SINGULAR VALUES OF GAUSSIAN MATRICES AND PERMANENT ESTIMATORS

MARK RUDELSON AND OFER ZEITOUNI

ABSTRACT. We present estimates on the small singular values of a class of matrices with independent Gaussian entries and inhomogeneous variance profile, satisfying a strong-connectedness condition. Using these estimates and concentration of measure for the spectrum of Gaussian matrices with independent entries, we prove that for a large class of graphs satisfying an appropriate expansion property, the Barvinok–Godsil-Gutman estimator for the permanent achieves sub-exponential errors with high probability.

1. INTRODUCTION

Recall that the permanent of an n-by-n matrix A is defined as

$$\mathrm{per}(A) = \sum_{\pi \in \mathcal{S}_n} a_{1,\pi(1)} a_{2,\pi(2)} \cdots a_{n,\pi(n)} \,,$$

where the summation is over all permutations of n elements. In this paper we consider only matrices A with non-negative entries. This includes in particular matrices with 0–1 entries, for which the evaluation of the permanent is fundamental in combinatorial counting problems. For general 0–1 matrices, the evaluation of the permanent is a $\#P$-complete problem [19]. Thus, the interest is in obtaining algorithms that compute approximations to the permanent, and indeed a polynomial running time Markov Chain Monte Carlo randomized algorithm that evaluates $\mathrm{per}(A)$ (up to $(1+\epsilon)$ multiplicative errors, with complexity polynomial in ϵ) is available [10]. In practice, however, the running time of such an algorithm, which is $O(n^{10})$, still makes it challenging to implement for large n. (An alternative, faster MCMC algorithm is presented in [3], with claimed running time of $O(n^7(\log n)^4)$.)

Date: January 26, 2013. Revised December 24, 2013.

M.R.: Department of Mathematics, University of Michigan. Partially supported by NSF grant DMS 1161372.

O. Z.: Faculty of Mathematics, Weizmann Institute and Courant Institute, New York University. Partially supported by NSF grant DMS 1106627 and a grant from the Israel Science Foundation.

An earlier simple probabilistic algorithm for the evaluation of $\mathrm{per}(A)$ is based on the following observation: if $x_{i,j}$ are i.i.d. zero mean variables with unit variance and X in an $n \times n$ matrix with entries $x_{i,j}$, then an easy computation shows that

$$\mathrm{per}(A) = \mathbb{E}(\det(A_{1/2} \odot X))^2 \,, \tag{1.1}$$

where for any two $n \times m$ matrices A, B, $D = A \odot B$ denotes their Hadamard, or Schur, product, i.e. the $n \times m$ matrix with entries $d_{i,j} = a_{i,j} \cdot b_{i,j}$, and where $A_{1/2}(i,j) = A(i,j)^{1/2}$. Thus, $\det(A_{1/2} \odot X)^2$ is an unbiased estimator of $\mathrm{per}(A)$. This algorithm was proposed (with $x_{i,j} \in \{-1, 1\}$) in [7], and takes advantage of the fact that the evaluation of determinants is computationally easy via Gaussian elimination. While we do not discuss computational issues in this article, we note that the evaluation of the determinant requires at most $o(n^3)$ arithmetic operations; in terms of bit complexity, for matrices with integer entries of k bits, there exist algorithms with complexity $O(n^\alpha k^{1+o(1)})$, with $\alpha < 3$, see e.g. [11] for a review and the value $\alpha \sim 2.7$. To avoid rounding errors in the case of real valued random variables one needs to take $k = n^{1+o(1)}$, yielding a total bit-complexity in that case smaller than $o(n^4)$.

Thus, the main question concerning the above algorithm is the approximation error, and in particular the concentration of the random variable $\det^2(A_{1/2} \odot X)$ around its mean. For general matrices A with non-negative entries, Barvinok showed that using standard Gaussian variables $x_{i,j}$, with probability approaching one, the resulting multiplicative error is at most exponential in n, with sharp constant. (The constant cannot be improved, as the example of A being the identity matrix shows.)

For restricted classes of matrices, better performance is possible. Thus, in [6], the authors analyzed a variant of the Godsil-Gutman algorithm due to [12] and showed that for certain dense, *random* $0-1$ matrices, a multiplicative $(1+\epsilon)$ error is achieved in time $O(n\omega(n)\epsilon^{-2})$. In [5], it is shown that for a restricted class of non-random matrices, the performance achieved by the Barvinok–Godsil-Gutman estimator is better than in the worst-case scenario. Indeed, if for some fixed constants $\alpha, \beta > 0$ one has $a_{i,j} \in [\alpha, \beta]$, then for any $\delta > 0$, with G denoting the standard Gaussian matrix,

$$\mathbb{P}\left(\frac{1}{n}\left|\log \frac{\det(A_{1/2} \odot G)^2}{\mathrm{per}(A)}\right| > \delta\right) \to_{n\to\infty} 0 \,,$$

uniformly in A; that is, for such matrices this estimator achieves subexponentional (in n) errors, with $o(n^3)$ (arithmetic) running time. An

improved analysis in presented in [4], where it is shown that the approximation error in the same set of matrices is only exponential in $n^{2/3}\log n$.

The class of matrices considered in [5] is somewhat restricted - first, it does not include incidence matrices of non-trivial graphs, and second, for such matrices, as noted in [5], a polynomial error deterministic algorithm with running time $O(n^4)$ is available by adapting the algorithm in [14]. Our goal in this paper is to better understand the properties of the Barvinok–Godsil-Gutman estimator, and show that in fact the same analysis applies for a class of matrices that arise from (δ,κ)-*strongly connected* graphs, i.e. graphs with good expansion properties (see Definition 2.1 for a precise definition). Our first main result concerning permanent estimators reads as follows.

Theorem 1.1. *There exist C, C', c depending only on δ,κ such that for any $\tau \geq 1$ and any adjacency matrix A of a (δ,κ)-strongly connected graph,*

$$\mathbb{P}\left(\left|\log\det^2(A_{1/2}\odot G) - \mathbb{E}\log\det^2(A_{1/2}\odot G)\right| > C(\tau n\log n)^{1/3}\right) \leq \exp(-\tau) + \exp\left(-c\sqrt{n}/\log n\right). \tag{1.2}$$

and

$$\mathbb{E}\log\det^2(A_{1/2}\odot G) \leq \log\mathrm{per}(A) \leq \mathbb{E}\log\det^2(A_{1/2}\odot G) + C'\sqrt{n\log n}.$$

For a more refined probability bound see Theorem 7.1. Combining the two inequalities of Theorem 1.1, we obtain the concentration of the Barvinok-Godsil-Gutman estimator around the permanent.

Corollary 1.2. *Under the assumptions of Theorem 1.1,*

$$\mathbb{P}\left(\left|\log\frac{\det^2(A_{1/2}\odot G)}{\mathrm{per}(A)}\right| > 2C'\sqrt{n\log n}\right) \leq \exp\left(-c\sqrt{n}/\log n\right).$$

This corollary implies the uniform convergence in probability if we consider a family of (δ,κ)-strongly connected $n\times n$ bipartite graphs with $n\to\infty$.

Corollary 1.3. *Let $SC_{\delta,\kappa,n}$ denote the collection of adjacency matrices of (δ,κ)-strongly connected $n\times n$ bipartite graphs. Let $\{\tau_n\}_{n=1}^{\infty}$ be a sequence of positive numbers such that $\tau_n\to\infty$. Set $\mathtt{s}_n = \tau_n\sqrt{n\log n}$. Then for any $\varepsilon>0$,*

$$\lim_{n\to\infty}\sup_{A\in \mathrm{SC}_{\delta,\kappa,n}} \mathbb{P}\left(\frac{1}{\mathtt{s}_n}\left|\log\frac{\det^2(A_{1/2}\odot G)}{\mathrm{per}(A)}\right| > \varepsilon\right) = 0. \tag{1.3}$$

We remark that the error estimate (1.3) in Corollary 1.3 is probably not optimal. Indeed, in the special case $A_{i,j} \equiv 1$, a consequence of the distributional results concerning matrices with i.i.d. Gaussian entries [20], see also [8], is that (1.3) holds with $\mathtt{s}_n$ satisfying $\mathtt{s}_n/\log n \to \infty$. As Theorem 1.1 shows, the main source of error is the discrepancy between $\mathbb{E}\log\det^2(A_{1/2} \odot G)$ and $\log \mathbb{E}\det^2(A_{1/2} \odot G)$.

Our second main result pertains to graphs whose adjacency matrix A satisfies $\mathrm{per}(A) > 0$. For such matrices, there exists a (polynomial time) scaling algorithm that transforms A into an (almost) doubly stochastic matrix, see [14, Pages 552-553]. In particular, there exists a deterministic algorithm (with running time $O(n^4)$) that outputs non-negative diagonal matrices D_1, D_2 so that $B = D_1 A D_2$ is an approximately doubly stochastic matrix, i.e. $\sum_i B_{i,j} \in [1/2, 2], \sum_j B_{i,j} \in [1/2, 2]$. (Much more can be achieved, but we do not use that fact.) Since $\mathrm{per}(A) = \mathrm{per}(B) \cdot \prod_i (D_1(i,i) D_2(i,i))$, evaluating $\mathrm{per}(A)$ thus reduces to the evaluation of $\mathrm{per}(B)$. The properties of the Barvinok–Godsil-Gutman estimator for such matrices are given in the following theorem.

Theorem 1.4. *Let $r > 0$. There exist c, C, C' depending only on r, δ, κ with the following property. Let $0 \le b_n \le n$ be a given sequence. Let B be an $n \times n$ matrix with entries $0 \le b_{i,j} \le b_n/n$ such that*

$$\sum_{i=1}^{n} b_{i,j} \le 1 \qquad \text{for all } j \in [n];$$

$$\sum_{j=1}^{n} b_{i,j} \le 1 \qquad \text{for all } i \in [n].$$

Define the bipartite graph $\Gamma = \Gamma_B$ connecting the vertices i and j whenever $b_{i,j} \ge r/n$, and assume that Γ is (δ, κ)-strongly connected. Then for any $\tau \ge 1$

$$\mathbb{P}\left(\left| \log\det^2(B_{1/2} \odot G) - \mathbb{E}\log\det^2(B_{1/2} \odot G) \right| > C(\tau b_n n)^{1/3} \log^{c'} n \right)$$

(1.4)

$$\le \exp(-\tau) + \exp\left(-c\sqrt{n}/\log^c n\right)$$

and

(1.5)

$$\mathbb{E}\log\det^2(B_{1/2} \odot G) \le \log \mathrm{per}(B) \le \mathbb{E}\log\det^2(B_{1/2} \odot G) + C'\sqrt{b_n n}\log^{c'} n.$$

As in Theorem 1.1, we can derive the concentration around the permanent and the uniform convergence in probability.

Corollary 1.5. *Under the assumptions of Theorem 1.4,*

$$\mathbb{P}\left(\left|\log\frac{\det^2(B_{1/2}\odot G)}{\mathrm{per}(B)}\right| > 2C'\sqrt{b_n n}\log^{c'} n\right) \le \exp\left(-c\sqrt{n}/\log^{c} n\right).$$

Corollary 1.6. *Let $GSC_{c,\delta,\kappa,n}$ denote the collection of $n\times n$ matrices B with properties as in Theorem 1.4. Then there exists a constant $\bar{C}=\bar{C}(c,\delta,\kappa)$ so that with $\mathtt{s}_n=(nb_n\log^{\bar{C}} n)^{1/2}$, and any $\varepsilon>0$,*

$$\lim_{n\to\infty}\sup_{B\in \mathrm{GSC}_{c,\delta,\kappa,n}}\mathbb{P}\left(\frac{1}{\mathtt{s}_n}\left|\log\frac{\det^2(B_{1/2}\odot G)}{\mathrm{per}(B)}\right|>\varepsilon\right)=0.$$

Corollary 1.5 applies, in particular, to approximately doubly stochastic matrices B whose entries satisfy $c/n\le b_{i,j}\le 1$ for all i,j. For such matrices the graph Γ_A is complete, so the strong connectedness condition is trivially satisfied. Note that if such matrix contains entries of order $\Omega(1)$, then the algorithm of [14] estimates the permanent with an error exponential in n. In this case, $b_n=\Omega(n)$, and Corollary 1.5 is weaker than Barvinok's theorem in [2]. This is due to the fact that we do not have a good bound for the gap between $\mathbb{E}\log\det^2(B_{1/2}\odot G)$ and $\log\mathrm{per}(B)$, see (1.5). However, this bound cannot be significantly improved in general, even for well-connected matrices. As we show in Lemma 7.3, the gap between these values is of order $\Omega(n)$ for a matrix with all diagonal entries equal 1 and all off-diagonal entries equal c/n. For such a matrix, the Barvinok–Godsil-Gutman estimator will fail consistently, i.e., it will be concentrated around a value, which is $\exp(cn)$ far away from the permanent. Thus, we conclude that for almost doubly stochastic matrices with a strongly connected graph the Barvinok–Godsil-Gutman estimator either approximates the permanent up to $\exp(o(n))$ with high probability, or yields an exponentially big error with high probability.

As in [5], Theorems 1.3 and 1.6 depend on concentration of linear statistics of the spectrum of random (inhomogeneous) Gaussian matrices; this in turn require a good control on small singular values of such matrices. Thus, the first part of the current paper deals with the latter question, and proceeds as follows. In Section 2 we define the notion of strongly connected bipartite graphs, and state our main results concerning small singular values of Gaussian matrices, Theorems 2.3 and 2.4; we also state applications of the latter theorems to both adjacency graphs and to "almost" doubly stochastic matrices, see Theorems 2.5 and 2.7. Section 3 is devoted to several preliminary lemmas involving ε-net arguments. In Section 4 we recall the notion of compressible vectors and obtain estimate on the norm of Gaussian matrices restricted to compressible vectors. The control of the minimal

singular value (that necessitates the study of incompressible vectors) is obtained in Section 5, while Section 6 is devoted to the study of intermediate singular values. In Section 7, we retun to the analysis of the Barvinok–Godsil-Gutman estmator, and use the control on singular values together with an improved (compared to [5]) use of concentration inequalities to prove the applications and the main theorems in the introduction.

Acknowledgment We thank A. Barvinok and A. Samorodnitsky for sharing with us their knowledge of permanent approximation algorithms, and for useful suggestions. We also thank U. Feige for a useful suggestion.

2. Definitions and results

For a matrix A we denote its operator norm by $\|A\|$, and set $\|A\|_\infty = \max |a_{i,j}|$. By $[n]$ we denote the set $\{1, \dots, n\}$. By $\lfloor t \rfloor$ we denote the integer part of t.

Let $J \subset [m]$. Denote by $\mathbb{R}^J$ and S^J the coordinate subspace of $\mathbb{R}^m$ corresponding to J and its unit sphere.

For a left vertex $j \in [m]$ and a right vertex $i \in [n]$ of a bipartite graph $\Gamma = ([m], [n], E)$ we write $j \to i$ if j is connected to i.

Definition 2.1. Let $\delta, \kappa > 0$, $\delta/2 > \kappa$. Let Γ be an $m \times n$ bipartite graph. We will say that Γ is (δ, κ)-strongly connected if

(1) $\deg(i) \geq \delta m$ for all $i \in [n]$;
(2) $\deg(j) \geq \delta n$ for all $j \in [m]$;
(3) for any set $J \subset [m]$ the set of its strongly connected neighbors

$$I(J) = \{i \in [n] \mid j \to i \text{ for at least } \lfloor (\delta/2) \cdot |J| \rfloor \text{ numbers } j \in J\}$$

has cardinality $|I(J)| \geq \min\big((1+\kappa)|J|, n\big)$.

We fix the numbers δ, κ and call such graph strongly connected. Property (3) in this definition is similar to the expansion property of the graph. In the argument below we denote by C, c, etc. constants depending on the parameters δ, κ and r appearing in Theorems 2.3 and 2.4. The values of these constants may change from line to line.

Although condition (3) is formulated for all sets $J \subset [m]$, it is enough to check it only for sets with cardinality $|J| \leq (1 - \delta/2)m$. Indeed, if $|J| > (1 - \delta/2)m$, then any $i \in [n]$ is strongly connected to J.

Definition 2.2. Let A be an $m \times n$ matrix. Define the graph $\Gamma_A = ([m], [n], E)$ by setting $j \to i$ whenever $a_{j,i} \neq 0$.

We will prove two theorems bounding the singular values of a matrix with normal entries. In the theorems, we allow for non-centered entries

because it will be useful for the application of the theorem in the proof of Theorem 2.7

Theorem 2.3. *Let W be an $n \times n$ matrix with independent normal entries $w_{i,j} \sim N(b_{i,j}, a_{i,j}^2)$. Assume that*

(1) *$a_{i,j} \in \{0\} \cup [r, 1]$ for some constant $r > 0$ and all i, j;*
(2) *the graph Γ_A is strongly connected;*
(3) *$\|\mathbb{E}W\| \le K\sqrt{n}$ for some $K \ge 1$.*

Then for any $t > 0$

$$\mathbb{P}\left(s_n(W) \le ctK^{-C}n^{-1/2}\right) \le t + e^{-c'n}.$$

Theorem 2.4. *Let $n/2 < m \le n - 4$, and let W be an $n \times m$ matrix with independent normal entries $w_{i,j} \sim N(b_{i,j}, a_{i,j}^2)$. Assume that*

(1) *$a_{i,j} \in \{0\} \cup [r, 1]$ for some constant $r > 0$ and all i, j;*
(2) *the graph Γ_A is strongly connected;*
(3) *$\|\mathbb{E}W\| \le K\sqrt{n}$.*

Then for any $t > 0$

$$\mathbb{P}\left(s_m(W) \le ctK^{-C} \cdot \frac{n-m}{\sqrt{n}}\right) \le t^{(n-m)/4} + e^{-c'n}.$$

In Theorems 2.3, 2.4 we assume that the graph Γ_A is strongly connected. This condition can be relaxed. In fact, property (3) in the definition of strong connectedness is used only for sets J of cardinality $|J| \ge (r^2\delta/6)m$ (see Lemmas 4.1 and 4.2 for details).

We apply Theorems 2.3 and 2.4 to two types of matrices. Consider first the situation when the matrix A is an adjacency matrix of a graph, and $\mathbb{E}W = 0$.

Theorem 2.5. *Let Γ be a strongly connected $n \times n$ bipartite graph, and let A be its adjacency matrix. Let G be the $n \times n$ standard Gaussian matrix. Then for any $t > 0$*

$$\mathbb{P}\left(s_n(A \odot G) \le ctn^{-1/2}\right) \le t + e^{-c'n},$$

and for any $n/2 < m < n - 4$

$$\mathbb{P}\left(s_m(A \odot G) \le ct \cdot \frac{n-m}{\sqrt{n}}\right) \le t^{(n-m)/4} + e^{-c'n}.$$

Theorem 2.5 is also applicable to the case when Γ is an unoriented graph with n vertices. In this case we denote by A its adjacency matrix, and assume that the graph Γ_A is strongly connected.

Remark 2.6. With some additional effort the bound $m < n - 4$ in Theorem 2.5 can be eliminated, and the term $t^{(n-m)/4}$ in the right hand side can be replaced with t^{n-m+1}.

The second application pertains to "almost" doubly stochastic matrices, i.e. matrices with uniformly bounded norms of rows and columns.

Theorem 2.7. *Let W be an $n \times n$ matrix with independent normal entries $w_{i,j} \sim N(0, a_{i,j}^2)$. Assume that the matrix of variances $(a_{i,j}^2)_{i,j=1}^n$ satisfies the conditions*

(1) $\sum_{i=1}^n a_{i,j}^2 \le C$ *for any* $j \in [n]$, *and*
(2) $\sum_{j=1}^n a_{i,j}^2 \le C$ *for any* $i \in [n]$.

Consider an $n \times n$ bipartite graph Γ defined as follows:

$$i \to j, \text{ whenever } \frac{c}{n} \le a_{i,j}^2,$$

and assume that Γ is strongly connected. Then for any $t > 0$

$$\mathbb{P}\left(s_n(W) \le ctn^{-1}\log^{-C'} n\right) \le t + \exp(-C\log^4 n),$$

and for any $n/2 < m < n-4$

$$\mathbb{P}\left(s_m(W) \le ct \cdot \frac{n-m}{n\log^{C'} n}\right) \le t^{(n-m)/4} + \exp(-C\log^4 n).$$

Note that the condition on the variance matrix in Theorem 2.7 does not exclude the situation where several of its entries $a_{i,j}^2$ are of the order $\Omega(1)$. Also, $\exp(-C\log^4 n)$ in the probability estimate can be replaced by $\exp(-C\log^p n)$ for any p. Of course, the constants C, C', c would then depend on p.

3. Matrix norms and the ε-net argument

We prepare in this section some preliminary estimates that will be useful in bounding probabilities by ε-net arguments. First, we have the following bound on the norm of a random matrix as an operator acting between subspaces of $\mathbb{R}^n$. This will be useful in the proof of Theorem 2.4.

Lemma 3.1. *Let A be an $n \times n$ matrix with $\|A\|_\infty \le 1$, and let G be an $n \times n$ standard Gaussian matrix. Then for any subspaces $E, F \subset \mathbb{R}^n$ and any $s \ge 1$,*

$$\mathbb{P}\left(\|P_F(A \odot G) : E \to \mathbb{R}^n\| \ge cs(\sqrt{\dim E} + \sqrt{\dim F})\right) \le \exp(-Cs^2(\dim E + \dim F)),$$

where P_F is the orthogonal projection onto F.

Proof. When $a_{i,j} \equiv 1$, the lemma is a direct consequence of the rotational invariance of the Gaussian measure, and standard concentration estimates for the top singular value of a Wishart matrix [17, Proposition 2.3]. For general A satisfying the assumptions of the lemma,

the claim follows from the contraction argument in e.g. [18, Lemma 2.7], since the collection of entries $\{g_{i,j}\}$ so that $\|A \odot G : E \to F\| \leq cs(\sqrt{\dim E} + \sqrt{\dim F}))$ is a convex symmetric set. We give an alternative direct proof: let $A'_{i,j} = \sqrt{1 - A_{i,j}^2}$, and note that G equals in distribution $A \odot G_1 + A' \odot G_2$ where G_1, G_2 are independent copies of G. On the event

$$\mathcal{A}_1 := \left\{ \|P_F(A \odot G_1) : E \to \mathbb{R}^n\| \geq cs(\sqrt{\dim E} + \sqrt{\dim F}) \right\},$$

there exist unit vectors $v_{G_1} \in F, w_{G_1} \in E$ so that $|v_{G_1}^T A \odot G_1 w_{G_1}| \geq cs(\sqrt{\dim E} + \sqrt{\dim F})$. On the other hand, for any fixed v, w, $v^T A' \odot G_2 w$ is a Gaussian variable of variance bounded by 1, and hence the event

$$\mathcal{A}_2(v, w) := \left\{ |v^T A' \odot G_2 w| \geq cs(\sqrt{\dim E} + \sqrt{\dim F})/2 \right\}$$

has probability bounded above by

$$\exp(-Cs^2(\sqrt{\dim E} + \sqrt{\dim F})^2) \leq \exp(-Cs^2(\dim E + \dim F)).$$

The proof is completed by noting that

$$\begin{aligned} \mathbb{P}(\mathcal{A}_1) \quad &\leq \quad \mathbb{E}\mathbb{P}(\mathcal{A}_2(v_{G_1}, w_{G_1}) \mid \mathcal{A}_1)) \\ &\qquad + \mathbb{P}(\|P_F G : E \to \mathbb{R}^n\| \geq cs(\sqrt{\dim E} + \sqrt{\dim F})/2). \end{aligned}$$

□

To prove Theorem 2.7 we will need an estimate of the norm of the matrix, which is based on a result of Riemer and Schütt [15].

Lemma 3.2. *Let A be an $n \times n$ matrix satisfying conditions (1) and (2) in Theorem 2.7. Then*

$$\mathbb{P}(\|A \odot G\| \geq C \log^2 n) \leq \exp(-C \log^4 n).$$

Proof. Write $X = A \odot G$. By [15, Theorem 1.2],

$$\mathbb{E}\,\|A \odot G\| \leq C(\log^{3/2} n)\mathbb{E}(\max_{i=1,\ldots,n} \|(X_{i,j})_{j=1}^n\|_2 + \|(X_{i,j})_{i=1}^n\|_2). \tag{3.1}$$

Set $\eta_i = \|(X_{i,j})_{j=1}^n\|_2$, $i = 1, \ldots, n$ and $\Delta_i = \sum_{j=1}^n a_{i,j}^2 \leq C$. Define $\beta_{i,j} = a_{i,j}^2/\Delta_i \leq 1$. For $\theta \leq 1/4C$ one has that

$$\log \mathbb{E}e^{\theta \eta_i^2} = -\frac{1}{2} \sum_{j=1}^n \log(1 - 2\beta_{i,j}\theta\Delta_i) \leq c\theta,$$

for some constant c depending only on C. In particular, the independent random variables η_i possess uniform (in i, θ, n) subgaussian tails, and therefore, $\mathbb{E} \max_{i=1,\ldots,n} \eta_i \leq c'(\log n)^{1/2}$. Arguing similarly

for $\mathbb{E}(\max_{i=1,\dots,n} \|(X_{i,j})_{i=1}^n\|_2))$ and substituting in (3.1), one concludes that

$$\mathbb{E}\,\|A \odot G\| \le C \log^2 n.$$

The lemma follows from the concentration for the Gaussian measure, since $F : \mathbb{R}^{n^2} \to \mathbb{R}$, $F(B) = \|A \odot B\|$ is a 1-Lipschitz function, see e.g. [13]. □

Throughout the proofs below we will repeatedly use the easiest form of the ε-net argument. For convenience, we will formulate it as a separate lemma.

Lemma 3.3. *Let V be a $n \times m$ random matrix. Let $\mathcal{L} \subset S^{m-1}$ be a set contained in an l-dimensional subspace of $\mathbb{R}^m$. Assume that there exists $\varepsilon > 0$ such that for any $x \in \mathcal{L}$*

$$\mathbb{P}\left(\|Vx\|_2 < \varepsilon\sqrt{n}\right) \le p.$$

Denote by $\mathcal{L}_\alpha$ the α-neighborhood of $\mathcal{L}$ in $\mathbb{R}^m$. Then

$$\mathbb{P}\left(\exists x \in \mathcal{L}_{\varepsilon/(4K)} :\ \|Vx\| < (\varepsilon/2)\cdot\sqrt{n} \text{ and } \|V\| \le K\sqrt{n}\right) \le \left(\frac{6K}{\varepsilon}\right)^l \cdot p.$$

Proof. Let $\mathcal{N} \subset \mathcal{L}$ be an $(\varepsilon/(4K))$-net in $\mathcal{L}$. By the volumetric estimate, we can choose $\mathcal{N}$ of cardinality

$$|\mathcal{N}| \le \left(\frac{6K}{\varepsilon}\right)^l.$$

Assume that there exists $y \in \mathcal{L}_{\varepsilon/(4K)}$ such that $\|Vy\|_2 < (\varepsilon/2)\cdot\sqrt{n}$. Choose $x \in \mathcal{N}$ for which $\|y - x\|_2 < \varepsilon/(2K)$. If $\|V\| \le K\sqrt{n}$, then

$$\|Vx\|_2 \le (\varepsilon/2)\sqrt{n} + \|V\| \cdot \frac{\varepsilon}{2K} \le \varepsilon\sqrt{n}.$$

Therefore, by the union bound,

$$\mathbb{P}\left(\exists y \in \mathcal{L}_{\varepsilon/(4K)} :\ \|Vy\| < (\varepsilon/2)\sqrt{n} \text{ and } \|V\| \le K\sqrt{n}\right)$$
$$\le \mathbb{P}\left(\exists x \in \mathcal{N} :\ \|Vx\| \le \varepsilon\sqrt{n}\right) \le \left(\frac{6K}{\varepsilon}\right)^l \cdot p.$$

□

4. Compressible vectors

As developed in detail in [16, 17], when estimating singular values it is necessary to handle separately the action of random matrices on compressible, e.g., close to sparse, vectors. We begin with a basic small ball estimate.

Lemma 4.1. *Let $m, n \in \mathbb{N}$. Let A, B be (possibly random) $n \times m$ matrices, and let $W = A \odot G + B$, where G is the $n \times m$ Gaussian matrix, independent of A, B. Assume that, a.s.,*

(1) *$a_{i,j} \in \{0\} \cup [r, 1]$ for some constant $r > 0$ and all i, j;*
(2) *the graph Γ_A satisfies $deg(j) \geq \delta n$ for all $j \in [m]$.*

Then for any $x \in S^{m-1}$, $z \in \mathbb{R}^n$ and for any $t > 0$

$$\mathbb{P}\left(\|Wx - z\|_2 \leq t\sqrt{n}\right) \leq (Ct)^{cn}.$$

Proof. Let $x \in S^{m-1}$. Set $I = \{i \in [n] \mid \sum_{j=1}^m a_{i,j}^2 x_j^2 \geq r^2\delta/2\}$. Let $\Gamma = \Gamma_{A^T}$ be the graph of the matrix A^T. The inequality

$$\sum_{i=1}^n \sum_{j=1}^m a_{i,j}^2 x_j^2 \geq \sum_{j=1}^m r^2 \deg_\Gamma(j) x_j^2 \geq r^2\delta n \sum_{j=1}^m x_j^2 = r^2\delta n$$

implies

$$\sum_{i \in I} \left(\sum_{j=1}^m a_{i,j}^2 x_j^2 \right) \geq \sum_{i=1}^n \sum_{j=1}^m a_{i,j}^2 x_j^2 - r^2\delta n/2 \geq r^2\delta n/2.$$

On the other hand, we have the reverse inequality

$$\sum_{i \in I} \left(\sum_{j=1}^m a_{i,j}^2 x_j^2 \right) \leq |I| \left(\sum_{j=1}^m x_j^2 \right) = |I|,$$

and so $|I| \geq r^2\delta n/2$.

For any $i \in I$ the independent normal random variables $w_i = \sum_{j=1}^m (a_{i,j} g_{i,j} + b_{i,j}) x_j$ have variances at least $r^2\delta/2$. Estimating the Gaussian measure of a ball by its Lebesgue measure, we get that for any $\tau > 0$

$$\begin{aligned} &\mathbb{P}\left(\|Wx - z\|_2^2 \leq \tau^2 (r^2\delta/2)^2 \cdot n\right) \\ &\qquad \leq \mathbb{P}\left(\sum_{i \in I} (w_i - z_i)^2 \leq \tau^2 (r^2\delta/2) \cdot |I| \right) \leq (C\tau)^{|I|}. \end{aligned}$$

Setting $t = \tau r^2\delta/2$ finishes the proof. □

We now introduce the notion of compressible and incompressible vectors. The compressible vectors will be easier to handle by an ε-net argument, keeping track of the degree of compressibility. This is the content of the next three lemmas in this section.

For $u, v < 1$ denote

$$\mathrm{Sparse}(u) = \{x \in S^{m-1} \mid |\mathrm{supp}(x)| \leq um\}.$$

and

$$\mathrm{Comp}(u,v) = \{x \in S^{m-1} \mid \exists y \in \mathrm{Sparse}(u),\ \|x-y\|_2 \le v\},$$
$$\mathrm{Incomp}(u,v) = S^{m-1} \setminus \mathrm{Comp}(u,v).$$

We employ the following strategy. In Lemma 4.2, we show that the matrix W is well invertible on the set of highly compressible vectors. Lemma 4.3 asserts that if the matrix is well invertible on the set of vectors with a certain degree of compressibility, then we can relax the compressibility assumption and show invertibility on a larger set of compressible vectors. Finally, in Lemma 4.4, we prove that the matrix W is well invertible on the set of all compressible vectors. This is done by using Lemma 4.2 for highly compressible vectors, and extending the set of vectors using Lemma 4.3 in finitely many steps. The number of these steps will be independent of the dimension.

Lemma 4.2. *Let $m, n \in \mathbb{N}$, $m \le (3/2)n$. Let A, B, W be $n \times m$ matrices satisfying the conditions of Lemma 4.1. Let $K \ge 1$. Then there exist constants c_0, c_1, c_2 such that, for any $z \in \mathbb{R}^n$,*

$$\mathbb{P}\Big(\exists x \in \mathrm{Comp}\left(c_0, c_1/K^2\right) : \|Wx - z\|_2 \le (c_1/K)\sqrt{n} \text{ and } \|W\| \le K\sqrt{n}\Big) \le e^{-c_2 n}.$$

Proof. Let c be the constant from Lemma 4.1. Without loss of generality, we may and will assume that $c < 1$. Let $t > 0$ be a number to be chosen later. For any set $J \subset [m]$ of cardinality $|J| = l = \lfloor cm/3 \rfloor$ Lemmas 4.1 and 3.3 imply

$$\mathbb{P}\left(\exists x \in (S^J)_{t/(4K)} :\ \|Wx\|_2 < (t/2)\sqrt{n} \text{ and } \|W\| \le K\sqrt{n}\right) \le \left(\frac{6K}{t}\right)^l \cdot (Ct)^{cn}.$$

(Recall that S^J is the unit sphere of the coordinate subspace of $\mathbb{R}^m$ corresponding to J.) Since $\mathrm{Comp}(c/3, t/(4K)) \subset \bigcup_{|J|=l} (S^J)_{t/(4K)}$, the union bound yields

$$\mathbb{P}\left(\exists x \in \mathrm{Comp}(c/3, t/(4K)) :\ \|Wx\| < (t/2)\sqrt{n} \text{ and } \|W\| \le K\sqrt{n}\right) \le \binom{m}{l} \cdot \left(\frac{6K}{t}\right)^l \cdot (Ct)^{cn} \le \left(\frac{CK}{t}\right)^{cm/3} \cdot (Ct)^{cn},$$

which does not exceed $e^{-cn/3}$ provided that $t = c''/K$ for an appropriately chosen $c'' > 0$. This proves the lemma if we set $c_0 = c/3$, $c_1 = c''/4$. □

Lemma 4.3. *Let $m, n \in \mathbb{N}$, $n \le 2m$. Let A, B be (possibly random) $n \times m$ matrices, and set $W = A \odot G + B$, where G is the standard $n \times m$ Gaussian matrix, independent of A, B. Assume that, a.s.,*

(1) *$a_{i,j} \in \{0\} \cup [r, 1]$ for some constant $r > 0$ and all i, j;*
(2) *the graph Γ_{A^T} is strongly connected.*

Then for any c_0 and any $u, v > 0$, such that $u \ge c_0$ and $(1+\kappa/2)u < 1$, and for any $z \in \mathbb{R}^n$

$$\begin{aligned}\mathbb{P}\,\big(&\exists x \in \mathrm{Comp}((1+\kappa/2)u, (v/K)^{C+1}) \setminus \mathrm{Comp}(u, v) : \\ &\|Wx - z\|_2 \le cv(v/K)^C\sqrt{n} \text{ and } \|W\| \le K\sqrt{n}\big) \\ &\le e^{-cn}.\end{aligned}$$

where $c = c(c_0, \kappa, \delta, r)$.

Proof. Let $S(u, v) = \mathrm{Sparse}((1+\kappa/2)u) \setminus \mathrm{Comp}(u, v)$. Fix any $x \in \mathbb{R}^n$ and denote by J the set of all coordinates $j \in [m]$ such that $|x_j| \ge v/\sqrt{m}$. For any $x \in S(u, v)$ $|J| \ge um$, since otherwise $x \in \mathrm{Comp}(u, v)$. Since the graph Γ_{A^T} is strongly connected, this implies that $|I(J)| \ge (1+\kappa)um$.

If $i \in I(J)$, then $w_i = \sum_{j=1}^m a_{i,j} g_{i,j} x_j$ is a centered normal random variable with variance

$$\sigma_i^2 = \sum_{j=1}^m a_{i,j}^2 x_j^2 \ge \frac{v^2}{m} \cdot \sum_{j \in J} a_{i,j}^2 \ge \frac{v^2}{m} \cdot r^2(\delta/2)|J| \ge v^2 r^2 u\delta/2.$$

Hence, for any $t > 0$,

$$\begin{aligned}&\mathbb{P}\left(\|Wx - z\|_2 \le tvru \cdot \sqrt{\delta n}\right) \le \mathbb{P}\left(\|Wx - z\|_2 \le tvru \cdot \sqrt{\delta m/2}\right) \\ &\le \mathbb{P}\left(\sum_{i \in I(J)} (w_i - z_i)^2 \le t^2 v^2 r^2 u(\delta/2) \cdot |I(J)|\right) \le (ct)^{|I(J)|} \le (ct)^{(1+\kappa)um},\end{aligned}$$

where the third inequality is obtained by the same reasoning as at the end of the proof of Lemma 4.1. Let $\Delta \subset [m]$ be any set of cardinality $l = \lfloor (1+\kappa/2)um \rfloor$, and denote $\Phi^\Delta = S^\Delta \cap S(u, v)$. Set $\varepsilon = tvru \cdot \sqrt{\delta}$. By Lemma 3.3,

$$\begin{aligned}\mathbb{P}\left(\exists x \in (\Phi^\Delta)_{\varepsilon/(4K)} : \|Wx - z\|_2 \le t\frac{vru\sqrt{\delta n}}{2} \text{ and } \|W\| \le K\sqrt{n}\right) \\ \le (ct)^{(1+\kappa)um} \cdot \left(\frac{6K}{\varepsilon}\right)^l\end{aligned}$$

We have

$$\mathrm{Comp}\left(\left(1+\frac{\kappa}{2}\right)u, \frac{\varepsilon}{4K}\right) \setminus \mathrm{Comp}(u,v) \subset \bigcup_{|\Delta|=l} (\Phi^{\Delta})_{\varepsilon/(4K)}.$$

Therefore, the union bound yields

$$\mathbb{P}\left(\exists x \in \mathrm{Comp}\left(\left(1+\frac{\kappa}{2}\right)u, \frac{tvru\cdot\sqrt{\delta}}{4K}\right) \setminus \mathrm{Comp}(u,v) : \right.$$
$$\left. \|Wx-z\|_2 \le t\frac{vru\sqrt{\delta n}}{2} \text{ and } \|W\| \le K\sqrt{n}\right)$$
$$\le \binom{m}{l}\cdot(ct)^{(1+\kappa)um}\cdot\left(\frac{6K}{\varepsilon}\right)^{l} \le (ct)^{(1+\kappa)um}\cdot\left(\frac{CK}{u^2\cdot tvr\cdot\sqrt{\delta}}\right)^{(1+\kappa/2)um}$$
$$\le \left[\left(\frac{C'K}{v}\right)^{4/\kappa} t\right]^{\kappa um/2}.$$

This does not exceed $e^{-\kappa um/2}$ if we choose

$$t = e^{-1}\cdot\left(\frac{C'K}{v}\right)^{-4/\kappa}.$$

Substituting this t into the estimate above proves the lemma. □

Lemma 4.4. *Let $m,n \in \mathbb{N}$, $(2/3)m \le n \le 2m$. Let A, B be an $n\times m$ matrices, and set $W = A\odot G + B$, where G is the standard $n\times m$ Gaussian matrix, independent of A, B. Assume that*

(1) *$a_{i,j} \in \{0\}\cup[r,1]$ for some constant $r>0$ and all i,j;*
(2) *the graph Γ_{A^T} is strongly connected.*

Then for all $z \in \mathbb{R}^n$

$$\mathbb{P}\Big(\exists x \in \mathrm{Comp}(1-\kappa/2, K^{-C}) :$$
$$\|Wx-z\|_2 \le K^{-C}\sqrt{n} \text{ and } \|W\| \le K\sqrt{n}\Big) \le e^{-cn}.$$

Proof. Set $u_0 = c_0$, $v_0 = c_1K^{-2}$, where c_0, c_1 are the constants from Lemma 4.2. Let L be the smallest natural number such that

$$u_0(1+\kappa/2)^L > 1-\kappa/2.$$

Note that $u_0(1+\kappa/2)^L \le (1-\kappa/2)\cdot(1+\kappa/2) < 1$. Define by induction $v_{l+1} = (v_l/K)^{C+1}$, where C is the constant from Lemma 4.3. Then

$v_L = K^{-C'}$ for some $C' > 0$ depending only on the parameters δ, κ and r. We have

$$\mathrm{Comp}(1-\kappa/2, v_L) \subset \mathrm{Comp}(u_0, v_0) \cup \bigcup_{l=1}^{L} \mathrm{Comp}(u_0(1+\kappa/2)^l, v_l) \setminus \mathrm{Comp}(u_0(1+\kappa/2)^{l-1}, v_{l-1}).$$

The result now follows from Lemmas 4.2 and 4.3. □

5. Smallest singular value

To estimate the smallest singular value, we need the following result from [16, Lemma 3.5], that handles incompressible vectors.

Lemma 5.1. *Let W be an $n \times n$ random matrix. Let $W_1, \dots, W_n$ denote the column vectors of W, and let H_k denote the span of all column vectors except the k-th. Then for every $a, b \in (0,1)$ and every $t > 0$, one has*

(5.1)
$$\mathbb{P}\big(\inf_{x \in \mathrm{Incomp}(a,b)} \|Wx\|_2 < tbn^{-1/2}\big) \le \frac{1}{an} \sum_{k=1}^{n} \mathbb{P}\big(\mathrm{dist}(W_k, H_k) < t\big).$$

Now we can derive the first main result.

Proof of Theorem 2.3. Set $B = \mathbb{E}W$ and let $A = (a_{i,j})$, where $a_{i,j}^2 = \mathrm{Var}(w_{i,j})$, so

$$W = A \odot G + B,$$

where G is the $n \times n$ standard Gaussian matrix.

Without loss of generality, assume that $K > K_0$, where $K_0 > 1$ is a constant to be determined. Applying Lemma 4.2 to the matrix W, we obtain

$$\mathbb{P}\Big(\exists x \in \mathrm{Comp}\left(c_0, c_1 K^{-2}\right) : \|Wx\|_2 \le (4c_1/K)\sqrt{n} \text{ and } \|W\| \le K\sqrt{n}\Big) \le e^{-cn}.$$

Therefore, for any $t > 0$

$$\mathbb{P}\left(s_n(W) \le ctK^{-C} n^{-1/2}\right) \le e^{-cn} + \mathbb{P}\left(\|W\| \ge K\sqrt{n}\right) + \mathbb{P}\left(\exists x \in \mathrm{Incomp}\ \left(c_0, c_1 K^{-2}\right) :\ \|Wx\|_2 \le (4c_1/K)\sqrt{n}\right).$$

By Lemma 3.1,

$$\mathbb{P}\left(\|W\| > 2K\sqrt{n}\right) \le \mathbb{P}\left(\|A \odot G\| > K\sqrt{n}\right) \le e^{-cn},$$

provided that $K > K_0$ with K_0 taken large enough, thus determining K_0. By Lemma 5.1, it is enough to bound $\mathbb{P}\left(\mathrm{dist}(W_k, H_k) < ct\right)$ for all $k \in [n]$. Consider, for example, $k = 1$. In the discussion that follows, let $h \in S^{n-1}$ be a vector such that $h^T W_j = 0$ for all $j = 2, \dots, n$. Then

$$\mathrm{dist}(W_1, H_1) \geq |h^T W_1|.$$

Let $\tilde{A}$ be the $(n-1)\times n$ matrix whose rows are the columns of A^T, except the first one, i.e. $\tilde{A}^T = (A_2, A_3, \dots, A_n)$. Define the $(n-1)\times n$ matrices $\tilde{B}, \tilde{W}$ in the same way. The condition on h can now be rephrased as $\tilde{W}h = 0$.

Since the graph Γ_A is strongly connected, the graph $\Gamma_{\tilde{A}^T}$ is strongly connected with slightly smaller parameters and in particular with parameters $\delta/2$ and $\kappa/2$. Since $\mathrm{Comp}(1-\kappa/2, (2K)^{-C}) \subset \mathrm{Comp}(1-\kappa/4, (2K)^{-C})$, we get from Lemma 4.4 applied to $\tilde{W}$, $z = 0$, and with K replaced by $2K$, that

$$\begin{aligned}
&\mathbb{P}\left(\exists h \in \mathrm{Comp}(1-\kappa/2, (2K)^{-C}),\ \tilde{W}h = 0\right) \\
&\quad \leq \mathbb{P}\left(\exists h \in \mathrm{Comp}(1-\kappa/2, (2K)^{-C}), \left\|\tilde{W}h\right\|_2 \leq (2K)^{-C'}\sqrt{n}\right. \\
&\qquad \left. \text{and } \left\|\tilde{W}\right\| \leq 2K\sqrt{n}\right) + \mathbb{P}\left(\left\|\tilde{W}\right\| > 2K\sqrt{n}\right) \\
&\qquad \leq e^{-cn} + \mathbb{P}\left(\left\|\tilde{W}\right\| > 2K\sqrt{n}\right).
\end{aligned}$$

The last term is exponentially small:

$$\mathbb{P}\left(\left\|\tilde{W}\right\| > 2K\sqrt{n}\right) \leq \mathbb{P}\left(\|W\| > 2K\sqrt{n}\right) \leq e^{-cn}$$

Hence,

$$\mathbb{P}\left(\exists h \in \mathrm{Comp}(1-\kappa/2, (2K)^{-C}) :\ \tilde{W}h = 0\right) \leq e^{-c'n}.$$

Note that the vector h is independent of W_1. Therefore,

$$\begin{aligned}
&\mathbb{P}\left(\mathrm{dist}(W_1, H_1) < t(2K)^{-C}\right) \\
&\leq \mathbb{P}\left(|h^T W_1| \leq t(2K)^{-C},\ \tilde{W}h = 0, \text{ and } h \in \mathrm{Comp}(1-\kappa/2, (2K)^{-C})\right) \\
&\quad + \mathbb{P}\left(|h^T W_1| \leq t(2K)^{-C},\ \tilde{W}h = 0, \text{ and } h \notin \mathrm{Comp}(1-\kappa/2, (2K)^{-C})\right) \\
&\leq e^{-c'n} + \mathbb{E}\mathbb{P}_{W_1}(|h^T W_1| \leq t(2K)^{-C} \mid h \notin \mathrm{Comp}(1-\kappa/2, (2K)^{-C}) \\
&\leq e^{-c'n} + \sup_{u \in \mathrm{Incomp}(1-\kappa/2, (2K)^{-C})} \mathbb{P}\left(|u^T W_1| \leq ctK^{-C}\right)
\end{aligned}$$

Assume that $u \in \mathrm{Incomp}(1-\kappa/2, (2K)^{-C})$. Let $J = \{j \in [n] : |u_j| \geq (2K)^{-C} n^{-1/2}\}$. Then $|J| \geq (1-\kappa/2)n$. Hence, if $J' = \{j \in [n] : |a_{1j}| \geq rn^{-1/2}$, then $|J \cap J'| \geq (\delta - \kappa/2)n > (\delta/2)n$. Therefore, $u^T W_1$ is a

centered normal random variable with variance $\sigma^2 \geq r^2(2K)^{-2C} \cdot \delta/2$, and so

$$\mathbb{P}\left(|u^T W_1| \leq t(2K)^{-C}\right) \leq C't.$$

This means that

$$\mathbb{P}\left(\mathrm{dist}(W_1, H_1) < t(2K)^{-C}\right) \leq t + e^{-cn},$$

and the same estimate holds for $\mathrm{dist}(W_j, H_j),\ j > 1$, so the theorem follows from Lemma 5.1.

□

6. Intermediate singular value

The next elementary lemma allows one to find a set of rows of a fixed matrix with big ℓ_2 norms, provided that the graph of the matrix has a large minimal degree.

Lemma 6.1. *Let $k < n$, and let A be an $n \times n$ matrix. Assume that*

(1) *$a_{i,j} \in \{0\} \cup [r, 1]$ for some constant $r > 0$ and all i, j;*
(2) *the graph Γ_A satisfies deg$(j) \geq \delta n$ for all $j \in [n]$.*

Then for any $J \subset [n]$ there exists a set $I \subset [n]$ of cardinality

$$|I| \geq (r^2\delta/2)n,$$

such that for any $i \in I$

$$\sum_{j \in J} a_{i,j}^2 \geq (r^2\delta/2) \cdot |J|.$$

Proof. By the assumption on A,

$$\sum_{i=1}^{n} \sum_{j \in J} a_{i,j}^2 \geq r^2 \delta n \cdot |J|.$$

Let $I = \{i \in [n] \mid \sum_{j \in J} a_{i,j}^2 \geq r^2\delta \cdot |J|/2\}$. Then

$$|J| \cdot |I| \geq \sum_{i \in I} \sum_{j \in J} a_{i,j}^2 \geq r^2 \delta n \cdot |J| - \sum_{i \in I^c} \sum_{j \in J} a_{i,j}^2 \geq \frac{r^2\delta|J|}{2} \cdot n.$$

□

We also need the following lemma concerning the Gaussian measure in $\mathbb{R}^n$.

Lemma 6.2. *Let E, F be linear subspaces of $\mathbb{R}^n$. Let P_E, P_F be the orthogonal projections onto E and F, and assume that for some $\tau > 0$,*

$$\forall y \in F,\ \|P_E y\|_2 \geq \tau \|y\|_2 .$$

Let g_E be the standard Gaussian vector in E. Then for any $t > 0$

$$\mathbb{P}\left(\|P_F g_E\|_2 \le t\right) \le \left(\frac{ct}{\tau\sqrt{\dim F}}\right)^{\dim F}.$$

Proof. Let $E_1 = P_E F$. Then (because $\tau > 0$), the linear operator $P_E : F \to E_1$ has a trivial kernel and hence is a bijection. Denote by g_H the standard Gaussian vector in the space $H \subset \mathbb{R}^n$. Let $U : \mathbb{R}^n \to \mathbb{R}^n$ be an isometry such that $UE_1 = F$ and $UF = E_1$. Then $P_F = UP_{E_1}U$ and Ug_{E_1} has the same distribution as g_F. Therefore, integrating over the coordinates of g_E orthogonal to E_1, we get

$$\begin{aligned}\mathbb{P}\left(\|P_F g_E\|_2 \le t\right) &\le \mathbb{P}\left(\|UP_{E_1}Ug_{E_1}\|_2 \le t\right) \\ &= \mathbb{P}\left(\|P_{E_1} g_F\|_2 \le t\right) \le \mathbb{P}\left(\|g_F\|_2 \le t/\tau\right).\end{aligned}$$

The lemma follows from the standard density estimate for the Gaussian vector.

□

Let $J \subset [m]$. For levels $Q > q > 0$ define the set of totally spread vectors

$$\mathcal{S}^J_{q,Q} := \Big\{ y \in S^J : \ \frac{q}{\sqrt{|J|}} \le |y_k| \le \frac{Q}{\sqrt{|J|}} \quad \text{for all } k \in J \Big\}. \tag{6.1}$$

Lemma 6.3. *Let $\delta, \rho \in (0,1)$. There exist $Q > q > 0$ and $\alpha, \beta > 0$, which depend polynomially on δ, ρ, such that the following holds. Let $d \le m \le n$ and let W be an $n \times m$ random matrix with independent columns. For $I \subset [m]$ denote by H_I the linear subspace spanned by the columns W_i, $i \in I$. Let J be a uniformly chosen random subset of $[n]$ of cardinality d. Then for every $\varepsilon > 0$*

$$\begin{aligned}\mathbb{P}\Big(\inf_{x \in \mathrm{Incomp}(\delta,\rho)} \|Wx\|_2 < \alpha\varepsilon\sqrt{\frac{d}{n}} \Big) \\ \le \beta^d \cdot \mathbb{E}_J \mathbb{P}\big(\inf_{z \in \mathcal{S}^J_{q,Q}} \mathrm{dist}(Wz, H_{J^c}) < \varepsilon \big).\end{aligned} \tag{6.2}$$

Remark 6.4. Lemma 6.3 was proved in [17] for random matrices with i.i.d. entries (see Lemma 6.2 there). However, that proof can be extended to the general case without any changes.

Proof of Theorem 2.4. Set $B = \mathbb{E}W$ and let $A = (a_{i,j})$, where $a_{i,j}^2 = \mathrm{Var}(w_{i,j})$, so

$$W = A \odot G + B,$$

where G is the $n \times n$ standard Gaussian matrix. Without loss of generality assume that

$$\kappa \le \frac{r^2 \delta}{2}. \tag{6.3}$$

If this inequality doesn't hold, we can redefine κ as the right hand side of this inequality, and note that the strong connectedness property is retained when κ gets smaller.

Let $C > 0$ be as in Lemma 4.4. Decomposing the sphere into compressible and incompressible vectors, we write

$$\begin{aligned} & \mathbb{P}\left(s_m(W) \le ctK^{-C} \cdot \frac{n-m}{\sqrt{n}} \right) \\ \le\ & \mathbb{P}\left(\inf_{x \in \mathrm{Comp}(c_0, c_1 K^{-2})} \|Wx\|_2 \le ctK^{-C} \cdot \sqrt{n} \right) \\ & + \mathbb{P}\left(\inf_{x \in \mathrm{Incomp}(c_0, c_1 K^{-2})} \|Wx\|_2 \le ctK^{-C} \cdot \frac{n-m}{\sqrt{n}} \right). \end{aligned} \tag{6.4}$$

By Lemma 4.2, the first term in the right side of (6.4) does not exceed

$$e^{-c_2 n} + \mathbb{P}\left(\|W\| \ge 2K\sqrt{n} \right).$$

By Lemma 3.1, the last term in the last expression is smaller than e^{-cn}, if K is large enough.

To estimate the second term in the right side of (6.4) we use Lemma 6.3. Recall that by that lemma, we can assume that $q = K^{-C'}$ and $Q = K^{C'}$ for some constant C'. Then the lemma reduces the problem to estimating

$$\mathbb{P}\Big(\inf_{z \in \mathcal{S}^J_{q,Q}} \mathrm{dist}(Wz, H_{J^c}) < \varepsilon \Big)$$

for these q, Q and for a fixed subset $J \subset [m]$ of cardinality

$$d = \left\lfloor \frac{n-m}{2} \right\rfloor,$$

and with a properly chosen ε, see (6.8) below.

Since we do not control the norm of the submatrix matrix B corresponding to J, we will reduce the dimension further to eliminate this matrix. Set $H_0 = B\mathbb{R}^J \subset \mathbb{R}^n$, and let $F = (H_{J^c} \cup H_0)^\perp$. Then F is a linear subspace of $\mathbb{R}^n$ independent of $\{W_j,\ j \in J\}$, and

$$n - m \le \dim F \le n - m + d \le 2(n-m)\,. \tag{6.5}$$

Since $P_F B\mathbb{R}^J = \{0\}$, we get

(6.6)

$$\mathbb{P}\left(\exists z \in \mathcal{S}^J_{q,Q} : \ \mathrm{dist}(Wz, H_{J^c}) < \varepsilon\right) \le \mathbb{P}\left(\exists z \in \mathcal{S}^J_{q,Q} : \ \|P_F W z\|_2 < \varepsilon\right)$$
$$= \mathbb{P}\left(\exists z \in \mathcal{S}^J_{q,Q} : \ \|P_F (A \odot G) z\|_2 < \varepsilon\right)$$

for any $\varepsilon > 0$.

We start with bounding the small ball probability for a fixed vector $z \in \mathcal{S}^J_{q,Q}$. The i-th coordinate of the vector $(A \odot G)z$ is a normal random variable with variance

$$\sigma_i^2 = \sum_{j \in J} a_{i,j}^2 x_j^2 \ge \frac{q^2}{d} \sum_{j \in J} a_{i,j}^2.$$

Let $I \subset [n]$ be the set constructed in Lemma 6.1. Then for any $i \in I$ we have $\sigma_i \ge cq = c' K^{-C'}$. Let E be the subspace of $\mathbb{R}^n$ spanned by the vectors e_i, $i \in I$.

Since $P_E(A \odot G)z$ and $P_{E^\perp}(A \odot G)z$ are independent Gaussian vectors,

(6.7)

$$\mathbb{P}\left(\|P_F(A \odot G)z\|_2 < \varepsilon\right)$$
$$= \mathbb{E}_{P_{E^\perp}(A \odot G)} \mathbb{P}\left(\|P_F P_E (A \odot G)z + P_F P_{E^\perp}(A \odot G)z\|_2 < \varepsilon \mid P_{E^\perp}(A \odot G)\right)$$
$$\le \mathbb{P}\left(\|P_F P_E (A \odot G)z\|_2 < \varepsilon\right) \le \mathbb{P}\left(\|P_F g_E\|_2 < cK^{C'}\varepsilon\right).$$

Here g_E is the standard Gaussian vector in E. The first inequality in (6.7) is a consequence of Anderson's inequality [1, Theorem 1], applied to the convex symmetric function $f(x) = \mathbf{1}_{\|x\|_2 < \varepsilon}$ and the Gaussian random vector $P_F P_E (A \odot G)x$. The last inequality in (6.7) follows since $P_E(A \odot G)z$ is a vector with independent normal coordinates with variances greater than $c' K^{-C'}$.

Now we have to check that the spaces E and F satisfy the conditions of Lemma 6.2, with high probability. Let $\tilde{A}, \tilde{B}, \tilde{G}$, and $\tilde{W}$ be $(m - d) \times n$ matrices whose rows coincide with the columns of the matrices A, B, G, and W corresponding to the set J^c. Then the condition $F \perp \mathrm{span}(W_j, \ j \in J^c)$ can be rewritten as $F \subset \mathrm{Ker}(\tilde{W})$. By Lemma 4.4 and (6.3),

$$\mathbb{P}\left(F \cap S^{n-1} \not\subset \mathrm{Incomp}(1 - r^2\delta/4, K^{-C})\right)$$
$$\le \mathbb{P}\left(\exists x \in \mathrm{Comp}(1 - \kappa/2, K^{-C}) : \tilde{W}x = 0\right) \le e^{-c_3 n}.$$

Assume that $F \cap S^{n-1} \subset \mathrm{Incomp}(1 - r^2\delta/4, K^{-C})$. Since $\dim E = |I| \ge (r^2\delta/4)n$, the incompressibility means that for any $y \in F \cap$

S^{n-1}, $\|P_E y\|_2 \geq \tau = K^{-C}$. Hence, by (6.7) and Lemma 6.2, for $z \in \mathcal{S}^J_{q,Q}$,

$$\mathbb{P}\left(\|P_F(A \odot G)z\|_2 < \varepsilon \text{ and } F \cap S^{n-1} \subset \mathrm{Incomp}(1 - r^2\delta/4, K^{-C})\right) \leq \left(\frac{c' K^{C'} \varepsilon}{\tau\sqrt{n-m}}\right)^{n-m} \leq \left(\frac{c' K^{C''} \varepsilon}{\sqrt{n-m}}\right)^{n-m}.$$

By Lemma 3.1 and (6.5),

$$\mathbb{P}\left(\left\|P_F(A \odot G) : \mathbb{R}^J \to \mathbb{R}^n\right\| \geq C_0 t^{-1/2}\sqrt{n-m}\right) \leq e^{-t^{-1}(n-m)}.$$

Let

$$\eta = \frac{\varepsilon\sqrt{t}}{2C_0\sqrt{n-m}}.$$

By the volumetric estimate we can find an η-net $\mathcal{N}^J \subset \mathcal{S}^J_{q,Q}$ of cardinality

$$|\mathcal{N}^J| \leq \left(\frac{3}{\eta}\right)^d.$$

For η chosen above we have

$$\mathbb{P}\left(\exists z \in \mathcal{N}^J : \|P_F(A \odot G)z\|_2 < \varepsilon \text{ and } F \cap S^{n-1} \subset \mathrm{Incomp}(1 - r^2\delta/4, K^{-C})\right) \leq \left(\frac{3}{\eta}\right)^d \cdot \left(\frac{c' K^{C''} \varepsilon}{\sqrt{n-m}}\right)^{n-m} \leq \left(\frac{c'' K^{2C''} \varepsilon}{\sqrt{t}\sqrt{n-m}}\right)^{(n-m)/2}.$$

This does not exceed $t^{(n-m)/4}$, if we set

$$\varepsilon = cK^{-C}\sqrt{n-m} \cdot t. \tag{6.8}$$

Assume now that

- $\forall z \in \mathcal{N}^J \ \ \|P_F(A \odot G)z\|_2 \geq \varepsilon$;
- $F \cap S^{n-1} \subset \mathrm{Incomp}(1 - r^2\delta/4, K^{-C})$;
- $\left\|P_F(A \odot G) : \mathbb{R}^J \to \mathbb{R}^n\right\| \leq C_0 t^{-1/2}\sqrt{n-m}$.

The previous proof shows that these conditions are satisfied with probability at least

$$1 - t^{(n-m)/4} - e^{-c_3 n} - e^{-t^{-1}(n-m)} \geq 1 - 2t^{(n-m)/4} - e^{-c_3 n}$$

Let $z' \in \mathcal{S}^J_{q,Q}$ and let $z \in \mathcal{N}^J$ be an η-approximation of z': $\|z' - z\|_2 < \eta$. Then, on the event above,

$$\|P_F(A \odot G)z'\|_2 \geq \|P_F(A \odot G)z\|_2 - \left\|P_F(A \odot G) : \mathbb{R}^J \to \mathbb{R}^n\right\| \cdot \eta \geq \varepsilon - C_0 t^{-1/2}\sqrt{n-m} \cdot \eta \geq \varepsilon/2.$$

We thus have proved that

$$\mathbb{P}\left(\exists z \in \mathcal{S}^J_{q,Q} : \ \|P_F(A \odot G)z\|_2 < cK^{-C}\sqrt{n-m}\cdot t\right) \\ \leq 2t^{(n-m)/4} + e^{-c_3 n}.$$

Combining this with (6.2), (6.6), and (6.8) we obtain

$$\begin{aligned}
&\mathbb{P}\left(\inf_{x\in \mathrm{Incomp}(c_0,c_1K^{-2})} \|Wx\|_2 \leq \alpha\cdot ctK^{-C}\cdot\frac{n-m}{\sqrt{n}}\right) \\
&\leq \beta^d \cdot \max_{J\subset[n],\ |J|=d} \mathbb{P}\left(\exists z\in\mathcal{S}^J_{q,Q} : \ \mathrm{dist}(Wz, H_{J^c}) < \alpha\cdot ctK^{-C}\cdot\sqrt{n-m}\right) \\
&\leq \beta^d \cdot \max_{J\subset[n],\ |J|=d} \mathbb{P}\left(\exists z\in\mathcal{S}^J_{q,Q}\ \ \|P_F(A\odot G)z\|_2 < \alpha\cdot ctK^{-C}\cdot\sqrt{n-m}\right) \\
&\leq \beta^d\cdot\left(2t^{(n-m)/4} + e^{-c_3 n}\right).
\end{aligned}$$

Recall that $d = \lfloor (n-m)/2 \rfloor$, $\alpha = K^{-c}$, and $\beta = K^c$. Replacing t by $\beta^2 t$ in the inequality above to eliminate the coefficient β^d in the right hand side, we complete the proof of the theorem.

□

7. Applications

7.1. **Singular value bounds.** The bound on the smallest singular value in Theorem 2.5 follows immediately from Theorem 2.3 and Lemma 3.1. To bound $s_m(A \odot G)$ we apply Theorem 2.4 to the matrix W consisting of the first m columns of $A \odot G$ and note that $s_m(W) \leq s_m(A \odot G)$.

To prove Theorem 2.7, decompose the matrix W by writing $W = W^{(1)} + W^{(2)}$ where $W^{(1)}$ and $W^{(2)}$ are independent centered Gaussian matrices with independent entries and

$$\mathrm{Var}(W^{(1)}_{i,j}) = \begin{cases} c/n, & \text{if } c/n \leq a_{i,j} \\ 0, & \text{otherwise,} \end{cases}, \qquad \mathrm{Var}(W^{(2)}_{i,j}) = a_{i,j} - \mathrm{Var}(W^{(1)}_{i,j}).$$

Let Ω be the event $\left\|W^{(2)}\right\| \leq C'\log^2 n$. By Lemma 3.2, for appropriate constants C', C'', one has

$$\mathbb{P}(\Omega^c) \leq \mathbb{P}(\|W\| \geq C'\log^2 n) \leq \exp(-C''\log^4 n).$$

On the other hand,

$$\begin{aligned}
&\mathbb{P}\left(s_n(W) \leq ctn^{-1}\log^{-c'} n \text{ and } \Omega\right) \\
&\qquad \leq E_{W^{(2)}}\mathbb{P}\left(s_n(W^{(1)} + W^{(2)}) \leq ctn^{-1}\log^{-C'} n \mid \Omega\right) \\
&\qquad\qquad \leq \sup_{X:\|X\|\leq C'\log^2 n} \mathbb{P}\left(s_n(W^{(1)} + X) \leq ctn^{-1}\log^{-c'} n\right).
\end{aligned}$$

By Theorem 2.3 applied to $\sqrt{n}W^{(1)}$ and $B = \sqrt{n}X$ with $K = C' \log^2 n$, the last probability is at most $t + e^{-cn}$. The second estimate in Theorem 2.7 is proved by the same argument.

7.2. **Permanent estimates.** We turn next to the proof of the theorems in the introduction. We begin with a refinement of Theorem 1.1.

Theorem 7.1. *There exist $\tilde{C}, c, c'$ depending only on δ, κ such that for any $\tau \geq 1$ and any adjacency matrix A of a (δ, κ)-strongly connected graph*

$$(7.1)\ \mathbb{P}\left(\left|\log \det^2(A_{1/2} \odot G) - \mathbb{E} \log \det^2(A_{1/2} \odot G)\right| > \tilde{C}(\tau n \log n)^{1/3}\right)$$
$$\leq \ 6\exp(-\tau) + 3\exp\left(-c\tau^{1/3} n^{1/3} \log^{-2/3} n\right) + 9\exp\left(-c' n\right).$$

and

(7.2)

$$\mathbb{E} \log \det^2(A_{1/2} \odot G) \leq \log \operatorname{per}(A) \leq \mathbb{E} \log \det^2(A_{1/2} \odot G) + C'\sqrt{n \log n}.$$

Theorem 1.1 follows from Theorem 7.1 since the right side of (7.1) does not exceed $9\exp(-\tau) + 12\exp(-c\sqrt{n}/\log n)$. The coefficients 9 and 12 can be removed by adjusting the constants $\tilde{C}$ and c'.

Proof. The proof of Theorem 7.1 is partially based on the ideas of [5, Pages 1563–1566]. We would like to apply the Gaussian concentration inequality to the logarithm of the determinant of the matrix $A_{1/2} \odot G$, which can be written as the sum of the logarithms of its singular values. However, since the logarithm is not a Lipschitz function, we will have to truncate it in a neighborhood of zero in order to be able to apply the concentration inequality. This truncation is introduced in Section 7.2.1.

The singular values will be divided into two groups. For the large values of $n-l$ we use the concentration of the (sums of subsets) singular values $s_{n-l}(A_{1/2} \odot G)$ around their mean. In contrast to [5], we do not use the concentration inequality once, but rather divide the range of singular values to several subsets, and apply separately the concentration inequality in each subset. The definition of the subsets, introduced in Section 7.2.1, will be chosen to match the singular values estimates of Theorem 2.4.

On the other hand, when $n - l$ becomes small, the concentration doesn't provide an efficient estimate. In that case we use the lower bounds for such singular values obtained in Theorem 2.3. Because the number of singular values treated this way is small, their total

contribution to the sum of the logarithms will be small as well. This computation is described in Section 7.2.2.

Getting rid of the truncation of the logarithm requires an a-priori rough estimate on the second moment of $\log \det^2(A_{1/2} \odot G)$, which is presented in Lemma 7.2 and proved in Section 7.3. With this, we arrive in Section 7.2.3, to the control of the deviations of $\log \det^2(A_{1/2} \odot G)$ from $\mathbb{E} \log \det^2(A_{1/2} \odot G)$ that is presented in (7.1).

To complete the proof of the Theorem, we will need to relate $\mathbb{E} \log \det^2(A_{1/2} \odot G)$ to $\log \mathbb{E} \det^2(A_{1/2} \odot G) = \mathrm{perm}(A)$. This is achieved in Section 7.2.4 by again truncating the log (at a level different than that used before) and employing an exponential inequality.

7.2.1. *Construction of the truncated determinant.* Let $k_* \in \mathbb{N}$ be a number to be specified later. We choose truncation dimensions n_k and the truncation levels ε_k for large codimensions first. For $k = 0, \dots, k_*$ set

$$\begin{aligned} n_k &= n \cdot 2^{-4k}; \\ t_k &= \sqrt{\tau} \cdot 2^{k+k_*}; \\ \varepsilon_k &= c_0 \frac{n_k}{\sqrt{n}} = c_0 \sqrt{n} \cdot 2^{-4k}. \end{aligned}$$

Here, c_0 is a fixed constant to be chosen below. We also set $l_* = n_{k_*}$.

For any $n \times n$ matrix V define the function $f(V)$ by

$$f(V) = \sum_{k=1}^{k_*} f_k(V), \quad \text{where } f_k(V) = \sum_{l=n-n_{k-1}}^{n-n_k-1} \log_{\varepsilon_k}(s_{n-l}(V)),$$

where $\log_\varepsilon(x) = \log(x \vee \varepsilon)$. Recall that the function $S : \mathbb{R}^{n^2} \to \mathbb{R}^n_+$ defined by $S(V) = (s_1(V), \dots, s_n(V))$ is 1-Lipschitz. Hence, each function f_k is Lipschitz with Lipschitz constant

$$L_k \le \frac{\sqrt{n_{k-1} - n_k}}{\varepsilon_k} \le c' \cdot 2^{2k}.$$

Denote $W = A_{1/2} \odot G$. The concentration of the Gaussian measure implies that for an appropriately chosen constant C, one has

$$\mathbb{P}\left(|f_k(W) - \mathbb{E} f_k(W)| > C t_k\right) \le 2 \exp\left(-\frac{c t_k^2}{L_k^2}\right) \le 2 \exp\left(-2^{2(k_*-k)}\tau\right).$$

(For this version, see e.g. [13, Formula (2.10)].) Therefore,

(7.3)

$$\mathbb{P}\left(|f(W) - \mathbb{E} f(W)| > C \sum_{k=1}^{k_*} t_k\right) \le 2 \sum_{k=1}^{k_*} \exp\left(-2^{2(k_*-k)}\tau\right) \le 4e^{-\tau}.$$

Here

$$\sum_{k=1}^{k_*} t_k = \sum_{k=1}^{k_*} \sqrt{\tau} 2^{k+k_*} \le 2\sqrt{\tau} 2^{2k_*} = 2\sqrt{\tau} \cdot \sqrt{\frac{n}{l_*}}.$$

We similarly handle singular values s_{n-l} for $l \ge n - l_*$. Define the function $g(V) = \sum_{l=n-n_{k_*}}^{n} \log_{\varepsilon_{k_*}}(s_{n-l}(V))$, whose Lipschitz constant is bounded by $\sqrt{l_*}/\varepsilon_{k^*} = c_0^{-1}\sqrt{n/l_*}$, and therefore

$$\mathbb{P}\left(|g(W) - \mathbb{E}g(W)| \ge c_1\sqrt{\tau} \cdot \sqrt{\frac{n}{l_*}}\right) \le 2e^{-\tau}. \tag{7.4}$$

Set

$$\varepsilon(l) = \begin{cases} \varepsilon_k, & l \in [n_k + 1, n_{k-1}] \\ \varepsilon_{k_*}, & l \le n_{k_*} = l_*. \end{cases}$$

Define

$$\widetilde{\det}(W, l_*) = \prod_{l=0}^{n-1} (s_{n-l}(W) \vee \varepsilon(l))^2.$$

We include l_* as the second argument to emphasize the dependence on the truncation level. From (7.3) and (7.4), we obtain the large deviation bound for the logarithm of the truncated determinant:

$$\mathbb{P}\left(|\log \widetilde{\det}(W, l_*) - \mathbb{E}\log \widetilde{\det}(W, l_*)| \ge c_2\sqrt{\tau}\sqrt{n/l_*}\right) \le 6e^{-\tau}. \tag{7.5}$$

7.2.2. *Basic concentration estimate for* $\log \det^2(W)$. Our next goal is to get rid of the truncation, i.e., to relate $\widetilde{\det}(W, l_*)$ to $\det^2(W)$. Toward this end, define the set of $n \times n$ matrices $\mathcal{W}_1$ as follows:

$$\mathcal{W}_1 = \{V \mid \exists k,\ 1 \le k \le k_*,\ s_{n-n_k}(V) < \varepsilon_k\}.$$

Then by Theorem 2.4,

$$\mathbb{P}(W \in \mathcal{W}_1) \le \sum_{k=1}^{k_*} \left(c_0\varepsilon_k \cdot \frac{\sqrt{n}}{n_k}\right)^{n_k} + k_* e^{-cn} \le 2e^{-n_{k_*}/4},$$

with an appropriate choice of the constant c_0.

For codimensions smaller than $l_* = n_{k_*}$ we simply estimate the total contribution of small singular values. For $0 \le l \le l_*$ set

$$d_{n-l} = n^{-\frac{l_*}{(l+1)\log l_*} - \frac{1}{2}}.$$

Let $\mathcal{W}_2$ be the set of $n \times n$ matrices defined by

$$\mathcal{W}_2 = \{W \mid \exists l \le l_*,\ s_{n-l}(W) \le d_{n-l}\}.$$

Applying Theorem 2.3 for $0 \le l < 4$ and 2.4 for $4 \le l \le l_*$, we obtain

$$\begin{aligned}\mathbb{P}\left(W \in \mathcal{W}_2\right) &\le \sum_{l=0}^{3} c\sqrt{n} \cdot d_{n-l} + \sum_{l=4}^{l_*} \left(c\frac{\sqrt{n}}{l} \cdot d_{n-l}\right)^{l/4} + (l_*+1)e^{-cn} \\ &\le C l_* \cdot n^{-\frac{l_*}{4\log l_*}} \le C l_* \cdot \exp(-l_*/4) \le \exp(-l_*/8).\end{aligned}$$

Assume that $V \notin \mathcal{W}_2$. Then

$$\sum_{l=0}^{l_*} \log s_{n-l}^{-1}(V) \le \sum_{l=0}^{l_*} \log n \cdot \left(\frac{1}{2} + \frac{l_*}{(l+1)\log l_*}\right) \le \frac{3}{2} l_* \log n.$$

Let $\mathcal{W}_3$ denote the set of all $n \times n$ matrices V such that $\|V\| \ge n$. Then $\mathbb{P}\left(W \in \mathcal{W}_3\right) < e^{-n}$. If $V \notin \mathcal{W}_3$, then

$$\sum_{l=0}^{l_*} \log s_{n-l}(V) \le l_* \log n.$$

Therefore, for any $V \in (\mathcal{W}_2 \cup \mathcal{W}_3)^c$,

$$-\frac{3}{2} l_* \log n \le \sum_{l=0}^{l_*} \log s_{n-l}(V) \le \sum_{l=0}^{l_*} \log(s_{n-l}(V) \vee \varepsilon_{k_*}) \le l_* \log n\,.$$

We thus obtain that if $W \in (\mathcal{W}_1 \cup \mathcal{W}_2 \cup \mathcal{W}_3)^c$ then

$$|\log \det^2(W) - \log \widetilde{\det}(W, l_*)| \le \frac{3}{2} l_* \log n \tag{7.6}$$

Note that the event $W \in (\mathcal{W}_1 \cup \mathcal{W}_2 \cup \mathcal{W}_3)^c$ has probability larger than $1 - 3e^{-l_*/8}$.

Setting

$$Q(l_*) = \mathbb{E} \log \widetilde{\det}(W, l_*),$$

we thus conclude from (7.5) that

$$\begin{aligned}\mathbb{P}\left(\left|\log \det^2(W) - Q(l_*)\right| \ge \frac{3}{2} l_* \log n + c_2\sqrt{\tau n / l_*}\right) \\ \le 6\exp(-\tau) + 3\exp(-l_*/8)\,.\end{aligned} \tag{7.7}$$

This is our main concentration estimate. We will use it with l_* depending on τ to obtain an optimized concentration bound. Also, we will use special choices of l_* to relate a hard to evaluate quantity $Q(l_*)$ to the characteristics of the distribution of $\det^2(W)$, namely to $\mathbb{E}\log\det^2(W)$ and $\log \mathbb{E}\det^2(W)$. This will be done by comparing $\mathbb{E}\log\det^2(W)$ to $Q(l_1)$ and $\log\mathbb{E}\det^2(W)$ to $Q(l_2)$ for different values l_1 and l_2. This means that we also have to compare $Q(l_1)$ and $Q(l_2)$. The last comparison requires only (7.7).

Let $100 \le l_1, l_2 \le n/2$. For $j = 1, 2$, denote

$$\tilde{W}_j = \left\{ V \Big| |\log \det^2(V) - Q(l_j)| \le \frac{3}{2} l_j \log n + 4c_2 \sqrt{n/l_j} \right\}.$$

Using (7.7) with $\tau = 16$, we show that $\mathbb{P}(\tilde{W}_j) > 1/2$ for $j = 1, 2$. This means that $\tilde{W}_1 \cap \tilde{W}_2 \neq \emptyset$. Taking $V \in \tilde{W}_1 \cap \tilde{W}_2$, we obtain

$$\begin{aligned} |Q(l_1) - Q(l_2)| &\le |Q(l_1) - \log \det^2(V)| + |\log \det^2(V) - Q(l_2)| \\ &\le \frac{3}{2}(l_1 + l_2) \log n + cn^{1/2}(l_1^{-1/2} + l_2^{-1/2}). \end{aligned} \tag{7.8}$$

7.2.3. *Comparing $Q(l_*)$ to $\mathbb{E} \log \det^2(W)$.* Our next task is to relate $\mathbb{E} \log \det^2(W)$ to $Q(l_*)$ for some $l_* = l_1$. Toward this end we optimize the left side of (7.7) for $\tau = 8$ by choosing $l_* = l_1$, where

$$2n^{1/3} \log^{-2/3} n \le l_1 = n \cdot 2^{-4k_1} < 32 n^{1/3} \log^{-2/3} n.$$

Then we get from (7.7) that there exists $c > 0$ such that for all $\tau \ge 1$,

$$\begin{aligned} \mathbb{P}\left(\left| \log \det^2(W) - Q(l_1) \right| \ge c\tau^{1/2} (n \log n)^{1/3} \right) \\ \le 6 \exp(-\tau) + 3 \exp(-l_1/8). \end{aligned} \tag{7.9}$$

Let $\mathcal{W}_4$ be the set of all $n \times n$ matrices V such that $|\log \det(V)^2 - Q(l_1)| > \sqrt{n}$. The inequality (7.9) applied with $\tau = c' l_1$ for an appropriate c' reads

$$\mathbb{P}(W \in \mathcal{W}_4) \le \exp(-c l_1) = \exp\left(-C n^{1/3} \log^{-2/3} n \right). \tag{7.10}$$

We have

$$\begin{aligned} &|\mathbb{E} \log \det^2(W) - Q(l_1)| \le \mathbb{E} |\log \det^2(W) - Q(l_1)| \\ =&\mathbb{E} |\log \det^2(W) - Q(l_1)| \cdot \mathbf{1}_{\mathcal{W}_4^c}(W) + \mathbb{E} |\log \det^2(W) - Q(l_1)| \cdot \mathbf{1}_{\mathcal{W}_4}(W). \end{aligned}$$

The first term here can be estimated by integrating the tail in (7.9):

$$\begin{aligned} &\mathbb{E} |\log \det^2(W) - Q(l_1)| \cdot \mathbf{1}_{\mathcal{W}_4^c}(W) \\ &\le c(n \log n)^{1/3} + \int_{c(n \log n)^{1/3}}^{\sqrt{n}} \mathbb{P}\left(|\log \det^2(W) - Q(l_1)| > x \right) dx \\ &\le c(n \log n)^{1/3} + \int_1^{c' l_1} 2 \exp\left(- \left(\frac{x}{c(n \log n)^{1/3}} \right)^2 \right) dx \le C(n \log n)^{1/3}. \end{aligned}$$

To bound the second term, we need the following rough estimate of the second moment of the logarithm of the determinant. The proof of this estimate will be presented in the next subsection.

Lemma 7.2. *Let $W = G \odot A'_{1/2}$, where G is the standard Gaussian matrix, and A' is a deterministic matrix with entries $0 \le a_{i,j} \le 1$ for all i, j having at least one generalized diagonal with entries $a'_{i,\pi(i)} \ge c/n$ for all i. Then*

$$\mathbb{E} \log^2 \det^2(W) \le \bar{C} n^3.$$

Since A is the matrix of a (δ, κ)-strongly connected graph, it satisfies the conditions of Lemma 7.2. The estimate of the second term follows from Lemma 7.2, (7.9), and the Cauchy–Schwarz inequality:

$$\begin{aligned}
&\mathbb{E}|\log \det^2(W) - Q(l_1)| \cdot \mathbf{1}_{\mathcal{W}_4}(W) \\
&\le \left(\mathbb{E}|\log \det^2(W) - Q(l_1)|^2\right)^{1/2} \cdot \mathbb{P}^{1/2}(W \in \mathcal{W}_4) \\
&\le (\bar{C} n^3 + 2Q^2(l_1))^{1/2} \cdot \exp\left(-(C/2) n^{1/3} \log^{-2/3} n\right).
\end{aligned}$$

Combining the bounds for $\mathcal{W}_4$ and $\mathcal{W}_4^c$, we get

$$\begin{aligned}
&|\mathbb{E} \log \det^2(W) - Q(l_1)| \\
&\le C(n \log n)^{1/3} + (\bar{C} n^3 + 2Q^2(l_1))^{1/2} \cdot \exp\left(-(C/2) n^{1/3} \log^{-2/3} n\right),
\end{aligned}$$

which implies

$$|\mathbb{E} \log \det^2(W) - Q(l_1)| \le C'(n \log n)^{1/3}. \tag{7.11}$$

7.2.4. *Comparing* $\log \mathbb{E} \det^2(W)$ *to* $\mathbb{E} \log \det^2(W)$. We start with relating $Q(l_1)$ and $\log \mathbb{E}\det^2(W) = \log \operatorname{perm}(A)$. To this end we will use a different value of l_*. Namely, choose l_2 so that

$$\sqrt{n/\log n} \le l_2 = n \cdot 2^{4k_2} < 16\sqrt{n/\log n}.$$

The reasons for this choice will become clear soon. Denote for brevity

$$U := \log \widetilde{\det}(W, l_2) - \mathbb{E} \log \widetilde{\det}(W, l_2).$$

We deduce from (7.5) that

$$\begin{aligned}
\mathbb{E}(e^U) &\le \mathbb{E}(e^{|U|}) \le 1 + \int_0^\infty e^t \mathbb{P}\left(|U| \ge t\right) dt \\
&\le 1 + 6 \int_0^\infty e^t e^{-t^2 l_2 / c_2 n} \, dt \le 1 + c_3 e^{c_4 n / l_2}.
\end{aligned}$$

Taking logarithms, we conclude that

$$\begin{aligned}
\log \mathbb{E}\det^2(W) &\le \log \mathbb{E}\widetilde{\det}(W, l_2) \\
&\le \mathbb{E} \log \widetilde{\det}(W, l_2) + \log(1 + c_3 e^{c_3 n / l_2}) \\
&\le Q(l_2) + c_4 n / l_2 .
\end{aligned}$$

The inequality (7.8) implies

$$\begin{aligned}(7.12)\qquad &\log \mathbb{E}\det^2(W)\\ &\le Q(l_1) + c_4 n/l_2 + \frac{3}{2}(l_1+l_2)\log n + cn^{1/2}(l_1^{-1/2}+l_2^{-1/2})\\ &\le Q(l_1) + c_5\sqrt{n\log n}.\end{aligned}$$

The value of l_2 was selected to optimize the inequality (7.12). To bound $Q(l_1) - \log \mathbb{E}\det^2(W)$ from above, we use (7.9) with $\tau = 4$ to derive

$$\begin{aligned}(7.13)\quad \mathbb{P}\left(|\log\det^2(W) - Q(l_1)| \le 2c(n\log n)^{1/3}\right)\\ \ge 1 - 6e^{-4} - 3e^{-l_1/8} > \frac{1}{2}.\end{aligned}$$

On the other hand, Chebyshev's inequality applied to the random variable $\det^2(W)/\mathbb{E}\det^2(W)$ implies that

$$\mathbb{P}\left(\det^2(W) \le 2\mathbb{E}\det^2(W)\right) \ge \frac{1}{2},$$

and therefore

$$(7.14)\qquad \mathbb{P}\left(\log\det^2(W) - \log\mathbb{E}\det^2(W) \le \log 2\right) \ge \frac{1}{2}.$$

This means that the events in (7.13) and (7.14) intersect, and so

$$Q(l_1) - \log\mathbb{E}\det^2(W) \le 2c(n\log n)^{1/3} + \log 2.$$

Together with (7.12) this provides a two-sided bound

$$\begin{aligned}|Q(l_1) - \log\mathbb{E}\det^2(W)| &\le \max\left(c_5\sqrt{n\log n},\ 2c(n\log n)^{1/3} + \log 2\right)\\ &= c_5\sqrt{n\log n}\end{aligned}$$

for a sufficiently large n. The combination of this inequality with (7.11) yields

$$|\mathbb{E}\log\det^2(W) - \log\mathbb{E}\det^2(W)| \le c_6\sqrt{n\log n}.$$

7.2.5. *Concentration around* $\mathbb{E}\log\det^2(W)$. To finish the proof we have to derive the concentration inequality. This will be done by choosing the truncation parameter l_* depending on τ. Namely, assume first that $1 \le \tau \le n^2\log^2 n$ and define l_* by

$$2^{-8}\tau^{1/3}n^{1/3}\log^{-2/3}n < l_* = n\cdot 2^{-4k_*} \le 2^{-4}\tau^{1/3}n^{1/3}\log^{-2/3}n.$$

The constraint on τ is needed to guarantee that $k_* \geq 1$. Substituting this l_* in (7.7), we get

$$\mathbb{P}\left(\left|\log\det^2(W) - Q(l_*)\right| \geq \frac{3}{2} l_* \log n + c_2\sqrt{\tau n / l_*}\right) \leq 6\exp(-\tau) + 3\exp\left(-c\tau^{1/3} n^{1/3}\log^{-2/3} n\right).$$

By (7.11) and (7.8), for such τ we have

$$\begin{aligned}
&|\mathbb{E}\log\det^2(W) - Q(l_*)| \\
&\leq |\mathbb{E}\log\det^2(W) - Q(l_1)| + |Q(l_1) - Q(l_*)| \\
&\leq C'(n\log n)^{1/3} + \frac{3}{2}(l_1 + l_*)\log n + cn^{1/2}(l_1^{-1/2} + l_*^{-1/2}) \\
&\leq C''(\tau n \log n)^{1/3}.
\end{aligned}$$

Together with the previous inequality, this implies

$$\mathbb{P}\left(\left|\log\det^2(W) - \mathbb{E}\log\det^2(W)\right| \geq \tilde{C}(\tau n\log n)^{1/3}\right) \leq 6\exp(-\tau) + 3\exp\left(-c\tau^{1/3} n^{1/3}\log^{-2/3} n\right),$$

if the constant $\tilde{C}$ is chosen large enough.

If $\tau > \tau_0 := n^2\log^2 n$, we use the inequality above with $\tau = \tau_0$ and obtain

$$\mathbb{P}\left(\left|\log\det^2(W) - \mathbb{E}\log\det^2(W)\right| \geq \tilde{C}(\tau n\log n)^{1/3}\right) \leq 9\exp\left(-c' n\right),$$

Finally, for all $\tau \geq 1$, this implies

$$\mathbb{P}\left(\left|\log\det^2(W) - \mathbb{E}\log\det^2(W)\right| \geq \tilde{C}(\tau n\log n)^{1/3}\right) \leq 6\exp(-\tau) + 3\exp\left(-c\tau^{1/3} n^{1/3}\log^{-2/3} n\right) + 9\exp\left(-c' n\right).$$

which completes the proof of Theorem 7.1. □

7.3. **Second moment of the logarithm of the determinant.** It remains to prove Lemma 7.2. The estimate of the lemma, which was necessary in the proof of (1.2), is very far from being precise, so we will use rough, but elementary bounds.

Proof of Lemma 7.2. We will estimate the expectations of the squares of the positive and negative parts of the logarithm separately. Denote by $W_1, \dots, W_n$ the columns of the matrix W. By the Hadamard

inequality,

$$\mathbb{E}\log_+^2 \det(W)^2 \le \sum_{j=1}^n \mathbb{E}\log_+^2 \|W_j\|_2^2 \le n\sum_{j=1}^n\sum_{i=1}^n \mathbb{E}\log^2(1+w_{i,j}^2) \le Cn^3.$$

Here in the second inequality we used an elementary bound

$$\log_+(\sum_{i=1}^n u_i) \le \sum_{i=1}^n \log(1+u_i)$$

valid for all $u_1,\dots,u_n \ge 0$, and the Cauchy–Schwarz inequality. The last inequality holds since $w_{i,j}$ is a normal random variable of variance at most 1.

To prove the bound for $\mathbb{E}\log_-^2 \det(W)^2$, assume that $a'_{i,i} \ge c/n$ for all $i \in [n]$. Set $A'' = \sqrt{n/c}A'_{1/2}$, so $a''_{i,i} \ge 1$, and let $W'' = A'' \odot G$. Then $\mathbb{E}\log_-^2 \det(W)^2 \le \mathbb{E}\log_-^2 \det(W'')^2 + 2n\log n$. We will prove the following estimate by induction:

$$\mathbb{E}\log_-^2 \det(W'')^2 \le c'n^2, \tag{7.15}$$

where the constant c' is chosen from the analysis of the one-dimensional case.

For $n = 1$ this follows from the inequality

$$\mathbb{E}\log_-^2(w_{1,1}+x) \le c', \tag{7.16}$$

which holds for all $x \in \mathbb{R}$. Assume that (7.15) holds for n. Denote by $\mathbb{E}_1$ the expectation with respect to $g_{1,1}$ and by $\mathbb{E}'$ the expectation with respect to $G^{(1)}$, which will denote the other entries of G. Denote by $D_{1,1}$ the minor of W'' corresponding to the entry $(1,1)$. Note that $D_{1,1} \ne 0$ a.s. Decomposing the determinant with respect to the first row, we obtain

$$\begin{aligned}\mathbb{E}\log_-^2 \det(W'')^2 &= \mathbb{E}'\left(\mathbb{E}_1\left[\log_-^2(a''_{1,1}g_{1,1}D_{1,1}+Y) \mid G^{(1)}\right]\right)\\ &= \mathbb{E}'\left(\mathbb{E}_1\left[\left(\log_-(a''_{1,1}g_{1,1}+\frac{Y}{D_{1,1}}) + \log_-(D_{1,1})\right)^2 \mid G^{(1)}\right]\right).\end{aligned}$$

Since $Y/D_{1,1}$ is independent of $g_{1,1}$, inequality (7.16) yields

$$\mathbb{E}_1\left(\log_-^2\left((a''_{1,1}g_{1,1}+\frac{Y}{D_{1,1}}\right) \mid G^{(1)}\right) \le c.$$

Therefore, by Cauchy–Schwarz inequality,

$$\begin{aligned}&\mathbb{E}\log_-^2 \det(W'')^2\\ &\le \mathbb{E}'\left(c' + 2\mathbb{E}_1\left[\log_-\left(a''_{1,1}g_{1,1} + \frac{Y}{D_{1,1}}\right) \mid G^{(1)}\right]\cdot \log_-(D_{1,1}) + \log_-^2(D_{1,1})\right)\\ &\qquad\qquad \le \left(\sqrt{c'} + \sqrt{\mathbb{E}'\log_-^2(D_{1,1})}\right)^2.\end{aligned}$$

By the induction hypothesis, $\mathbb{E}'\log_-^2(D_{1,1}) \le c'n^2$, so

$$\mathbb{E}\log_-^2 \det(W'')^2 \le c'(n+1)^2.$$

This proves the induction step, and thus completes the proof of Lemma 7.2. □

Theorem 1.4 is proved similarly, using this time Theorem 2.7 instead of Theorem 2.5, and taking into account the degradation of the Lipschitz constant due to the presence of b_n. We omit further details.

7.4. **Concentration far away from permanent.** Consider an approximately doubly stochastic matrix B with all entries of order $\Omega(n^{-1})$, which has some entries of order $\Omega(1)$. For such matrices the conditions of Theorem 1.4 are satisfied with $\delta, \kappa = 1$, so the Barvinok–Godsil-Gutman estimator is strongly concentrated around $\mathbb{E}\log\det^2(B_{1/2}\odot G)$. Yet, the second inequality of this Theorem reads

$$\log \mathrm{per}(B) \le \mathbb{E}\log\det^2(B_{1/2}\odot G) + C'n\log^{c'} cn,$$

which is too weak to obtain a subexponential deviation of the estimator from the permanent. However, the next lemma shows that the inequality above is sharp up to a logarithmic term. This means, in particular, that the Barvinok–Godsil-Gutman estimator for such matrices can be concentrated around a value, which is $\exp(cn)$ away from the permanent.

Lemma 7.3. *Let $\alpha > 0$, and let B be an $n \times n$ matrix with entries*

$$b_{i,j} = \begin{cases} \alpha/n, & \text{for } i \ne j;\\ 1, & \text{for } i = j.\end{cases}$$

There exist constants $\alpha_0, \beta > 0$ so that if $0 < \alpha < \alpha_0$ then

$$\liminf_{n\to\infty} \frac{1}{n}\left|\mathbb{E}\log\det(B_{1/2}\odot G)^2 - \log\mathbb{E}\det(B_{1/2}\odot G)^2\right| \ge \beta. \tag{7.17}$$

Proof. Recall that from (1.4), we have that for any fixed $\alpha < 1$, the random variable

$$\frac{1}{n} \left| \mathbb{E} \log \det(B_{1/2} \odot G)^2 - \log \det(B_{1/2} \odot G)^2 \right|$$

converges to 0 (in probability and a.s.). Since

$$\mathbb{E} \det(B_{1/2} \odot G)^2 = \mathrm{per}(B) \geq 1 \,,$$

it thus suffices to show that, with constants as in the statement of the lemma,

$$\liminf_{n\to\infty} \frac{1}{n} \log \det(B_{1/2} \odot G)^2 \leq -\beta \,, \quad a.s.. \tag{7.18}$$

We rewrite the determinant as a sum over permutations with ℓ fixed points. We then have

$$\det(B_{1/2} \odot G) = \sum_{\ell=0}^{n} \sum_{F \subset [n], |F|=\ell} \left(\prod_{i \in F} G_{ii} \right) \frac{(-1)^{\sigma(F)} M_F \alpha^{(n-\ell)/2}}{n^{(n-\ell)/2}} =: \sum_{\ell=0}^{n} A_\ell \,,$$

where M_F is the determinant of an $(n-\ell) \times (n-\ell)$ matrix with i.i.d. standard Gaussian entries, $EM_F^2 = (n-\ell)!$, $\sigma(F)$ takes values in $\{-1, 1\}$ and M_F is independent of $\prod_{i \in F} G_{ii}$. (Note that M_{F_1} is not independent of M_{F_2} for $F_1 \neq F_2$.)

Recall that

$$\binom{n}{\ell} \leq e^{nh(\ell_n)} \,, \tag{7.19}$$

where $\ell_n = \ell/n$ and h is the entropy function, $h(x) = -x \log x - (1-x) \log(1-x) \leq \log 2$.

We will need the following easy consequence of Chebyshev's inequality: for any $y > 0$,

$$\mathbb{P}\left(\left| \prod_{i=1}^{\ell} G_{ii} \right| \geq e^{-y\ell} \right) \leq (\mathbb{E}|G_{11}|)^\ell e^{y\ell} = \left(\sqrt{\frac{2}{\pi}} e^y \right)^\ell \,. \tag{7.20}$$

It is then clear that there exist $\delta_1, \delta_2 > 0$ so that, for any $\ell_n > (1 - \delta_1)$, one has

$$\binom{n}{\ell} \mathbb{P}\left(\left| \prod_{i=1}^{\ell} G_{ii} \right| \geq e^{-\delta_2 n} \right) \leq \frac{1}{n^3} \,. \tag{7.21}$$

Choose now $\delta_1' \leq \delta_1$ positive so that

$$\delta_2 > 3h(1 - \delta_1') \,, \tag{7.22}$$

which is always possible since $h(\cdot)$ is continuous and $h(1) = 0$.

We will show that we can find $\alpha_0 > 0$ such that for any $\alpha < \alpha_0$, for all n large and any ℓ,

$$\mathbb{P}\left(|A_\ell| \geq e^{-\delta_2 n/2}\right) \leq \frac{2}{n^3}\,. \tag{7.23}$$

This would imply (7.18) and conclude the proof of the lemma.

To see (7.23), we argue separately for $\ell_n \geq (1-\delta_1')$ and $\ell_n < (1-\delta_1')$. In either case, we start with the inequality

$$\begin{aligned}\mathbb{P}\left(|A_\ell| \geq e^{-\delta_2 n/2}\right) \\ \leq \binom{n}{\ell}\mathbb{P}\left(\left(\frac{\alpha}{n}\right)^{n(1-\ell_n)/2}\left(\prod_{i=1}^{\ell}|G_{ii}|\right)|M_{[\ell]}| \geq \binom{n}{\ell}^{-1} e^{-\delta_2 n/2}\right).\end{aligned} \tag{7.24}$$

Considering first $\ell_n \geq (1-\delta_1')$, we estimate the right side in (7.24) by

$$\begin{aligned}\binom{n}{\ell}\mathbb{P}\left(\prod_{i=1}^{\ell}|G_{ii}| \geq e^{-\delta_2 n}\right) \\ +\binom{n}{\ell}\mathbb{P}\left(\left(\frac{\alpha}{n}\right)^{n(1-\ell_n)/2}|M_{[\ell]}| \geq \binom{n}{\ell}^{-1} e^{\delta_2 n/2}\right).\end{aligned} \tag{7.25}$$

The first term in (7.25) is bounded by $1/n^3$ by our choice of parameters, see (7.21). To analyze the second term we use Chebyshev's inequality and the fact that $\alpha < 1$:

$$\begin{aligned}&\binom{n}{\ell}\mathbb{P}\left(\left(\frac{\alpha}{n}\right)^{n(1-\ell_n)/2}|M_{[\ell]}| \geq \binom{n}{\ell}^{-1} e^{\delta_2 n/2}\right) \\ \leq &\binom{n}{\ell}^3 e^{-\delta_2 n}\frac{\alpha^{n(1-\ell_n)}(n-\ell)!}{n^{n-\ell}} \\ \leq &e^{n[3h(\ell_n)-\delta_2]} \leq e^{3h(1-\delta_1')-\delta_2} \leq \frac{1}{n^3}\,,\end{aligned}$$

where the last inequality is due to (7.22). This completes the proof of (7.23) for $\ell_n \geq (1-\delta_1')$, for any $\alpha \leq 1$.

It remains to analyze the case $\ell_n < (1-\delta_1')n$. This is where the choice of α_0 will be made. Starting from (7.24) we have by Chebyshev's inequality

$$\begin{aligned}\mathbb{P}\left(|A_\ell| \geq e^{-\delta_2 n/2}\right) &\leq \binom{n}{\ell}^3 e^{\delta_2 n}\left(\frac{\alpha}{n}\right)^{n(1-\ell_n)} E|M_{[\ell]}|^2 \\ &\leq \alpha^{n(1-\ell_n)} e^{3n\log 2} \leq e^{n[3\log 2+\delta_1'\log\alpha]}\,.\end{aligned}$$

Choosing $\alpha_0 < 1$ such that $3\log 2 + \delta_1' \log \alpha_0 < 0$ shows that the last term is bounded by $1/n^3$ for large n, and completes the proof of the lemma. □

References

[1] T. W. Anderson, *The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities*, Proc. Amer. Math. Soc. **6** (1955), pp. 170–176.

[2] A. Barvinok, *Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor*, Random Structures Algorithms **14** (1999), pp. 29–61.

[3] I. Bezáková, D. Štefankovič, V. V. Vazirani and E. Vigoda, *Accelerating simulated annealing for the permanent and combinatorial counting problems*, SIAM J. Comput. **37** (2008), pp. 1429–1454.

[4] K. P. Costello and V. Vu, *Concentration of random determinants and permanent estimators*, SIAM J. Discrete Math. **23** (2009), pp. 1356–1371.

[5] S. Friedland, B. Rider and O. Zeitouni, *concentration of permanent estimators for certain large matrices*, Annals Appl. Prob **14** (2004), pp. 1359–1576.

[6] A. Frieze and M. Jerrum, *An analysis of a Monte Carlo algorithm for estimating the permanent*, Combinatorica **15** (1995), pp. 67–83.

[7] C. D. Godsil and I. Gutman, *on the matching polynomial of a graph*, in *Algebraic methods in graph theory I-II* (L. Lóvasz and V. T. Sós, eds.), North-Holland, Amsterdam (1981), pp. 67–83.

[8] N. R. Goodman, *Distribution of the determinant of a complex Wishart distributed matrix*, Annals Stat. **34** (1963), pp. 178–180.

[9] A. Guionnet and O. Zeitouni, *Concentration of the spectral measure for large matrices*, Elec. Comm. Probab. **5** (2000), pp. 119–136.

[10] M. Jerrum, A. Sinclair and E. Vigoda, *A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries*, J. ACM **51** (2004), pp. 671–697.

[11] E. Kalfoten and G. Villard, *On the complexity of computing determinants*, Comput. Complexity **13** (2004), pp. 91–130.

[12] N. Karmarkar, R. Karp, R. Lipton, L. Lovász and M. Luby, *A Monte Carlo algorithm for estimating the permanent*, SIAM J. Comput. **22** (1993), pp. 284–293.

[13] M. Ledoux, *The concentration of measure phenomenon*, American Math. Soc. (2001).

[14] N. Linial, A. Samorodnitsky and A. Wigderson, *A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents*, Combinatorica **20** (2000), pp. 545–568.

[15] S. Riemer, C. Schütt, *On the expectation of the norm of random matrices with non-identically distributed entries*, Electron. J. Probab. **18** (2013), no. 29, 13 pp.

[16] M. Rudelson, R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*. Adv. Math **218** (2008), pp. 600–633.

[17] M. Rudelson, R. Vershynin, *The smallest singular value of a random rectangular matrix*. Comm. Pure Appl. Math. **62** (2009), pp. 1707–1739.

[18] S. Szarek, *Spaces with large distance to ℓ_∞^n and random matrices*. Amer. J. Math. **112** (1990), pp. 899–942.

[19] L. Valiant, *The complexity of evaluating the permanent*, Theoret. Comput. Sci. **8** (1979), pp. 189–201.

[20] S. S. Wilks, *Moment generating operators for determinants of product moments in samples from a normal system*, Ann. Math **35** (1934), pp. 312–340.