adding the bounds in the above three cases, we have

$$\varepsilon_k \leq n^{-\nu}k^{-\nu k}.$$

Then the lemma follows. □

Then we use the above bound to prove Eq. (3.13).

Lemma 7.2. *With parameters satisfying Eq. (3.4), with probability* $1-o(1)$,

$$\left|\sum_{k=0}^{t}V_k z^k - \sum_{k=0}^{t}V'_k z^k\right| = \mathcal{O}(n^{c-\nu}).$$

Proof. Let $M = \frac{\ln n}{\ln \ln n}$. Using Lemma 7.1, with probability $1-o(1)$,

(7.38)
$$\begin{aligned} \left|\sum_{k=0}^{t}\varepsilon_k z^k\right| \leq & \sum_{k=0}^{M}\varepsilon_k |z|^k + \sum_{k=M+1}^{t}\varepsilon_k |z|^k \\ \leq & n^{-\nu}\sum_{k=0}^{M}|z|^k + n^{-\nu}\sum_{k=M+1}^{t}k^{-\nu k}|z|^k. \end{aligned}$$

For large n,

$$\sum_{k=0}^{M}|z|^k \leq 2(\ln n)^{cM} \leq 2n^c.$$

On the other hand, since $M = \frac{\ln n}{\ln \ln n}$ and $c < \nu$ as in Eq. (3.4), it holds for large n that

$$\begin{aligned} \frac{(k+1)^{-\nu(k+1)}|z|^{k+1}}{k^{-\nu k}|z|^k} &= (k+1)^{-\nu}|z|\left(1+\frac{1}{k}\right)^{-\nu k} \\ &\leq \frac{|z|}{(M+1)^{\nu}} < \frac{1}{2}. \end{aligned}$$

Thus, for large n,

(7.39)
$$\sum_{k=M+1}^{t}k^{-\nu k}|z|^k \leq M^{-\nu M}|z|^M \leq n^c,$$

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited

which means

$$\sum_{k=0}^{t} \varepsilon_k \, |z|^k = \mathcal{O}(n^{c-\nu}). \tag{7.40}$$

□

References

[AA13] Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. *Theory OF Computing*, 9(4):143–252, 2013. 1, 1.1

[AAB+19] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, and Joseph C. Bardin et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574:505–510, 2019. 1, 1.1

[Bar16] Alexander Barvinok. Computing the permanent of (some) complex matrices. *Foundations of Computational Mathematics*, 16(2):329–342, 2016. 1

[Bar19] Alexander Barvinok. Computing permanents of complex diagonally dominant matrices and tensors. *Israel Journal of Mathematics*, 232(2):931–945, 2019. 1

[BCS13] Peter Bürgisser, Michael Clausen, and Mohammad A Shokrollahi. *Algebraic complexity theory*, volume 315. Springer Science & Business Media, 2013. 1

[BFNV18] Adam Bouland, Bill Fefferman, Chinmay Nirkhe, and Umesh Vazirani. Quantum supremacy and the complexity of random circuit sampling. arXiv:1803.04402, 2018. 1.1

[BJS10] Michael J. Bremner, Richard Jozsa, and Dan J. Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proceedings of the Royal Society A*, 467, 2010. 1.1

[CCL10] Jin-Yi Cai, Xi Chen, and Dong Li. Quadratic lower bound for permanent vs. determinant in any characteristic. *computational complexity*, 19(1):37–56, 2010. 1

[CPS99] Jin-Yi Cai, Aduri Pavan, and D. Sivakumar. On the hardness of permanent. In *STACS 99, 16th Annual Symposium on Theoretical Aspects of Computer Science, Trier, Germany, March 4-6, 1999, Proceedings*, pages 90–99, 1999. 1

[Edm65] Jack Edmonds. Maximum matching and a polyhedron with 0, 1-vertices. *Journal of research of the National Bureau of Standards B*, 69(125-130):55–56, 1965. 1

[EM18] Lior Eldar and Saeed Mehraban. Approximating the permanent of a random matrix with vanishing mean. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 23–34. IEEE, 2018. 1, 1.1, 2, 2

[FFK94] Stephen A. Fenner, Lance J. Fortnow, and Stuart A. Kurtz. Gap-definable counting classes. *Journal of Computer and System Sciences*, 48(1):116 – 148, 1994. 1

[JKP19] Matthew Jenssen, Peter Keevash, and Will Perkins. Algorithms for #BIS-hard problems on expander graphs. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2235–2247. Society for Industrial and Applied Mathematics, 2019. 1

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM (JACM)*, 51(4):671–697, 2004. 1

[LLLM19] Chao Liao, Jiabao Lin, Pinyan Lu, and Zhenyu Mao. Counting independent sets and colorings on random regular bipartite graphs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2019, September 20-22, 2019, Massachusetts Institute of Technology, Cambridge, MA, USA.*, pages 34:1–34:12, 2019. 1

[MB19] Ryan L. Mann and Michael J. Bremner. Approximation Algorithms for Complex-Valued Ising Models on Bounded Degree Graphs. *Quantum*, 3:162, July 2019. 1

[MR04] Thierry Mignon and Nicolas Ressayre. A quadratic bound for the determinant and permanent problem. *International Mathematics Research Notices*, 2004(79):4241–4253, 2004. 1

[RW04] Grzegorz A Rempała and Jacek Wesołowski. Limit theorems for random permanents with exchangeable structure. *Journal of multivariate analysis*, 91(2):224–239, 2004. 1.1, 2

[Val79a] Leslie G Valiant. Completeness classes in algebra. In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 249–261. ACM, 1979. 1

[Val79b] Leslie G Valiant. The complexity of computing the permanent. *Theoretical computer science*, 8(2):189–201, 1979. 1

[Val08] Leslie G Valiant. Holographic algorithms. *SIAM Journal on Computing*, 37(5):1565–1594, 2008. 1

[WQD+19] Hui Wang, Jian Qin, Xing Ding, Ming-Cheng Chen, Si Chen, Xiang You, Yu-Ming He, Xiao Jiang, Z. Wang, L. You, J. J. Renema, Sven Hoefling, Chao-Yang Lu, and Jian-Wei Pan. Boson sampling with 20 input photons in 60-mode interferometers at 1014 state spaces. arXiv:1910.09930, 2019. 1

Downloaded 01/01/22 to 143.110.155.208 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Copyright © 2021 by SIAM
Unauthorized reproduction of this article is prohibited