On the permanent of a random symmetric matrix

Matthew Kwan[*] Lisa Sauermann[†]

Abstract

Let M_n denote a random symmetric $n \times n$ matrix, whose entries on and above the diagonal are i.i.d. Rademacher random variables (taking values ± 1 with probability $1/2$ each). Resolving a conjecture of Vu, we prove that the permanent of M_n has magnitude $n^{n/2+o(n)}$ with probability $1 - o(1)$. Our result can also be extended to more general models of random matrices. In our proof, we build on and extend some techniques introduced by Tao and Vu, studying the evolution of permanents of submatrices in a random matrix process.

Keywords: random symmetric matrix, permanent.
MSC Subject Classification: 60B20, 15A15.

1 Introduction

Two of the most basic matrix parameters are the *determinant* and the *permanent*: for an $n \times n$ matrix $M = (x_{i,j})_{i,j}$, define

$$\det(M) = \sum_{\pi \in S_n} \operatorname{sign}(\pi) \prod_{i=1}^{n} x_{i,\pi(i)} \quad \text{and} \quad \operatorname{per}(M) = \sum_{\pi \in S_n} \prod_{i=1}^{n} x_{i,\pi(i)}. \tag{1.1}$$

A central direction of research in probabilistic combinatorics and random matrix theory is to understand the determinant and permanent of different types of random matrices. For example, let A_n be an $n \times n$ matrix whose entries are i.i.d. Rademacher-distributed random variables, taking values ± 1 with probability $1/2$ each (this is often called a *random Bernoulli matrix*). A classical theorem of Komlós [19] (perhaps the foundational theorem in discrete random matrix theory) is that $\Pr(\det A_n = 0) = o(1)$ as $n \to \infty$. That is to say, A_n is *asymptotically almost surely* nonsingular. Since then, there has been intensive effort to refine our understanding of the singularity probability (see [18, 31, 32, 28, 5]), culminating in a recent breakthrough of Tikhomirov [36] proving that $\Pr(\det A_n = 0) = 2^{-n+o(n)}$. The problem of estimating the order of magnitude of $\det A_n$ has also received significant attention: Tao and Vu [31] proved that with probability $1 - o(1)$ we have $|\det A_n| = n^{n/2+o(n)}$, and later Nguyen and Vu [26] proved a central limit theorem for $\log |\det A_n|$ (see also [15, 16]).

Most of the above-mentioned results generalise readily to more general types of random matrices, where the entries are independently sampled from any subgaussian distribution. There has also been intensive interest in random matrices with dependence between the entries. Perhaps the most prominent examples are *symmetric* random matrices. Let M_n be the random matrix whose entries on and above the diagonal are independent Rademacher random variables, and the entries below the diagonal are chosen to make the matrix symmetric (equivalently, we can choose a random matrix A_n as above and condition on the event that A_n is symmetric). The study of random symmetric matrices has necessitated the development of new tools, but by now there is a fairly complete understanding of the determinant of a random symmetric matrix with Rademacher entries. The fact that $\Pr(\det M_n = 0) = o(1)$ was first proved by Costello, Tao and Vu [10] (see also [12]), and stronger estimates on $\Pr(\det M_n = 0)$ were obtained by several authors [24, 38, 13, 9]. It is also known that with probability $1 - o(1)$ we have $|\det M_n| = n^{n/2+o(n)}$ (this follows from work on the least singular value of M_n due to Nguyen [25] and Rudelson [38], together with

[*] Department of Mathematics, Stanford University, Stanford, CA. Email: mattkwan@stanford.edu. Research supported by NSF Award DMS-1953990.

[†] School of Mathematics, Institute for Advanced Study, Princeton, NJ. Email: lsauerma@mit.edu. Research supported by NSF Grant CCF-1900460 and NSF Award DMS-1953772. Part of this work was completed while this author was a Szegő Assistant Professor at Stanford University.

Wigner's celebrated semicircle law [43, 44]), and a central limit theorem for $\log|\det M_n|$ was proved by Bourgade and Mody [4] (see also [34]).

It is widely believed that for all the above-mentioned theorems concerning determinants of random matrices (symmetric or not), there should be analogous theorems for permanents. However, the permanent appears to be a much more challenging parameter to study. For example, while the determinant encodes information about linear dependence and can be interpreted as the (signed) volume of a certain parallelepiped, it only seems possible to attack the permanent from a "combinatorial" point of view, directly considering the formal definition in (1.1). The fact that the permanent is harder to study is maybe not surprising, since (in contrast to the determinant) the permanent of a matrix is #P-hard to compute (as was famously proved by Valiant [37]). Even the analogue of the singularity problem, to show that $\Pr(\operatorname{per} A_n = 0) = o(1)$, was open for a long time. In 2009, Tao and Vu [33] finally resolved this problem and also estimated the typical magnitude of $\operatorname{per} A_n$: namely, they proved that asymptotically almost surely $|\operatorname{per} A_n| = n^{n/2+o(n)}$. Perhaps surprisingly, permanents of random matrices turn out to be of interest in quantum computing: Aaronson and Arkhipov [2] proved that quantum computers cannot be efficiently simulated by classical computers, conditional on a conjecture which strengthens the aforementioned Tao–Vu permanent theorem (see also [1, 11, 22]).

In this paper we study the permanent of random *symmetric* matrices. More precisely, we estimate the typical magnitude of the permanent of a random symmetric matrix with Rademacher entries.

Theorem 1.1. *Asymptotically almost surely,* $|\operatorname{per} M_n| = n^{n/2+o(n)}$.

The study of the permanent of a random symmetric matrix seems to have first been explicitly suggested by Tao and Vu (see [33, Remark 1.6]). They observed that their arguments for the permanent of a (not necessarily symmetric) random matrix "do not seem to easily yield any non-trivial result for the permanent of a random *symmetric* Bernoulli matrix". The statement of Theorem 1.1 has been conjectured by Vu in 2009 (see [39]). He also mentioned the conjecture in a recent survey [40, Conjecture 6.11], and described it as "the still missing piece of the picture" regarding determinants and permanents of random discrete matrices.

Theorem 1.1 is actually a combination of two different results. First, the following proposition gives an upper bound on $|\operatorname{per} M_n|$.

Proposition 1.2. *For any $\varepsilon > 0$, if n is sufficiently large with respect to ε, we have*

$$\Pr\Big(|\operatorname{per} M_n| \geq n^{n/2+\varepsilon n}\Big) \leq n^{-\varepsilon n}.$$

Proposition 1.2 is easily proved using an estimate for $\mathbb{E}[(\operatorname{per} M_n)^2]$ and Markov's inequality; see Section 2. The main role of this paper is to prove the following lower bound on $|\operatorname{per} M_n|$.

Theorem 1.3. *There is a positive constant $c > 0$ such that for any $\varepsilon > 0$ the following holds. If n is sufficiently large with respect to ε, we have*

$$\Pr\Big(|\operatorname{per} M_n| \leq n^{n/2-\varepsilon n}\Big) \leq n^{-c}.$$

We remark that the constant c in Theorem 1.3 can be made explicit. For example, $c = 1/150$ certainly suffices, though this can be improved substantially simply by optimising constants throughout the proof. However, without new ideas our methods do not seem to be capable of proving the statement of Theorem 1.3 with any $c \geq 1/2$. This state of affairs is essentially the same as for the non-symmetric case previously considered by Tao and Vu [33].

Of course, Theorem 1.3 also shows that the probability of having $\operatorname{per} M_n = 0$ is polynomially small. Before this paper no nontrivial bounds for this probability were known, except when $n = 2^m - 1$ with $m \in \mathbb{N}$, where for elementary number-theoretic reasons it is actually impossible to have $\operatorname{per} M_n = 0$ (see [30][1]).

Finally, we remark that the methods used to prove Theorem 1.1 are quite robust, and analogous theorems can be proved for much more general distributions. For example, consider any fixed real probability

[1] Actually, this result is part of an extensive body of research concerning permanents of (non-random) matrices with ±1 entries; see for example [41, 6, 7, 8, 30, 42, 20, 29, 21].

distributions μ and ν, and let $M_n^{\mu,\nu}$ be the random symmetric matrix whose diagonal entries have distribution ν and whose off-diagonal entries have distribution μ (and whose entries on and above the diagonal are mutually independent). With some fairly mild assumptions on μ and ν it is a routine matter to prove an analogue of Proposition 1.2 for $M_n^{\mu,\nu}$, and in Section 8 we sketch how to make some minor adaptations to the proof of Theorem 1.3 to obtain a version that holds for very general distributions (we only require that μ has nontrivial support). In particular, one can prove an analogue of Theorem 1.1 for random symmetric Gaussian matrices such as the Gaussian Orthogonal Ensemble (GOE).

Notation. In this paper, we use the notation $\mathbb{N} = \{1, 2, \dots\}$ for the positive integers. All logarithms are to base e. For functions $f : \mathbb{N} \to \mathbb{R}$ and $g : \mathbb{N} \to \mathbb{R}_{>0}$, we write $f = o(g)$ if $f(n)/g(n) \to 0$ as $n \to \infty$.

2 Proof of Proposition 1.2: the second moment of the permanent

In this section we provide the simple proof of Proposition 1.2. It will be an immediate consequence of Markov's inequality and the following lemma.

Lemma 2.1. $\mathbb{E}[(\operatorname{per} M_n)^2] \le n^{n+o(n)}$.

Proof. Write $x_{i,j}$ for the (i,j) entry of M_n. For a permutation $\pi \in S_n$, let $X_\pi = \prod_{i=1}^n x_{i,\pi(i)}$, so that $\operatorname{per} M_n = \sum_\pi X_\pi$. It will not be necessary for the proof, but we remark that

$$\mathbb{E}X_\pi = \begin{cases} 1 & \text{if } \pi \text{ consists only of 2-cycles,} \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, let $I_\pi \subseteq \{1, \dots, n\}$ be the set of indices which appear in 2-cycles of π, and let F_π be the family of sets $\{i, \pi(i)\}$, for $i \notin I_\pi$. Then for two permutations $\pi, \pi' \in S_n$, we have

$$\mathbb{E}[X_\pi X_{\pi'}] = \begin{cases} 1 & \text{if } (I_\pi, F_\pi) = (I_{\pi'}, F_{\pi'}), \\ 0 & \text{otherwise.} \end{cases}$$

For $k = 0, \dots, n$, let $\mathcal{Q}_k$ be the set of all permutations $\pi \in S_n$ satisfying $|I_\pi| = k$, and note that $|\mathcal{Q}_k| \le \binom{n}{k} k^{k/2}(n-k)! \le 2^n k^{k/2} n^{n-k}$. Indeed, for any choice of a set $I \subseteq \{1, \dots, n\}$ of size k, there are at most $k^{k/2}$ ways to partition I into 2-cycles, and at most $(n-k)!$ ways to choose a permutation of $\{1, \dots, n\} \setminus I$.

Now, for any $0 \le k \le n$ and any $\pi \in \mathcal{Q}_k$, there are at most $2^{n-k} k^{k/2}$ choices of π' satisfying $(I_\pi, F_\pi) = (I_{\pi'}, F_{\pi'})$. Indeed, for such π', the restriction of π' to I_π must be a permutation of I_π consisting only of 2-cycles (so there are at most $k^{k/2}$ ways to choose this restriction), and for each $i \notin I_\pi$ we must have $\pi'(i) \in \{\pi(i), \pi^{-1}(i)\}$. It follows that

$$\mathbb{E}[(\operatorname{per} M_n)^2] = \sum_{\pi, \pi' \in S_n} \mathbb{E}X_\pi X_{\pi'} \le \sum_{k=0}^n |\mathcal{Q}_k| \cdot 2^{n-k} k^{k/2} \le \sum_{k=0}^n 4^n k^k n^{n-k} \le n^{n+o(n)},$$

as claimed. □

Proof of Proposition 1.2. Let $\varepsilon > 0$, and suppose that n is sufficiently large such that the $o(n)$-term in Lemma 2.1 is at most εn. Then we have $\mathbb{E}[(\operatorname{per} M_n)^2] \le n^{n+\varepsilon n}$ and consequently by Markov's inequality

$$\Pr\Big(|\operatorname{per} M_n| \ge n^{n/2+\varepsilon n}\Big) = \Pr\big((\operatorname{per} M_n)^2 \ge n^{n+2\varepsilon n}\big) \le \frac{\mathbb{E}[(\operatorname{per} M_n)^2]}{n^{n+2\varepsilon n}} \le n^{-\varepsilon n},$$

as desired. □

3 Structure of the proof of Theorem 1.3

The rest of this paper is devoted to the proof of Theorem 1.3. In this section we outline the high-level strategy of the proof, stating two key lemmas and deducing Theorem 1.3 from them.

We couple the distributions of the matrices M_n for all $n \in \mathbb{N}$ by viewing each M_n as containing the first n rows and columns of an infinite random symmetric matrix (with Rademacher entries). Say that subsets of a given ground set are *complement-disjoint* if their complements are disjoint. For $A, B \subseteq \{1, \dots, n\}$, let $M_n[A, B]$ be the submatrix of M_n consisting of the rows in A and the columns in B. For $\lambda \geq 0$, we say that a matrix is λ-*heavy* if its permanent has absolute value at least λ.

The following lemma shows that with high probability there exists a heavy submatrix of M_n consisting of almost all the rows and columns of M_n, and moreover we have some control over which rows and columns are not included. Roughly speaking, it is proved by studying how permanents of submatrices evolve in the sequence of random matrices $M_1, \dots, M_n$.

Lemma 3.1. *There is a positive constant $c > 0$ such that for any $\varepsilon > 0$ the following holds. Let $n \in \mathbb{N}$ be sufficiently large with respect to ε, and let $L = \lfloor (\log n)/10 \rfloor$. Let X and Y be disjoint subsets of $\{1, \dots, n\}$ with sizes $|X| = L$ and $|Y| = 3L$. Then with probability at least $1 - (1/4) \cdot n^{-c}$ there is a set B satisfying $|B| = n - L$ and $\{1, \dots, n\} \setminus Y \subseteq B \subseteq \{1, \dots, n\}$, such that $M_n[\{1, \dots, n\} \setminus X, B]$ is $n^{(1-\varepsilon)n/2}$-heavy.*

By applying Lemma 3.1 with various different choices of X and Y, we can obtain many heavy submatrices $M_n[A_1, B_1], \dots, M_n[A_m, B_m]$ such that the sets $A_1, \dots, A_m, B_1, \dots, B_m$ are complement-disjoint. The next lemma states that in such a situation, if we sample M_{n+1} by adding a random row and column to M_n, then a large proportion of our submatrices can be transformed into larger submatrices without losing much of their heaviness. We will apply this lemma repeatedly, each step decreasing by 1 the number of rows and columns that our submatrices are missing.

Lemma 3.2. *Let $m \in \mathbb{N}$ be sufficiently large. Let $\lambda > 0$, let $1 \leq L < n$ be integers, and let $A_1, \dots, A_m, B_1, \dots, B_m$ be complement-disjoint subsets of $\{1, \dots, n\}$ of size $n - L$. Let us condition on an outcome of M_n such that all the submatrices $M_n[A_\ell, B_\ell]$, for $\ell = 1, \dots, m$, are λ-heavy.*

Then, with probability at least $1 - m^{-1/24}$, for $m' = \lceil m/36 \rceil$ there are complement-disjoint subsets $A'_1, \dots, A'_{m'}, B'_1, \dots, B'_{m'} \subseteq \{1, \dots, n+1\}$ of size $n - L + 2$, such that for all $\ell = 1, \dots, m'$ the submatrices $M_{n+1}[A'_\ell, B'_\ell]$ are $\lambda/(4n^4)$-heavy.

We now show how to deduce Theorem 1.3 from Lemmas 3.1 and 3.2.

Proof of Theorem 1.3. Choose an absolute constant $0 < c < 1/50$, such that the statement in Lemma 3.1 is satisfied. Fix $\varepsilon > 0$.

Let $L = \lfloor (\log n)/10 \rfloor$ and $m = \lfloor (n - L)/(4L) \rfloor$, and consider disjoint sets $X_1, \dots, X_m, Y_1, \dots, Y_m \subseteq \{1, \dots, n - L\}$ with $|X_1| = \dots = |X_m| = L$ and $|Y_1| = \dots = |Y_m| = 3L$.

For each $\ell = 1, \dots, m$, we apply Lemma 3.1 to the subsets $X_\ell, Y_\ell \subseteq \{1, \dots, n - L\}$. Each application fails with probability at most $n^{-c}/4$, so it follows from Markov's inequality (see for example Lemma 4.8) that with probability at least $1 - (1/2) \cdot n^{-c}$, at least $m/2$ applications succeed. That is to say, with $m' = \lceil m/2 \rceil$ and $\lambda = (n-L)^{(1-\varepsilon)(n-L)/2}$, we obtain complement-disjoint sets $A_1, \dots, A_{m'}, B_1, \dots, B_{m'} \subseteq \{1, \dots, n-L\}$ of size $n - 2L$ such that for all $\ell = 1, \dots, m'$, the matrices $M_{n-L}[A_\ell, B_\ell]$ are λ-heavy. Note that if n is sufficiently large with respect to ε, then $m' \geq n^{9/10}$ and $\lambda \geq n^{n/2 - (3/4)\varepsilon n}$.

Now, we wish to iteratively apply Lemma 3.2, L times in total. After each application, m' decreases by a factor of $36 < e^4$, so after $L = \lfloor (\log n)/10 \rfloor$ steps the value of m' will still be at least $\sqrt{n}$. Each of the L applications of Lemma 3.2 succeeds with probability at least $1 - n^{-1/48}$. Thus, with probability at least $1 - L \cdot n^{-1/48} \geq 1 - (1/2) \cdot n^{-c}$ (for sufficiently large n) we can indeed apply the lemma L times. In the end we obtain subsets $A', B' \subseteq \{1, \dots, n\}$ of size n such that the matrix $M_n[A', B']$ is λ'-heavy, where

$$\lambda' = \frac{\lambda}{4(n-L)^4 \cdot 4(n-L+1)^4 \cdots 4(n-1)^4} \geq \frac{\lambda}{(4n^4)^L} \geq \frac{\lambda}{n^{5L}} \geq n^{n/2 - (3/4)\varepsilon n - \log n/2} \geq n^{n/2 - \varepsilon n}$$

(again assuming that n is sufficiently large with respect to ε). But note that we must have $A' = B' = \{1, \dots, n\}$, so this means that M_n itself is λ'-heavy. In summary, with probability at least $1 - n^{-c}$ we have $|\operatorname{per} M_n| \geq \lambda' \geq n^{n/2 - \varepsilon n}$, as desired. □

We remark that the overall structure of our proof is similar to the work of Tao and Vu [33] on permanents of (not necessarily symmetric) random matrices. Indeed, Tao and Vu's proof can also be broken up into two parts analogous to Lemmas 3.1 and 3.2. However, in Tao and Vu's setting, all entries of the random matrix are independent, allowing them to expose the entries row by row. After exposing k rows, they consider $k \times k$ submatrices that consist of all the k exposed rows (and of k of the n columns). When exposing the k-th row, the permanent of any such $k \times k$ submatrix can be described as a linear polynomial in some of the entries of the new row, where the coefficients are given by the permanents of certain $(k-1) \times (k-1)$ submatrices in the first $k-1$ rows. In contrast, in our setting with the random symmetric matrix M_n, we are forced to expose the entries of our matrix in a different way: at the k-th step we reveal the entries in M_k that are not present in M_{k-1} (that is, we add a new random row and column, with equal entries, to the matrix considered so far[2]). Since there is only one $k \times k$ submatrix in M_k (namely M_k itself), in our setting we also need to consider the permanents of (substantially) smaller submatrices of M_k.

This more intricate strategy introduces significant challenges. Most notably, the permanents of the submatrices of M_k are described by *quadratic* polynomials in the new matrix entries, where the coefficients depend on the permanents of certain submatrices of M_{k-1} (this is in contrast to Tao and Vu's setting, where the permanents are described by linear polynomials in the entries of the new row). This necessitates the use of some more sophisticated probabilistic tools. Furthermore, there can be certain types of cancellations within these quadratic polynomials, which are not possible for the linear polynomials in the Tao–Vu setting. For example, even if all submatrices of M_{k-1} have non-zero permanent, it can happen that the polynomial describing the permanent of some submatrix of M_k has only very few nonzero coefficients. Handling these types of cancellations requires key new ideas.

Organization of the rest of the paper. Lemma 3.1 will be proved in Section 6, and Lemma 3.2 will be proved in Section 7. As preparation, in Section 4 we collect some probabilistic tools that we will use in the proofs, and in Section 5 we collect some lemmas that can be obtained by studying permanent expansion formulas.

4 Probabilistic tools

This section collects some theorems and simple facts that will be needed for proving Lemmas 3.1 and 3.2. We start with some basic anti-concentration estimates for linear forms. The first of these is the famous Erdős–Littlewood–Offord inequality (see for example [35, Corollary 7.8]).

Theorem 4.1. *Let $t \geq 1$ be a real number, and let f be a linear polynomial in n variables, in which at least m degree-1 coefficients have absolute value at least r. Then for uniformly random $\boldsymbol{\xi} \in \{-1,1\}^n$ we have*

$$\Pr(|f(\boldsymbol{\xi})| \leq t \cdot r) \leq (\lceil t \rceil + 1) \cdot \binom{m}{\lfloor m/2 \rfloor} \cdot 2^{-m} \leq \frac{3t}{\sqrt{m}}.$$

We will also need the following very easy fact.

Fact 4.2. *Let f be a linear polynomial in n variables, which has at least one coefficient with absolute value at least r. Then for uniformly random $\boldsymbol{\xi} \in \{-1,1\}^n$ we have*

$$\Pr(|f(\boldsymbol{\xi})| < r) \leq \frac{1}{2}.$$

Proof. First, suppose that the constant coefficient of f has absolute value at least r, and suppose without loss of generality that $f(\boldsymbol{\xi}) = a_1\xi_1 + \cdots + a_n\xi_n + c$ with $c \geq r$. Then, by symmetry we have that $\Pr(a_1\xi_1 + \cdots + a_n\xi_n < 0) \leq 1/2$ and therefore $\Pr(f(\boldsymbol{\xi}) < r) \leq 1/2$.

Otherwise, for some $i \in \{1, \ldots, n\}$, the coefficient of ξ_i in $f(\boldsymbol{\xi})$ has absolute value at least r, and suppose without loss of generality that $i = 1$. Condition on any outcomes of the variables $\xi_2, \ldots, \xi_n$, and observe that then we can have $|f(\boldsymbol{\xi})| < r$ for at most one of the two possible outcomes of ξ_1. □

We will also need counterparts of both the above statements for quadratic polynomials. The quadratic counterpart of Fact 4.2 is again easy to prove.

[2] This type of exposure is standard in the study of symmetric random matrices (see for example [10]).

Fact 4.3. *Let f be a quadratic polynomial in n variables, which has at least one multilinear degree-2 coefficient with absolute value at least r. Then for uniformly random $\boldsymbol{\xi} \in \{-1,1\}^n$ we have*

$$\Pr(|f(\boldsymbol{\xi})| < r) \leq \frac{3}{4}.$$

Proof. We may assume that f is multilinear (every term of the form ξ_i^2 can be replaced by the constant 1 without changing the behaviour of $f(\boldsymbol{\xi})$). Suppose without loss of generality that the coefficient a_{12} of $\xi_1\xi_2$ satisfies $|a_{12}| \geq r$, and write $f(\xi_1,\dots,\xi_n) = \xi_1 \cdot (a_{12}\xi_2 + g(\xi_3,\dots,\xi_n)) + h(\xi_2,\dots,\xi_n)$. Conditioning on any outcomes of $\xi_3,\dots,\xi_n$, with probability at least 1/2 we have $|a_{12}\xi_2 + g(\xi_3,\dots,\xi_n)| \geq r$. Then, conditioning on such an outcome of ξ_2, we have $|f(\boldsymbol{\xi})| \geq r$ with probability at least 1/2. □

It is more delicate to generalise the Erdős–Littlewood–Offord inequality to quadratic polynomials. For a multilinear quadratic polynomial f in the variables $x_1,\dots,x_n$ and for a real number $r > 0$, let $G^{(r)}(f)$ be the graph with vertex set $\{1,\dots,n\}$ having an edge ij whenever the coefficient of x_ix_j in f has absolute value at least r. Let $\nu(G)$ be the matching number[3] of a graph G. The following is a special case of a theorem proved by Meka, Nguyen and Vu [23, Theorem 1.6].

Theorem 4.4. *Let $r > 0$, let f be a multilinear quadratic polynomial in n variables, and let $\nu = \nu(G^{(r)}(f)) \geq 3$. Then for uniformly random $\boldsymbol{\xi} \in \{-1,1\}^n$ we have*

$$\Pr(|f(\boldsymbol{\xi})| \leq r) \leq \frac{(\log \nu)^C}{\nu^{1/2}},$$

where C is an absolute constant.

The following concentration inequality is a special case of the Azuma–Hoeffding martingale concentration inequality, and is sometimes known as McDiarmid's inequality (see for example [35, Lemma 1.34]).

Lemma 4.5. *Let $c > 0$ and let X be a random variable defined in terms of independent random variables $\xi_1,\dots,\xi_n$, having the property that varying any individual ξ_i affects the value of X by at most c. Then for any $t \geq 0$ we have*

$$\Pr(|X - \mathbb{E}X| \geq t) \leq 2e^{-t^2/(2nc^2)}.$$

The next inequality is a one-sided version of the Azuma–Hoeffding inequality for supermartingales (see [33, Lemma 2.3]).

Lemma 4.6. *Let $c > 0$. In a probability space, let $Z_1,\dots,Z_n$ be a sequence of random objects, and let $W_1,\dots,W_n$ be a sequence of random variables, such that for each k, all of $Z_1,\dots,Z_k,W_1,\dots,W_k$ are fully determined by Z_k, and such that $|W_{k+1} - W_k| \leq c$ for all $k = 1,\dots,n-1$. Suppose that the supermartingale property $\mathbb{E}[W_{k+1}|Z_k] \leq W_k$ is satisfied for $k = 1,\dots,n-1$. Then for any $t > 0$ we have*

$$\Pr(W_n - W_1 \geq t) \leq e^{-t^2/(2nc^2)}.$$

Recall that for $0 < p < 1$, a Bernoulli random variable $\chi \sim \mathrm{Ber}(p)$ is a random variable taking values 0 and 1 with $\Pr(\chi = 1) = p$ and $\Pr(\chi = 0) = 1 - p$. The following lemma is a version of the Chernoff concentration bound for sums of Bernoulli random variables (see for example [3, Theorem A.1.4]).

Lemma 4.7. *Let $\chi_1,\dots,\chi_m$ be independent Bernoulli random variables, where for each $i = 1,\dots,m$ we have $\chi_i \sim \mathrm{Ber}(p_i)$ for some $0 < p_i < 1$. Then for any $t > 0$ the sum $X = \chi_1 + \dots + \chi_m$ satisfies*

$$\Pr(X - \mathbb{E}X > t) < e^{-2t^2/n} \quad \textit{and} \quad \Pr(X - \mathbb{E}X < -t) < e^{-2t^2/n}.$$

Sometimes we will encounter random variables that *stochastically dominate* a sum of Bernoulli random variables (we say that a random variable X stochastically dominates another random variable Y if there is a coupling of X and Y such that we always have $X \geq Y$). For example, consider a random process with n steps where each step satisfies a certain property with probability at least 1/2, even when conditioning on any outcome of the previous steps. Then the number X of steps with this property stochastically dominates a sum of n independent $\mathrm{Ber}(1/2)$ random variables. Denoting this sum by Y, we therefore have

[3] The matching number of a graph G is the largest number ν such that one can find ν disjoint edges in G.

$\Pr(X-(n/2)<-t)\le\Pr(Y-(n/2)<-t)<e^{-2t^2/n}$ for any $t>0$ by Lemma 4.7. Hence we can use Lemma 4.7 to show that a random variable is very likely reasonably large if it stochastically dominates a sum of Bernoulli random variables.

Finally, the following lemma is an easy consequence of Markov's inequality (see, for example, [33, Lemma 2.1]).

Lemma 4.8. *Let $1>p>q>0$, and let $E_1,\dots,E_m$ be events (not necessarily independent), each of which occurs with probability at least p. Then the probability that at least qm of the events $E_1,\dots,E_m$ occur simultaneously is at least $(p-q)/(1-q)$.*

5 Permanent expansion formulas

Just as for the determinant, it is possible to expand the permanent of a matrix in terms of permanents of submatrices. Below we record two such expansions, which we will use in the proofs of Lemmas 3.1 and 3.2.

Fact 5.1. *Let M be an $n\times n$ matrix. Add a new row $(x_1,\dots,x_n)$ to obtain an $(n+1)\times n$ matrix M'. Then for any subsets $A,B\subseteq\{1,\dots,n\}$ with $|B|=|A|+1$, we have*

$$\operatorname{per}M'[A\cup\{n+1\},B]=\sum_{i\in B}x_i\operatorname{per}M[A,B\setminus\{i\}].$$

For a matrix M, let $M^{(i,j)}$ be the submatrix of M obtained by removing row i and column j.

Fact 5.2. *Let M be an $n\times n$ matrix. Add a new row $(x_1,\dots,x_n,z)$ and a new column $(y_1,\dots,y_n,z)$ to obtain an $(n+1)\times(n+1)$ matrix M'. Then for any subsets $A,B\subseteq\{1,\dots,n\}$ with $|A|=|B|$, we have*

$$\operatorname{per}M'[A\cup\{n+1\},B\cup\{n+1\}]=z\operatorname{per}M[A,B]+\sum_{i\in A,j\in B}x_jy_i\operatorname{per}M[A,B]^{(i,j)}.$$

We will use Fact 5.1 in combination with the linear anti-concentration inequalities in Fact 4.2 and Theorem 4.1, and we will use Fact 5.2 in combination with the quadratic anti-concentration inequalities in Fact 4.3 and Theorem 4.4

Observe in particular that the formula in Fact 5.2 gives an expression for the permanent of a $(k+1)\times(k+1)$ matrix in terms of permanents of $(k-1)\times(k-1)$ submatrices (and one $k\times k$ submatrix). This means that, for example, when we add a new row and column to a matrix, the size of the largest submatrix with nonzero permanent can increase by two. This observation will be crucial for the proof of Lemma 3.2.

In the proof of Lemma 3.2, we will apply Fact 5.2 with the symmetric matrices $M=M_n$ and $M'=M_{n+1}$. That is to say, $(x_1,\dots,x_n,z)=(y_1,\dots,y_n,z)$, so the formula in Fact 5.2 can be interpreted as a quadratic polynomial in $x_1,\dots,x_n$ (after conditioning on the value of z). In order to apply Fact 4.3 to this polynomial, we need this polynomial to have a multilinear degree-2 coefficient with large absolute value. For this, it suffices that $\operatorname{per}M[A,B]^{(i,j)}+\operatorname{per}M[A,B]^{(j,i)}$ has large absolute value for some $i\ne j$ (with $i,j\in A\cap B$). The following lemma will be useful for ensuring this condition.

Lemma 5.3. *Let M be an $n\times n$ matrix and let $A,B\subseteq\{1,\dots,n\}$ be subsets with $|A|=|B|$ such that $M[A,B]$ is λ-heavy. Suppose we are given an element $a\in B\setminus A$ and distinct elements $b_1,b_2\in A\setminus B$. Then there are distinct $i,j\in\{a,b_1,b_2\}$ such that*

$$\left|\operatorname{per}M[A',B']^{(i,j)}+\operatorname{per}M[A',B']^{(j,i)}\right|\ge\lambda/2,$$

where $A'=A\cup\{a\}$ and $B'=(B\setminus\{a\})\cup\{i,j\}$.

Proof. Suppose without loss of generality that $\operatorname{per}M[A,B]\ge\lambda$. If we have

$$\operatorname{per}M[(A\setminus\{b_s\})\cup\{a\},(B\setminus\{a\})\cup\{b_s\}]\ge-\lambda/2$$

for some $s\in\{1,2\}$, then we can take $i=b_s$ and $j=a$. Indeed, then we have $A'=A\cup\{a\}$ and $B'=B\cup\{b_s\}$, and obtain $\operatorname{per}M[A',B']^{(i,j)}+\operatorname{per}M[A',B']^{(j,i)}\ge(-\lambda/2)+\lambda=\lambda/2$.

Otherwise, if there is no such $s\in\{1,2\}$, we can take $i=b_1$ and $j=b_2$. Then we have $A'=A\cup\{a\}$ and $B'=(B\setminus\{a\})\cup\{b_1,b_2\}$, and obtain $\operatorname{per}M[A',B']^{(i,j)}+\operatorname{per}M[A',B']^{(j,i)}<(-\lambda/2)+(-\lambda/2)=-\lambda$. □

We end this section with two simple lemmas that illustrate how to apply Facts 5.1 and 5.2 and Lemma 5.3 to "grow" heavy minors in a random symmetric matrix. Recall that M_n is a random symmetric matrix, and that M_{n-1} contains the first $n-1$ rows and columns of M_n.

Lemma 5.4. *Consider $A, B \subseteq \{1, \dots, n-1\}$ with $|A| = |B|$, and fix any nonempty $I \subseteq \{1, \dots, n-1\} \setminus B$. Consider any outcome M of M_{n-1} such that $M[A, B]$ is λ-heavy, for some $\lambda > 0$. Then*

$$\Pr\big(M_n[A \cup \{n\}, B \cup \{i\}] \text{ is } \lambda\text{-heavy for some } i \in I \,\big|\, M_{n-1} = M\big) \geq 1 - 2^{-|I|}.$$

Proof. We condition on $M_{n-1} = M$. Let $x_1, \dots, x_n$ be the entries in the last row of M_n. Let us also condition on any outcome of the variables x_b for $b \in B$. Now, by Fact 5.1, for each $i \in I$ we have

$$\operatorname{per} M_n[A \cup \{n\}, B \cup \{i\}] = x_i \operatorname{per} M_{n-1}[A, B] + \sum_{b \in B} x_b \operatorname{per} M_{n-1}[A, (B \cup \{i\}) \setminus \{b\}].$$

Since $|\operatorname{per} M_{n-1}[A, B]| \geq \lambda$, each $i \in I$ satisfies the desired condition $|\operatorname{per} M_n[A \cup \{n\}, B \cup \{i\}]| \geq \lambda$ with probability at least $1/2$, and (by our conditioning on the variables x_b for $b \in B$) this happens independently for all $i \in I$. □

Lemma 5.5. *Consider $A, B \subseteq \{1, \dots, n-1\}$ with $|A| = |B|$, and consider an outcome M of M_{n-1} such that $M[A, B]$ is λ-heavy, for some $\lambda > 0$. Then for any $a \in B \setminus A$ and any distinct $b_1, b_2 \in A \setminus B$, we can choose distinct $i, j \in \{a, b_1, b_2\}$ such that*

$$\Pr\big(M_n[A \cup \{a, n\}, (B \setminus \{a\}) \cup \{i, j, n\}] \text{ is } (\lambda/2)\text{-heavy} \,\big|\, M_{n-1} = M\big) \geq \frac{1}{4}.$$

Proof. Let us condition on $M_{n-1} = M$. By Lemma 5.3, we can choose distinct $i, j \in \{a, b_1, b_2\}$ such that

$$\left| \operatorname{per} M_{n-1}[A', B']^{(i,j)} + \operatorname{per} M_{n-1}[A', B']^{(j,i)} \right| \geq \lambda/2, \tag{5.1}$$

where $A' = A \cup \{a\}$ and $B' = (B \setminus \{a\}) \cup \{i, j\}$. Note that $\{i, j\} \subseteq \{a, b_1, b_2\} \subseteq A'$ and that clearly $\{i, j\} \subseteq B'$.

Now, $\operatorname{per} M_n[A \cup \{a, n\}, (B \setminus \{a\}) \cup \{i, j, n\}] = \operatorname{per} M_n[A' \cup \{n\}, B' \cup \{n\}]$, so it suffices to show that with probability at least $1/4$ we have $|\operatorname{per} M_n[A' \cup \{n\}, B' \cup \{n\}]| \geq \lambda/2$.

Let $(x_1, \dots, x_{n-1}, z)$ be the random entries of the last row (and the last column) of M_n. By Fact 5.2, we have

$$\operatorname{per} M_n[A' \cup \{n\}, B' \cup \{n\}] = z \operatorname{per} M_{n-1}[A', B'] + \sum_{k \in A', \ell \in B'} x_k x_\ell \operatorname{per} M_{n-1}[A', B']^{(k,\ell)}.$$

Note that this is a quadratic polynomial in the variables $x_1, \dots, x_{n-1}, z$, and the coefficient of $x_i x_j$ is precisely $\operatorname{per} M_{n-1}[A', B']^{(i,j)} + \operatorname{per} M_{n-1}[A', B']^{(j,i)}$. Recalling $i \neq j$ and (5.1), Fact 4.3 now implies that $\Pr(|\operatorname{per} M_n[A' \cup \{n\}, B' \cup \{n\}]| < \lambda/2) \leq 3/4$. This finishes the proof of Lemma 5.5. □

6 Proof of Lemma 3.1: growing a single heavy submatrix

In this section we prove Lemma 3.1. Recall that $L = \lfloor (\log n)/10 \rfloor$ and that $X, Y \subseteq \{1, \dots, n\}$ are disjoint subsets with $|X| = L$ and $|Y| = 3L$. By reordering the rows and columns, we can assume without loss of generality that $X = \{1, \dots, L\}$ and $Y = \{n - 3L + 1, \dots, n\}$.

Lemma 3.1 will be a consequence of the following two lemmas. The first of these lemmas is itself a weaker version of Lemma 3.1 (it also produces a heavy submatrix, but with less control over where it lies, and not with dimensions as close to $n \times n$).

Lemma 6.1. *For any fixed $0 < \delta < 1/16$, the following holds for all integers $n \in \mathbb{N}$ that are sufficiently large with respect to δ. Let $\lambda = n^{(1/2 - 8\delta)n}$ and suppose that $R \in \mathbb{N}$ satisfies $\delta n \leq R \leq 2\delta n$. Then with probability at least $1 - e^{-\delta^2 n}$ there is a subset $B \subseteq \{1, \dots, n\}$ of size $n - R$ such that $M_n[\{R+1, \dots, n\}, B]$ is λ-heavy.*

To prove Lemma 6.1 we adapt an argument in Tao and Vu's work [33], simultaneously tracking the propagation and growth of many heavy submatrices.

Our second lemma takes a heavy submatrix of a certain form with dimensions reasonably close to $n \times n$, and produces a slightly less heavy submatrix with dimensions much closer to $n \times n$, whose row and column sets satisfy the desired conditions in Lemma 3.1 (recall that we are assuming that $X = \{1, \dots, L\}$ and $Y = \{n - 3L + 1, \dots, n\}$). To be more precise, we actually start with a submatrix contained inside the $n' \times n'$ matrix $M_{n'}$ for some n' slightly smaller than n, and, conditioning on the outcome of $M_{n'}$, we only use the randomness from the additional rows and columns exposed when extending $M_{n'}$ to M_n

Lemma 6.2. *There is an absolute constant $c > 0$ such that the following holds for all sufficiently large integers $n \in \mathbb{N}$. Consider $\lambda > 0$ and integers L and R satisfying $(\log n)/20 < L < L^2 < R < (n - 5L^2 - 3L)/9$, and let $n' = n - 8R - 5L^2 - 3L$ and $\lambda' = \lambda/2^{R-L}$. Condition on an outcome of $M_{n'}$ for which there is a subset $B \subseteq \{1, \dots, n'\}$ of size $n' - R$ such that $M_{n'}[\{R+1, \dots, n'\}, B]$ is λ-heavy. Then with probability at least $1 - (1/8) \cdot n^{-c}$, there is a set B' of size $n - L$ with $\{1, \dots, n - 3L\} \subseteq B' \subseteq \{1, \dots, n\}$, such that $M_n[\{L + 1, \dots, n\}, B']$ is λ'-heavy.*

It is now easy to deduce Lemma 3.1 from the two lemmas above.

Proof of Lemma 3.1. Let $c > 0$ be the constant in Lemma 6.2. Recall that we are considering some $\varepsilon > 0$ and that we are assuming that n is sufficiently large with respect to ε. Then in particular $L = \lfloor (\log n)/10 \rfloor > (\log n)/20$. As mentioned at the beginning of this section, we may assume that $X = \{1, \dots, L\}$ and $Y = \{n - 3L + 1, \dots, n\}$.

Let $\delta = \varepsilon/32$, and $R = \lceil \delta n \rceil$, and note that by our assumption that n is large with respect to ε we have $L < L^2 < R < (n - 5L^2 - 3L)/9$. Now let $n' = n - 8R - 5L^2 - 3L \geq (1 - 9\delta)n$ and $\lambda = (n')^{(1/2 - 8\delta)n'} \geq n^{(1/2 - 15\delta)n}$ (again recalling that we assume n to be large with respect to ε). Note that then $\delta n' \leq R \leq 2\delta n'$.

By Lemma 6.1, with probability at least $1 - e^{-\delta^2 n'} \geq 1 - (1/8) \cdot n^{-c}$ (for n sufficiently large with respect to ε) there is a subset $B \subseteq \{1, \dots, n'\}$ of size $n' - R$ such that $M_{n'}[\{R + 1, \dots, n'\}, B]$ is λ-heavy. Then by Lemma 6.2 and our choice of c, with probability at least $1 - (1/8) \cdot n^{-c}$ there is a set B' of size $n - L$ with $\{1, \dots, n - 3L\} \subseteq B' \subseteq \{1, \dots, n\}$ such that $M_n[\{L + 1, \dots, n\}, B']$ is λ'-heavy, where $\lambda' = \lambda/2^{R-L} \geq n^{(1/2-15\delta)n}/2^{\delta n} \geq n^{(1/2-16\delta)n} = n^{(1-\varepsilon)n/2}$. Thus, the total probability that such a set B' exists is at least $1 - (1/4) \cdot n^{-c}$, as desired. □

Lemma 6.1 will be proved in Subsection 6.1 and Lemma 6.2 will be proved in Subsection 6.2.

6.1 Proof of Lemma 6.1: propagation of heavy submatrices

In this subsection we prove Lemma 6.1, adapting an argument from [33] to simultaneously track the propagation and growth of heavy submatrices as we expose more rows and columns of our random matrix. Roughly speaking, at each step we track submatrices of a certain form (with dimension growing by 1 at each step). At a given step, if we are guaranteed that many of our submatrices under consideration are heavy, then it is extremely likely that at the next step there will also be reasonably many heavy submatrices of the desired form. Moreover, depending on the structure of our random matrix at this step, one of the following is true, which will likely improve our situation in one of two ways. Either we have a good chance to dramatically increase the number of heavy submatrices in the next step, or we have a (very) good chance to have many submatrices in the next step which are much heavier than before. Lemma 6.3 below makes this precise.

After having proved Lemma 6.3, we will deduce Lemma 6.1 by iteratively applying Lemma 6.3, adding a new row and a new column to our random matrix at every step. Most likely, there will be many steps where our situation improves in one of the two ways described above. However, there is an upper bound for the number of heavy submatrices that we can have at the end of the process (simply by counting the total number of submatrices of the form that we consider). Hence the first type of improvement, which significantly increases the number of heavy submatrices, cannot occur too many times. So, among the two ways we can "improve the situation", the second type of improvement must happen most of the time. This means that during our process we get submatrices that are more and more heavy, and at the end we find a reasonably large number of very heavy submatrices in our final matrix M_n (in fact, we only need one such very heavy submatrix).

Lemma 6.3. *Fix $R, n \in \mathbb{N}$. For $k \in \mathbb{N}$ and real numbers $N > 0$ and $\lambda > 0$, let $E(k, N, \lambda)$ denote the event that there are at least N different subsets $B \subseteq \{1, \dots, k+R\}$ with $|B| = k$ such that the matrix $M_{k+R}[\{R+1, \dots, k+R\}, B]$ is λ-heavy.*

Then for any $k \in \mathbb{N}$ with $k + R \leq n$, and any real numbers $0 < \delta < 1/2$ as well as $K > 1$, $\lambda > 0$ and $N > 0$, there is a partition $E(k, N, \lambda) = E'(k, N, \lambda) \cup E''(k, N, \lambda)$ of the event $E(k, N, \lambda)$ such that the following holds. Let $N^+ = RN/(8K)$, $N^- = RN/(8n)$ and $\lambda^+ = K^{1/2-\delta}\lambda$, and let M, M', M'' be any possible outcomes of M_{k+R} satisfying $E(k, N, \lambda)$, $E'(k, N, \lambda)$ and $E''(k, N, \lambda)$ respectively. Then

$$\Pr\big(E(k+1, N^-, \lambda) \,\big|\, M_{k+R} = M\big) \geq 1 - 2e^{-R/8}. \tag{6.1}$$

$$\Pr\big(E(k+1, N^+, \lambda) \,\big|\, M_{k+R} = M'\big) \geq 1/3. \tag{6.2}$$

$$\Pr\big(E(k+1, N^-, \lambda^+) \,\big|\, M_{k+R} = M''\big) \geq 1 - 4K^{-\delta}. \tag{6.3}$$

Proof. We may assume without loss of generality that $N > 0$ is an integer (indeed, otherwise we can replace N by $\lceil N \rceil$, noting that the statement for $\lceil N \rceil$ implies the statement for N).

Let $x_1, \dots, x_{k+R+1}$ be the entries in the last row of M_{k+R+1}. For subsets $B \subseteq B' \subseteq \{1, \dots, k+R\}$ with sizes k and $k+1$ respectively, we say that B is a *parent* of B' and that B' is a *child* of B. For a subset $B' \subseteq \{1, \dots, k+R\}$ of size $k+1$, note that by Fact 5.1 we have

$$\operatorname{per} M_{k+R+1}[\{R+1, \dots, k+R+1\}, B'] = \sum_{B} \operatorname{per} M_{k+R}[\{R+1, \dots, k+R\}, B] \cdot x_{B' \setminus B}, \tag{6.4}$$

where the sum is over all parents B of B' (here, with slight abuse of notation we write $x_{\{i\}}$ instead of x_i for $i \in \{1, \dots, k+R\}$).

For each outcome of M_{k+R} such that $E(k, N, \lambda)$ holds, let us fix subsets $B_1, \dots, B_N$ as in the definition of $E(k, N, \lambda)$. Note that we always have $|\operatorname{per} M_{k+R}[\{R+1, \dots, k+R\}, B_i]| \geq \lambda$ for $i = 1, \dots, N$.

Furthermore, for each outcome of M_{k+R} satisfying $E(k, N, \lambda)$, let S_q denote the collection of subsets of $\{1, \dots, k+R\}$ of size $k+1$ which have exactly q parents among the sets $B_1, \dots, B_N$, and let $S = S_1 \cup \dots \cup S_n$ be the collection of all such subsets which have at least one parent among $B_1, \dots, B_N$. Furthermore, let $S_{\geq K} = S_{\lceil K \rceil} \cup \dots \cup S_n$ be the collection of all such subsets which have at least K parents among $B_1, \dots, B_N$. We say that $B' \in S$ is λ'-heavy for some $\lambda' > 0$ if $M_{k+R+1}[\{R+1, \dots, k+R+1\}, B']$ is λ'-heavy.

Since each of the sets $B_1, \dots, B_N$ is a parent of exactly R different sets $B' \in S$, a double-counting argument shows that we have

$$\sum_{q=1}^{n} q|S_q| = RN$$

for each outcome of M_{k+R} such that $E(k, N, \lambda)$ holds.

Now, let $E'(k, N, \lambda) \subseteq E(k, N, \lambda)$ be the event that $\sum_{q<K} q|S_q| \geq RN/2$, and condition on any outcome M' of M_{k+R} satisfying $E'(k, N, \lambda)$. Note that we then have $|S| \geq \sum_{q<K} |S_q| > RN/(2K)$. Furthermore note that for each $B' \in S$, at least one of the terms $\operatorname{per} M_{k+R}[\{R+1, \dots, k+R\}, B]$ on the left-hand side of (6.4) has absolute value at least λ (since B' has at least one parent among $B_1, \dots, B_N$). Hence, by Fact 4.2, each $B' \in S$ is λ-heavy with probability at least $1/2$, and (6.2) follows from Markov's inequality (to be precise, it follows from Lemma 4.8 applied with $p = 1/2$ and $q = 1/4$).

On the other hand, let $E''(k, N, \lambda) = E(k, N, \lambda) \setminus E'(k, N, \lambda)$ be the complementary event to $E'(k, N, \lambda)$ within $E(k, N, \lambda)$, i.e. the event that $\sum_{q \geq K} q|S_q| > RN/2$. Condition on any outcome M'' of M_{k+R} satisfying $E''(k, N, \lambda)$, and note that then $|S_{\geq K}| \geq RN/(2n)$. Also note that for each $B' \in S_{\geq K}$, at least K of the terms $\operatorname{per} M_{k+R}[\{R+1, \dots, k+R\}, B]$ on the left-hand side of (6.4) have absolute value at least λ. Hence, by the Erdős–Littlewood–Offord inequality (specifically, Theorem 4.1, applied with $m = K$, $r = \lambda$ and $t = K^{1/2-\delta}$), each $B' \in S_{\geq K}$ is λ^+-heavy with probability at least $1 - 3K^{-\delta}$. Then, (6.3) follows from Markov's inequality (specifically, we apply Lemma 4.8 with $p = 1 - 3K^{-\delta}$ and $q = 1/4$).

Finally, to prove (6.1), let us condition on any outcome M of M_{k+R} satisfying $E(k, N, \lambda)$. Say that for $i = 1, \dots, N$, the set B_i is *good* if at least $R/4$ of its R children $B' \in S$ are λ-heavy. We claim that each B_i is good with probability at least $1 - e^{-R/8}$. Indeed, consider some fixed $i \in \{1, \dots, N\}$, and condition on any outcome of the variables x_b for $b \in B_i$. Now for each child $B' \in S$ of B_i, the sum in

(6.4) depends only on the outcome of $x_{B'\setminus B_i}$ (since for all other elements of $b \in B'$ the corresponding variable x_b has already been fixed). Since $|\operatorname{per} M_{k+R}[\{R+1,\dots,k+R\},B_i]| \geq \lambda$, each child $B' \in S$ of B_i is λ-heavy with probability at least $1/2$, independently for all children $B' \in S$. So, by the Chernoff bound (Lemma 4.7), the set B_i is indeed good with probability at least $1 - e^{-2(R/4)^2/R} = 1 - e^{-R/8}$, as claimed.

Now, by Markov's inequality (specifically, Lemma 4.8, applied with $p = 1 - e^{-R/8}$ and $q = 1/2$), with probability at least $1 - 2e^{-R/8}$ at least $N/2$ of the sets $B_1, \dots, B_N$ are good. Whenever this is the case, there are at least $(N/2) \cdot (R/4)/n = RN/(8n)$ different λ-heavy sets $B' \in S$ (since each such set $B' \in S$ is a child of at most $k+1 \leq n$ different sets B_i). This proves (6.1). □

Now we deduce Lemma 6.1.

Proof of Lemma 6.1. As in the lemma statement, let $0 < \delta < 1/16$ and assume that $n \in \mathbb{N}$ is sufficiently large with respect to δ (sufficiently large to satisfy certain inequalities later in the proof). Let $R \in \mathbb{N}$ be an integer satisfying $\delta n \leq R \leq 2\delta n$, and let $K = n^{1-\delta}$. Furthermore, recall the notation from the statement of Lemma 6.3. We define random sequences $N_1, \dots, N_{n-R}$ and $\lambda_1, \dots, \lambda_{n-R}$ of positive real numbers by an iterative process. Let $N_1 = \lambda_1 = 1$ and for each $1 \leq k \leq n-R-1$ define N_{k+1} and λ_{k+1} as follows:

(i) if $E'(k, N_k, \lambda_k)$ and $E(k+1, N_k^+, \lambda_k)$ both hold, then let $N_{k+1} = N_k^+$ and $\lambda_{k+1} = \lambda_k$;

(ii) if $E''(k, N_k, \lambda_k)$ and $E(k+1, N_k^-, \lambda_k^+)$ both hold, then let $N_{k+1} = N_k^-$ and $\lambda_{k+1} = \lambda_k^+$;

(iii) if neither (i) nor (ii) holds, but $E(k, N_k, \lambda_k)$ and $E(k+1, N_k^-, \lambda_k)$ both hold, then let $N_{k+1} = N_k^-$ and $\lambda_{k+1} = \lambda_k$;

(iv) otherwise, abort (and then our sequences are not well-defined).

Note that the event $E(1, N_1, \lambda_1)$ always holds. If we do not abort at any point in the above process, then $E(k, N_k, \lambda_k)$ holds for each k, and in particular there is a subset $B \subseteq \{1, \dots, n\}$ of size $|B| = n - R$ such that $M_n[\{R+1, \dots, n\}, B]$ is λ_{n-R}-heavy. Thus, in order for the desired event in Lemma 6.1 to hold, it is sufficient that the process does not abort and that $\lambda_{n-R} \geq n^{(1/2-8\delta)n}$. We will show that this happens with probability at least $1 - e^{-\delta^2 n}$.

The main observation is that case (i) cannot occur too many times, simply because it is not possible for N_k to ever be larger than 2^n. Roughly speaking, it will follow from this observation and (6.2) that $E'(k, N_k, \lambda_k)$ is unlikely to occur too many times. This will then imply that case (ii) is likely to occur many times, meaning that λ_{n-R} is large.

Claim 6.4. *Case (i) in the above process occurs for fewer than δn different values of k.*

Proof. Note that whenever (i) holds, we have $N_{k+1}/N_k = N_k^+/N_k = R/(8K) \geq \delta n / 8n^{1-\delta} = n^\delta/8$. On the other hand, whenever (ii) or (iii) holds, we have $N_{k+1}/N_k = N_k^-/N_k = R/(8n) \geq \delta n/8n = \delta/8$. Now suppose for the purpose of contradiction that (i) holds for at least δn different k, and let us define $m = k+1$ for the last such value k. Note that then we have

$$N_m \geq (n^\delta/8)^{\delta n} \cdot (\delta/8)^{m-\delta n} \geq n^{\delta^2 n} \cdot (\delta/8)^m \geq n^{\delta^2 n} \cdot (\delta/8)^n > 2^n$$

for sufficiently large n. On the other hand, by our choice of m, case (i) holds for $k = m-1$, and so in particular the event $E(k+1, N_k^+, \lambda_k) = E(m, N_m, \lambda_m)$ holds. This means that there are at least $N_m > 2^n$ different subsets $B \subseteq \{1, \dots, m+R\} \subseteq \{1, \dots, n\}$ satisfying the conditions in the definition of the event $E(m, N_m, \lambda_m)$. But this is clearly a contradiction, since the total number of subsets of $\{1, \dots, n\}$ is only 2^n. □

The next observation is that if case (ii) occurs many times, then we are done.

Claim 6.5. *If we do not abort at any point, and case (ii) occurs for at least $n - 12\delta n$ different values of k, then $\lambda_{n-R} \geq n^{(1/2-8\delta)n}$.*

Proof. Whenever (ii) holds, we have $\lambda_{k+1}/\lambda_k = \lambda_k^+/\lambda_k = K^{1/2-\delta}$. On the other hand, whenever (i) or (iii) holds, we have $\lambda_{k+1} = \lambda_k$. So, if case (ii) occurs for at least $n - 12\delta n$ values of k, then

$$\lambda_{n-R} \geq (K^{1/2-\delta})^{n-12\delta n} \geq n^{(1-\delta)\cdot(1/2-\delta)\cdot(n-12\delta n)} \geq n^{(1/2-2\delta)\cdot(n-12\delta n)} \geq n^{(1/2-8\delta)n}. \qquad \square$$

It now suffices to show that with probability at least $1 - e^{-\delta^2 n}$, we do not abort and case (ii) occurs at least $n - 12\delta n$ times. To this end, we define an auxiliary random process $W_1, \dots, W_{n-R}$ that evolves in parallel with $N_1, \dots, N_{n-R}$ and $\lambda_1, \dots, \lambda_{n-R}$. Namely, let $W_1 = 0$, and for $1 \leq k \leq n - R - 1$ let

$$W_{k+1} = W_k + (1-\delta) - \begin{cases} 3 & \text{in case (i)}, \\ 1 & \text{in case (ii)}, \\ 0 & \text{in case (iii) or (iv)}. \end{cases}$$

Furthermore, if case (iv) occurs then let $W_{k+2} = W_{k+3} = \cdots = W_{n-R}$ all be equal to the value of W_{k+1} just defined (that is to say, we "freeze" the value of W_k after the process aborts).

Note that $W_1, \dots, W_k$ are fully determined by the random matrix M_{k+R} (which also determines its submatrices $M_{R+1}, \dots, M_{k+R}$). Moreover, this defines a supermartingale, in the sense that $\mathbb{E}[W_{k+1}|M_{k+R}] \leq W_k$ for each k (provided n is sufficiently large). To see this, consider any outcome of M_k for which we have not yet aborted (meaning in particular that the event $E(k, N_k, \lambda_k) = E'(k, N_k, \lambda_k) \cup E''(k, N_k, \lambda_k)$ holds). If $E'(k, N_k, \lambda_k)$ holds, then $\mathbb{E}[W_{k+1} - W_k|M_k] \leq (1-\delta) - (1/3) \cdot 3 \leq -\delta$ by (6.2). On the other hand, if $E''(k, N_k, \lambda_k)$ holds, then $\mathbb{E}[W_{k+1} - W_k|M_k] \leq (1-\delta) - (1 - 4K^{-\delta}) = -\delta + 4K^{-\delta} \leq 0$ for sufficiently large n, by (6.3). In addition, observe that $|W_i - W_{i-1}| \leq 3$ for each $1 < i \leq n - R$.

By Lemma 4.6 (with $Z_k = M_{k+R}$ for $k = 1, \dots, n-R$, and $c = 3$) we have $W_{n-R} \leq 5\delta n$ with probability at least $1 - e^{-(25/18)\delta^2 n} \geq 1 - (1/2)e^{-\delta^2 n}$. Also, by (6.1) and the union bound, the probability that we ever abort is bounded by $(n-R) \cdot 2e^{-R/8} \leq n \cdot 2e^{-\delta n/8} \leq (1/2)e^{-\delta^2 n}$. But note that if we never abort, then

$$W_{n-R} = (n - R - 1)(1-\delta) - 3X_{\text{(i)}} - X_{\text{(ii)}},$$

where $X_{\text{(i)}}$ is the number of times that case (i) occurs, and $X_{\text{(ii)}}$ is the number of times that case (ii) occurs. Recall that $X_{\text{(i)}} \leq \delta n$ by Claim 6.4. Hence, if $W_{n-R} \leq 5\delta n$ and the process does not abort, then case (ii) occurs $X_{\text{(ii)}} \geq (n-R-1)(1-\delta) - 3\delta n - 5\delta n \geq n - 12\delta n$ times, which by Claim 6.4 implies that $\lambda_{n-R} \geq n^{(1/2-8\delta)n}$. Thus, we have indeed shown that with probability at least $1 - e^{-\delta^2 n}$ the process does not abort and we have $\lambda_{n-R} \geq n^{(1/2-8\delta)n}$. □

6.2 Proof of Lemma 6.2: "filling out" a single heavy submatrix

In this subsection we prove Lemma 6.2. It will be a consequence of the following two lemmas, which (in two slightly different ways) "grow" a heavy submatrix by exposing a few additional rows and columns.

Lemma 6.6. *Let $1 \leq S < n$ and $\lambda > 0$, and condition on an outcome of M_n for which there is a subset $B \subseteq \{1, \dots, n\}$ of size $n - S$ such that $M_n[\{S+1, \dots, n\}, B]$ is λ-heavy. Then with probability at least $1 - 3S \cdot 2^{-S} - e^{-S/6}$, there is a set B' of size $n + 2S$ with $\{1, \dots, n\} \subseteq B' \subseteq \{1, \dots, n+3S\}$ such that the matrix $M_{n+3S}[\{S+1, \dots, n+3S\}, B']$ is λ-heavy.*

Lemma 6.7. *Let $2 \leq T < S < n$ and $\lambda > 0$, and condition on an outcome of M_n for which there is a set B of size $n - S$ with $\{1, \dots, S\} \subseteq B \subseteq \{1, \dots, n\}$ such that $M_n[\{S+1, \dots, n\}, B]$ is λ-heavy. Then with probability at least $1 - 5S \cdot 2^{-T} - e^{-S/40}$, there is a set B' of size $n + 5S - T$ with $\{1, \dots, T\} \subseteq B \subseteq \{1, \dots, n+5S\}$ such that $M_{n+5S}[\{T+1, \dots, n+5S\}, B']$ is $\lambda/2^{S-T}$-heavy.*

Before proving Lemmas 6.6 and 6.7, we deduce Lemma 6.2.

Proof of Lemma 6.2. Recall that $n' = n - 8R - 5L^2 - 3L$, and that we are conditioning on an outcome of $M_{n'}$ for which there is a subset $B \subseteq \{1, \dots, n'\}$ of size $n' - R$ such that $M_{n'}[\{R+1, \dots, n'\}, B]$ is λ-heavy.

First, by Lemma 6.6 (applied with $S = R < n'$), with probability at least $1 - 3R \cdot 2^{-R} - e^{-R/6}$, there is a set B_1 of size $n' + 2R$ with $\{1, \dots, R\} \subseteq \{1, \dots, n'\} \subseteq B' \subseteq \{1, \dots, n'+3R\}$ such that $M_{n'+3R}[\{R+1, \dots, n'+3R\}, B_1]$ is λ-heavy. Let us now condition on such an outcome for $M_{n'+3R}$.

Then, by Lemma 6.7 (applied with $T = L^2$ and $S = R < n' + 3R$), we obtain that with probability at least $1 - 5R\cdot 2^{-L^2} - e^{-R/40}$ there is a set B_2 of size $n' + 8R - L^2$ with $\{1, \dots, L^2\} \subseteq B_2 \subseteq \{1, \dots, n' + 8R\}$ such that that the matrix $M_{n'+8R}[\{L^2 + 1, \dots, n' + 8R\}, B_2]$ is $(\lambda/2^{R-L^2})$-heavy. Let us condition on such an outcome for $M_{n'+8R}$.

Applying Lemma 6.7 again (this time with $T = L$ and $S = L^2 < n' + 8R$), we now get that with probability at least $1 - 5L^2 \cdot 2^{-L} - e^{-L^2/40}$, there is a set B_3 of size $n' + 8R + 5L^2 - L$ with $\{1, \dots, L\} \subseteq B_3 \subseteq \{1, \dots, n' + 8R + 5L^2\}$ such that that $M_{n'+8R+5L^2}[\{L+1, \dots, n'+8R+5L^2\}, B_3]$ is λ'-heavy, where $\lambda' = (\lambda/2^{R-L^2})/2^{L^2-L} = \lambda/2^{R-L}$. Let us condition on such an outcome for $M_{n'+8R+5L^2}$

Finally, by Lemma 6.6 (applied with $S = L < n'+8R+5L^2$), with probability at least $1 - 3L\cdot 2^{-L} - e^{-L/6}$, there is a set B' of size $n' + 8R + 5L^2 + 2L = n - L$ with

$$\{1, \dots, n - 3L\} = \{1, \dots, n' + 8R + 5L^2\} \subseteq B' \subseteq \{1, \dots, n' + 8R + 5L^2 + 3L\} = \{1, \dots, n\}$$

such that $M_n[\{L + 1, \dots, n\}, B']$ is λ'-heavy.

The probability that all four steps succeed is at least

$$1-(3R\cdot 2^{-R}+e^{-R/6})-(5R\cdot 2^{-L^2}+e^{-R/40})-(5L^2\cdot 2^{-L}+e^{-L^2/40})-(3L\cdot 2^{-L}+e^{-L/6}) \geq 1-(1/8)\cdot n^{-c}$$

for some small constant $c > 0$ (recall that $(\log n)/20 < L < L^2 < R < n$ and that n is sufficiently large). □

We now prove Lemma 6.6.

Proof of Lemma 6.6. For any $m \in \mathbb{N}$ with $n \leq m \leq n + 3S$, define the random variable Q_m to be the minimum value of $|\{1, \dots, n\} \setminus B'|$ among all subsets $B' \subseteq \{1, \dots, m\}$ of size $m - S$ such that $M_m[\{S + 1, \dots, m\}, B']$ is λ-heavy. If no such subset B' exists, let $Q_m = \infty$.

Recall that in Lemma 6.6 we are conditioning on an outcome of M_n for which there is a subset $B \subseteq \{1, \dots, n\}$ of size $n-S$ such that $M_n[\{S+1, \dots, n\}, B]$ is λ-heavy. This means that $Q_n \leq |\{1, \dots, n\} \setminus B| = S$.

For $n < m \leq n + 3S$, we say that step m is a *failure* if $Q_m > Q_{m-1}$. We say that step m is *progress* if $Q_m < Q_{m-1}$ or if $Q_m = Q_{m-1} = 0$ or if $Q_{m-1} = \infty$.

For any $n < m \leq n + 3S$, when conditioning on any outcome of M_{m-1}, we claim that step m is a failure with probability at most 2^{-S}. Indeed, if $Q_{m-1} < \infty$, let $B' \subseteq \{1, \dots, m - 1\}$ be a subset of size $m - 1 - S$ with $|\{1, \dots, n\} \setminus B'| = Q_{m-1}$ such that $M_{m-1}[\{S + 1, \dots, m - 1\}, B']$ is λ-heavy. By applying Lemma 5.4 with $I = \{1, \dots, m - 1\} \setminus B'$, we see that with probability at least $1 - 2^{-S}$ there exists some $i \in \{1, \dots, m - 1\} \setminus B'$ such that $M_m[\{S + 1, \dots, m\}, B' \cup \{i\}]$ is λ-heavy (which in particular implies $Q_m \leq |\{1, \dots, n\} \setminus (B' \cup \{i\})| \leq Q_{m-1}$). On the other hand, if $Q_{m-1} = \infty$, then step m cannot be a failure. So this indeed shows that in any case (when conditioning on any outcome of M_{m-1}), step m is a failure with probability at most 2^{-S}.

Furthermore, for any $n < m \leq n + 3S$, when conditioning on any outcome of M_{m-1}, we claim that step m is progress with probability at least $1/2$. Indeed, if $Q_{m-1} \notin \{0, \infty\}$, let $B' \subseteq \{1, \dots, m - 1\}$ be a subset of size $m - 1 - S$ with $|\{1, \dots, n\} \setminus B'| = Q_{m-1}$ such that $M_{m-1}[\{S + 1, \dots, m - 1\}, B']$ is λ-heavy. We can then apply Lemma 5.4 with $I = \{1, \dots, n\} \setminus B'$, to see that with probability at least $1/2$ there exists some $i \in \{1, \dots, n\} \setminus B'$ such that $M_m[\{S + 1, \dots, m\}, B' \cup \{i\}]$ is λ-heavy (which in particular implies $Q_m \leq |\{1, \dots, n\} \setminus (B' \cup \{i\})| < Q_{m-1}$). If $Q_{m-1} = \infty$, by definition step m is always progress. If $Q_{m-1} = 0$, then step m is progress if and only if it is not failure, and we already showed that it is failure with probability at most $2^{-S} \leq 1/2$. This shows that in any case (when conditioning on any outcome of M_{m-1}), step m is progress with probability at least $1/2$.

Hence the number of progress steps among the $3S$ steps $m \in \{n + 1, \dots, n + 3S\}$ stochastically dominates a sum of $3S$ independent $\mathrm{Ber}(1/2)$ random variables. By the Chernoff bound (Lemma 4.7) such a sum is at least S with probability at least $1 - e^{-2(S/2)^2/(3S)} = 1 - e^{-S/6}$. Thus, the number of progress steps is also at least S with probability at least $1 - e^{-S/6}$. Furthermore, note that by the union bound, with probability at least $1 - 3S \cdot 2^{-S}$ none of the $3S$ steps $m \in \{n + 1, \dots, n + 3S\}$ is a failure.

If there are no failures and at least S progress steps, then we must have $Q_{n+3S} = 0$. Hence, with probability at least $1 - 3S\cdot 2^{-S} - e^{-S/6}$ there exists a subset $B' \subseteq \{1, \dots, n + 3S\}$ of size $n + 2S$ such that $M_{n+3S}[\{S + 1, \dots, n + 3S\}, B']$ is λ-heavy and $|\{1, \dots, n\} \setminus B'| = 0$ (meaning that $\{1, \dots, n\} \subseteq B'$). □

The proof of Lemma 6.7 follows a similar strategy as the proof of Lemma 6.6 above.

Proof of Lemma 6.7. For any $m \in \mathbb{N}$ with $n \leq m \leq n+5S$, define the random variable Q_m to be the minimal number $Q \in \{T, T+1, T+2, \dots\}$ such that there is a set B' of size $m-Q$ with $\{1,\dots,Q\} \subseteq B' \subseteq \{1,\dots,m\}$ such that the matrix $M_m[\{Q+1,\dots,m\},B']$ is $(\lambda/2^{S-Q})$-heavy. If there is no $Q \in \{T, T+1, T+2, \dots\}$ for which such a set B' exists, we define $Q_m = \infty$.

Recall that in Lemma 6.7, we are conditioning on an outcome of M_n for which there is a set B of size $n-S$ with $\{1,\dots,S\} \subseteq B \subseteq \{1,\dots,n\}$ such that that $M_n[\{S+1,\dots,n\},B]$ is λ-heavy. This means that $Q_n \leq S$ (recall that $S > T$).

For $n < m \leq n+5S$, we say that step m is a *failure* if $Q_m > Q_{m-1}$. We say that step m is *progress* if $Q_m < Q_{m-1}$ or if $Q_m = Q_{m-1} = T$ or if $Q_{m-1} = \infty$.

For any $n < m \leq n+5S$, when conditioning on any outcome of M_{m-1}, we claim that step m is a failure with probability at most 2^{-T}. Indeed, if $Q_{m-1} < \infty$, let B' be a set of size $m-1-Q_{m-1}$ with $\{1,\dots,Q_{m-1}\} \subseteq B' \subseteq \{1,\dots,m-1\}$ such that the matrix $M_{m-1}[\{Q_{m-1}+1,\dots,m-1\},B']$ is $(\lambda/2^{S-Q_{m-1}})$-heavy. By applying Lemma 5.4 with $I = \{1,\dots,m-1\} \setminus B'$, we obtain that with probability at least $1-2^{-Q_{m-1}} \geq 1-2^{-T}$ there exists some $i \in \{1,\dots,m-1\} \setminus B'$ such that the matrix $M_m[\{Q_{m-1}+1,\dots,m\},B' \cup \{i\}]$ is $(\lambda/2^{S-Q_{m-1}})$-heavy (which in particular implies that $Q_m \leq Q_{m-1}$). On the other hand, if $Q_{m-1} = \infty$, then step m cannot be a failure.

We furthermore claim that for any $n < m \leq n+5S$, when conditioning on any outcome of M_{m-1}, step m is progress with probability at least 1/4. First assume that $Q_{m-1} \notin \{T,\infty\}$, and let B' be a set of size $m-1-Q_{m-1}$ with $\{1,\dots,Q_{m-1}\} \subseteq B' \subseteq \{1,\dots,m-1\}$ such that $M_{m-1}[\{Q_{m-1}+1,\dots,m-1\},B']$ is $(\lambda/2^{S-Q_{m-1}})$-heavy. We can then apply Lemma 5.5 to the sets $\{Q_{m-1}+1,\dots,m-1\}$ and B'. Since $\{1,\dots,Q_{m-1}\} \subseteq B'$ and $|B'| = m-1-Q_{m-1}$, we have $|\{Q_{m-1}+1,\dots,m-1\} \setminus B'| = Q_{m-1} \geq T \geq 2$. So we can find two distinct elements $b_1, b_2 \in \{Q_{m-1}+1,\dots,m-1\} \setminus B'$. Let us furthermore take $a = Q_{m-1} \in B' \setminus \{Q_{m-1}+1,\dots,m-1\}$. By Lemma 5.5, there exist distinct elements $i,j \in \{a,b_1,b_2\}$ such that the set $B^* = (B' \setminus \{a\}) \cup \{i,j,m\}$ has the property that the matrix $M_m[\{Q_{m-1},\dots,m\},B^*]$ is $(\lambda/2^{S-Q_{m-1}+1})$-heavy with probability at least 1/4. Also note that $|B^*| = m-(Q_m-1)$ and $\{1,\dots,Q_{m-1}-1\} \subseteq B^* \subseteq \{1,\dots,m\}$. So we can conclude that with probability at least 1/4 we have $Q_m \leq Q_{m-1}-1$, meaning that step m is progress.

In order to finish proving the claim that for any $n < m \leq n+5S$ (when conditioning on any outcome of M_{m-1}), step m is progress with probability at least 1/4, it only remains to consider the cases $Q_{m-1} = \infty$ and $Q_{m-1} = T$. If $Q_{m-1} = \infty$, then step m is always progress. If $Q_{m-1} = T$, then step m is progress if and only if it is not failure, and we already proved that step m is failure with probability at most $2^{-T} \leq 1/4 \leq 3/4$.

Having proved this claim, we can now conclude that the number of progress steps among the $5S$ steps $m \in \{n+1,\dots,n+5S\}$ stochastically dominates a sum of $5S$ independent Ber(1/4) random variables. By the Chernoff bound (Lemma 4.7) such a sum is at least S with probability at least $1-e^{-2(S/4)^2/(5S)} = 1-e^{-S/40}$. Thus, the number of progress steps is also at least S with probability at least $1-e^{-S/40}$. Furthermore, note that by the union bound, with probability at least $1-5S\cdot 2^{-T}$ none of the $5S$ steps $m \in \{n+1,\dots,n+5S\}$ is a failure.

If there are no failures and at least $S \geq S-T$ progress steps, then we must have $Q_{n+5S} = T$. Hence, with probability at least $1-5S\cdot 2^{-T}-e^{-S/40}$ there is a set B' of size $n+5S-T$ with $\{1,\dots,T\} \subseteq B' \subseteq \{1,\dots,n+5S\}$ such that the matrix $M_{n+5S}[\{T+1,\dots,n+5S\},B']$ is $(\lambda/2^{S-T})$-heavy. □

7 Proof of Lemma 3.2: survival of heavy submatrices

In this section we prove Lemma 3.2. Recall that we call subsets $S_1,\dots,S_m$ of some ground set S complement-disjoint, if their complements $S \setminus S_1,\dots,S \setminus S_m$ are disjoint (and note that this condition is in particular satisfied if $S_1 = \dots = S_m = S$).

As in the lemma statement, let $\lambda > 0$ and let $A_1,\dots,A_m,B_1,\dots,B_m$ be complement-disjoint subsets of $\{1,\dots,n\}$ of size $n-L$. Recall that we are conditioning on an outcome of the matrix M_n such that we have $|\operatorname{per} M_n[A_\ell,B_\ell]| \geq \lambda$ for $\ell = 1,\dots,m$. Also recall that we are assuming that m is large.

Let $x_1, \dots, x_n, z$ be the entries of the new row and column in M_{n+1}, and let us condition on any fixed outcome of z (which we no longer view as being random).

First, starting from the complement-disjoint subsets $A_1, \dots, A_m, B_1, \dots, B_m \subseteq \{1, \dots, n\}$ of size $n-L$, we will construct certain complement-disjoint subsets $A_1^*, \dots, A_m^*, B_1^*, \dots, B_m^* \subseteq \{1, \dots, n\}$ of size $n-L+1$. The plan is then to choose the desired subsets $A_1', \dots, A_{m'}', B_1', \dots, B_{m'}'$ in Lemma 3.2 to each be of the form $A_i^* \cup \{n+1\}$ or $B_i^* \cup \{n+1\}$, for suitably chosen $i \in \{1, \dots, m\}$.

Claim 7.1. *We can find quadruples $(A_\ell^*, B_\ell^*, i_\ell, j_\ell)$ for $\ell \in \{1, \dots, m\}$, satisfying the following conditions.*

- *For each $\ell \in \{1, \dots, m\}$, we have $A_\ell^*, B_\ell^* \subseteq \{1, \dots, n\}$ and $|A_\ell^*| = |B_\ell^*| = n-L+1$, and furthermore $i_\ell, j_\ell \in A_\ell^* \cap B_\ell^*$.*
- *The elements $i_1, j_1, \dots, i_m, j_m \in \{1, \dots, n\}$ are distinct.*
- *The sets $A_1^*, B_1^*, \dots, A_m^*, B_m^*$ are complement-disjoint (over the ground set $\{1, \dots, n\}$).*
- *For each $\ell \in \{1, \dots, m\}$, if we view $\operatorname{per} M_{n+1}[A_\ell^* \cup \{n+1\}, B_\ell^* \cup \{n+1\}]$ as a polynomial in $x_1, \dots, x_n$, then the coefficient of $x_{i_\ell} x_{j_\ell}$ has absolute value at least $\lambda/2$.*

Proof. First consider the case $L = 1$. For every $\ell \in \{1, \dots, m\}$, let us take $A_\ell^* = B_\ell^* = \{1, \dots, n\}$, let i_ℓ be the single element of $\{1, \dots, n\} \setminus A_\ell$, and let j_ℓ be the single element of $\{1, \dots, n\} \setminus B_\ell$. Note that the second condition is satisfied by the complement-disjointness of the sets $A_1, B_1, \dots, A_m, B_m$. For the last condition, observe that by Fact 5.2 and the symmetry of our matrices, the coefficient of $x_{i_\ell} x_{j_\ell}$ in $\operatorname{per} M_{n+1}$ equals $\operatorname{per} M_n[A_\ell, B_\ell] + \operatorname{per} M_n[B_\ell, A_\ell] = 2 \operatorname{per} M_n[A_\ell, B_\ell]$, which has absolute value at least $2\lambda \geq \lambda/2$.

Now, we consider the case $L \geq 2$. For each $\ell \in \{1, \dots, m\}$, choose $a_\ell \in B_\ell \setminus A_\ell$ and distinct $b_\ell, b_\ell' \in A_\ell \setminus B_\ell$ (this is possible since A_ℓ and B_ℓ are complement-disjoint and have size at most $n-2$). Now let $i_\ell, j_\ell \in \{a_\ell, b_\ell, b_\ell'\}$ be as in Lemma 5.3, and let $A_\ell^* = A_\ell \cup \{a_\ell\}$ and $B_\ell^* = (B_\ell \setminus \{a_\ell\}) \cup \{i_\ell, j_\ell\}$. For the last condition, note that by Fact 5.2 the coefficient of $x_{i_\ell} x_{j_\ell}$ in $\operatorname{per} M_{n+1}[A_\ell^* \cup \{n+1\}, B_\ell^* \cup \{n+1\}]$ equals $\operatorname{per} M_n[A_\ell^*, B_\ell^*]^{(i_\ell, j_\ell)} + \operatorname{per} M_n[B_\ell^*, A_\ell^*]^{(j_\ell, i_\ell)}$, which has absolute value at least $\lambda/2$ by the conclusion of Lemma 5.3. □

Fix quadruples $(A_\ell^*, B_\ell^*, i_\ell, j_\ell)$ for $\ell \in \{1, \dots, m\}$ as in Claim 7.1. Let $I = \{i_1, j_1, \dots, i_m, j_m\} \subseteq \{1, \dots, n\}$, and let us condition on any outcome for all the variables x_i with $i \notin I$ (which we will no longer view as being random). For every $\ell = 1, \dots, m$, define $P_\ell = \operatorname{per} M_{n+1}[A_\ell^* \cup \{n+1\}, B_\ell^* \cup \{n+1\}]$, viewed as a polynomial in the variables x_i for $i \in I$, but with all quadratic terms x_i^2 replaced by 1 (recall that our variables x_i are chosen in $\{-1, 1\}$). Then P_ℓ is a multilinear quadratic polynomial and the coefficient of $x_{i_\ell} x_{j_\ell}$ has absolute value at least $\lambda/2$.

Now, after our conditioning, the only remaining randomness comes from the $2m$ variables x_i for $i \in I$. It suffices to show that for sufficiently large m we have

$$\Pr\big(|P_\ell| \geq \lambda/(4n^4) \text{ for at least } m/36 \text{ indices } \ell \in \{1, \dots, m\}\big) \geq 1 - m^{-1/24}. \tag{7.1}$$

Indeed, if the event in (7.1) holds, then we can take the sets $A_1', \dots, A_{m'}, B_1', \dots, B_{m'}$ in Lemma 3.2 to be the sets $A_\ell^* \cup \{n+1\}$ and $B_\ell^* \cup \{n+1\}$ for $m' = \lceil m/36 \rceil$ different indices $\ell \in \{1, \dots, n\}$ for which we have $|P_\ell| = |\operatorname{per} M_{n+1}[A_\ell^* \cup \{n+1\}, B_\ell^* \cup \{n+1\}]| \geq \lambda/(4n^4)$.

Let $\sigma = \lambda/(4n^2)$ and $\tau = \lambda/(4n^4)$. Using the notation introduced above Theorem 4.4, for each $\ell \in \{1, \dots, m\}$ we consider the graph $G_\ell = G^{(\tau)}(P_\ell)$ on the vertex set I whose edges correspond to the coefficients of the polynomial P_ℓ of absolute value at least τ. We say that the index $\ell \in \{1, \dots, m\}$ is *easy* if the graph G_ℓ has matching number $\nu(G_\ell) \geq m^{1/6}$. If there are many easy indices, then we can prove (7.1) using the Meka–Nguyen–Vu polynomial anti-concentration inequality (Theorem 4.4), as follows.

Claim 7.2. *If there are at least $m/3$ easy indices $\ell \in \{1, \dots, m\}$, then (7.1) holds.*

Proof. Recall that for each easy index ℓ, we have $\nu(G^{(\tau)}(P_\ell)) = \nu(G_\ell) \geq m^{1/6}$. Hence by Theorem 4.4 we have that

$$\Pr(|P_\ell| \geq \tau) \geq 1 - \frac{(\log \nu(G_\ell))^C}{\nu(G_\ell)^{1/2}} \geq 1 - \nu(G_\ell)^{-1/3} \geq 1 - m^{-1/18}$$

for sufficiently large m (where C is the absolute constant appearing in the statement of Theorem 4.4). Hence by Markov's inequality (specifically Lemma 4.8 applied with $p = 1 - m^{-1/18}$ and $q = 1/2$), with probability at least $1 - 2m^{-1/18} \geq 1 - m^{-1/24}$ we have $|P_\ell| \geq \tau = \lambda/(4n^4)$ for at least $(1/2) \cdot (m/3) = m/6$ easy indices $\ell \in \{1, \dots, m\}$ (again assuming that m is sufficiently large). This in particular proves (7.1). □

Let us from now on assume that there are at least $2m/3$ indices $\ell \in \{1, \dots, m\}$ which are not easy. For each of these non-easy ℓ, since G_ℓ has no large matching it must have a large vertex cover, as follows.

Claim 7.3. *For every non-easy index $\ell \in \{1, \dots, m\}$, there is a subset $S_\ell \subseteq I$ of size $|S_\ell| \leq 2m^{1/6}$ such that each edge of the graph G_ℓ contains at least one vertex in S_ℓ (in other words, S_ℓ is a vertex cover of the graph G_ℓ).*

Proof. Let us take a maximal collection of disjoint edges in G_ℓ (this collection consists of at most $\nu(G_\ell) < m^{1/6}$ edges), and let S_ℓ consist of all the vertices contained in one of these edges. Then by the maximality of the chosen edge collection, each edge of G_ℓ must contain at least one vertex in S_ℓ. □

For each non-easy ℓ, fix a subset $S_\ell \subseteq I$ as in Claim 7.3. We now briefly describe the strategy of the remainder of the proof. The idea is that all the degree-2 terms of P_ℓ whose coefficient has large absolute value must contain a variable whose index is in S_ℓ. So if we condition on outcomes of x_i for $i \in S_\ell$, then P_ℓ "essentially" becomes a linear polynomial (apart from some small terms that we can ignore). If this linear polynomial has many coefficients with large absolute value, then we can apply the Erdős–Littlewood–Offord inequality to show that $|P_\ell|$ is typically quite large. However, it is possible that for most of the non-easy ℓ, we end up with linear polynomials which have few coefficients of large absolute value (in which case we will not be able to use such an argument). It turns out that this is unlikely to happen unless for many ℓ the polynomial P_ℓ each essentially depend on only a few of the variables x_i (we will call such indices ℓ *short*). We will be able to handle the case that there are many such indices ℓ using the Azuma–Hoeffding inequality.

Let us say that a variable x_i with $i \in I$ is *bad* if we have $i \in S_\ell$ for at least $m^{1/3}$ non-easy indices $\ell \in \{1, \dots, m\}$. Note that by a simple counting argument, there are at most $m \cdot 2m^{1/6}/m^{1/3} = 2m^{5/6}$ bad variables. We say that a variable x_i with $i \in I$ is *good* if it is not bad. Let $I_{\text{good}} \subseteq I$ be the set of all $i \in I$ such that x_i is good (and note that $|I_{\text{good}}| \leq |I| = 2m$).

In addition to all of our previous conditioning, let us now also condition on any fixed outcome of all bad variables x_i. This means that at this point the only remaining randomness comes from the variables x_i with $i \in I_{\text{good}}$. After fixing the outcomes for the bad variables, we can interpret each polynomial P_ℓ (for $\ell \in \{1, \dots, m\}$) as a polynomial in the variables x_i with $i \in I_{\text{good}}$. It suffices to prove (7.1) with this additional conditioning on the outcomes of the bad variables.

Note that for each $\ell \in \{1, \dots, m\}$ which is not easy, for any distinct $i, j \in I_{\text{good}} \setminus S_\ell$ the coefficient of $x_i x_j$ in P_ℓ has absolute value less than τ. Indeed, this follows from the definition of the graph $G_\ell = G^{(\tau)}(P_\ell)$ and the fact that every edge of G_ℓ contains at least one vertex in S_ℓ.

Recall that $\sigma = \lambda/(4n^2)$. We say that an index $\ell \in \{1, \dots, m\}$ is *short* if there are at most $6m^{1/6}$ good variables x_i such that x_i appears in a term of P_ℓ whose coefficient has absolute value at least σ.

Claim 7.4. *If there are at least $m/3$ short indices $\ell \in \{1, \dots, m\}$, then (7.1) holds.*

Proof. Recall that we are viewing each P_ℓ as a polynomial in the variables x_i with $i \in I_{\text{good}}$. Now, let P'_ℓ be the polynomial obtained from P_ℓ by deleting all terms whose coefficients have absolute value less than σ. Note that for each short index $\ell \in \{1, \dots, m\}$, the polynomial P'_ℓ contains at most $6m^{1/6}$ different variables x_i. Also note that we always have $|P_\ell - P'_\ell| < \sigma n^2 = \lambda/4$ for any outcomes of the good variables $x_i \in \{-1, 1\}$ (this is because at most n^2 terms get deleted in P'_ℓ, each with absolute value less than σ).

We say that a good variable is *short-popular* if it appears in the polynomial P'_ℓ for at least $m^{1/3}$ different short indices $\ell \in \{1, \dots, m\}$. Note that there are at most $m \cdot 6m^{1/6}/m^{1/3} = 6m^{5/6}$ short-popular variables. For the remainder of the proof of this claim, let us now also condition on any fixed outcomes for all short-popular variables x_i, which we no longer view as being random.

Since there are at most $2m^{5/6}$ bad variables, and at most $6m^{5/6}$ short-popular variables, there are at least $m/3 - 8m^{5/6} \geq m/4$ short indices $\ell \in \{1, \dots, m\}$ for which both of the variables x_{i_ℓ} and x_{j_ℓ} are good and not short-popular (for the inequality here, we are assuming that m is sufficiently large). For any such index ℓ, the coefficient of $x_{i_\ell}x_{j_\ell}$ in the polynomial P'_ℓ has absolute value at least $\lambda/2$ (since this is also the case for P_ℓ, and $\lambda/2 > \sigma$). Hence we can apply Fact 4.3 to find that $\Pr(|P'_\ell| \geq \lambda/2) \geq 1/4$.

So, if Y is the number of short indices ℓ with $|P'_\ell| \geq \lambda/2$, then $\mathbb{E}Y \geq (m/4)\cdot(1/4) = m/16$. Recall that we already conditioned on outcomes of all the short-popular variables. So each of the remaining random variables occur in P'_ℓ for at most $m^{1/3}$ short indices ℓ, and hence varying each individual variable affects the value of Y by at most $m^{1/3}$. Therefore, by Lemma 4.5 we have $Y \geq m/32$ with probability at least

$$1 - 2\exp\left(-\frac{(m/32)^2}{2\cdot 2m\cdot m^{2/3}}\right) = 1 - 2\exp\left(-\frac{m^{1/3}}{2^{12}}\right) \geq 1 - m^{-1/24},$$

assuming that m is sufficiently large. Hence with probability at least $1 - m^{-1/24}$ there are at least $m/32$ short indices ℓ with $|P'_\ell| \geq \lambda/2$, which implies that $|P_\ell| \geq \lambda/2 - \lambda/4 \geq \lambda/(4n^4)$. This in particular proves (7.1). □

We may from now on assume that there are at least $m/3$ indices $\ell \in \{1,\dots,m\}$ which are not easy and not short. Let us call such indices *interesting*.

For every interesting index $\ell \in \{1,\dots,m\}$, recall that P_ℓ is a multilinear polynomial in the variables x_i with $i \in I_{\text{good}}$. Furthermore, recall that for any distinct $i, j \in I_{\text{good}} \setminus S_\ell$ the coefficient of x_ix_j in P_ℓ has absolute value less than τ. Let P^*_ℓ be the polynomial obtained from P_ℓ by deleting all terms of the form x_ix_j for $i, j \in I_{\text{good}} \setminus S_\ell$. Note that we always have $|P_\ell - P^*_\ell| \leq \tau n(n-1)/2 \leq \tau(n^2-1) = \sigma - \tau$ (for all outcomes of the x_i).

For every interesting $\ell \in \{1,\dots,m\}$, there are at least $6m^{1/6}$ good variables which appear in a term of P_ℓ whose coefficient has absolute value at least σ. Since only terms with coefficient less than $\tau < \sigma$ get deleted in P^*_ℓ, this means that there are also at least $6m^{1/6}$ good variables which appear in a term of P^*_ℓ whose coefficient has absolute value at least σ.

Now, for every interesting $\ell \in \{1,\dots,m\}$, let us interpret $P^*_\ell \in \mathbb{R}[x_i, i \in I_{\text{good}}]$ as a polynomial $Q_\ell \in \mathbb{R}[x_i, i \in S_\ell][x_i, i \in I_{\text{good}} \setminus S_\ell]$, i.e. as a polynomial in the variables x_i for $i \in I_{\text{good}} \setminus S_\ell$ whose coefficients are polynomials in the variables x_i for $i \in S_\ell$. Then Q_ℓ is a linear polynomial (in the variables x_i for $i \in I_{\text{good}} \setminus S_\ell$). Its constant coefficient is a quadratic polynomial, and its other coefficients are linear polynomials (in the variables x_i for $i \in S_\ell$).

Let T_ℓ be the number of degree-1 coefficients of the linear polynomial Q_ℓ (in the variables x_i for $i \in I_{\text{good}} \setminus S_\ell$) which have absolute value at least σ. This is a random variable depending on the outcomes of the x_i with $i \in S_\ell$.

Claim 7.5. *For each interesting index $\ell \in \{1,\dots,m\}$, we have $\Pr\big(T_\ell \geq m^{1/6}\big) \geq 1/3$.*

Proof. Fix an interesting $\ell \in \{1,\dots,m\}$. Recall that there are at least $6m^{1/6}$ good variables x_i which appear in a term of P^*_ℓ whose coefficient has absolute value at least σ. Since $|S_\ell| \leq 2m^{1/6}$, at least $4m^{1/6}$ of these variables satisfy $i \in I_{\text{good}} \setminus S_\ell$. For each such i, the coefficient of x_i in Q_ℓ is a linear polynomial in the variables x_j for $j \in S_\ell$, with at least one coefficient of absolute value at least σ. Thus, by Fact 4.2, with probability at least $1/2$ the coefficient of x_i in Q_ℓ has absolute value at least σ. By Markov's inequality (specifically Lemma 4.8 applied with $p = 1/2$ and $q = 1/4$), with probability at least $1/3$ there are at least $(1/4)\cdot 4m^{1/6} = m^{1/6}$ different $i \in I_{\text{good}} \setminus S_\ell$ such that the coefficient of x_i in Q_ℓ has absolute value at least σ. In other words, with probability at least $1/3$, we have $T_\ell \geq m^{1/6}$. □

Claim 7.6. *For each interesting index $\ell \in \{1,\dots,m\}$, we have $\Pr\big(T_\ell \geq m^{1/6}$ and $|P^*_\ell| < \sigma\big) \leq 3m^{-1/12}$.*

Proof. Fix an interesting $\ell \in \{1,\dots,m\}$. Recall that T_ℓ depends only on the variables x_i with $i \in S_\ell$. So, for the proof of this claim, let us condition on some outcome for the variables x_i with $i \in S_\ell$ such that we have $T_\ell \geq m^{1/6}$. Under this conditioning the polynomial P^*_ℓ becomes precisely the polynomial Q_ℓ, which is a linear polynomial in the remaining variables x_i for $i \in I_{\text{good}} \setminus S_\ell$, having $T_\ell \geq m^{1/6}$ degree-1 coefficients with absolute value at least σ. Now, by the Erdős–Littlewood–Offord inequality (Theorem 4.1, applied with $t = 1$) we indeed have $\Pr(|P^*_\ell| < \sigma) \leq 3/T_\ell^{1/2} \leq 3m^{-1/12}$. □

Next, define the random variable X as the number of interesting $\ell \in \{1, \ldots, m\}$ such that $T_\ell \geq m^{1/6}$. Furthermore, define the random variable Y as the number of interesting $\ell \in \{1, \ldots, m\}$ such that $T_\ell \geq m^{1/6}$ and $|P^*_\ell| < \sigma$.

Claim 7.7. *With probability at least* $1 - m^{-1/12}$ *we have* $X \geq m/18$.

Proof. First, note that by Claim 7.5 we have $\mathbb{E}X \geq (m/3) \cdot (1/3) = m/9$. Also recall that for each interesting ℓ, the event $T_\ell \geq m^{1/6}$ only depends on the outcomes of x_i for $i \in S_\ell$. This means that changing the outcome of any of the good random variables x_i can affect X by at most $m^{1/3}$ (recall that for each good x_i, we have $i \in S_\ell$ for at most $m^{1/3}$ different ℓ). Hence by Lemma 4.5, we have $X \geq m/18$ with probability at least

$$1 - 2\exp\left(-\frac{(m/18)^2}{2 \cdot 2m \cdot m^{2/3}}\right) = 1 - 2\exp\left(-\frac{m^{1/3}}{36^2}\right) \geq 1 - m^{-1/12}$$

(for sufficiently large m), as desired. □

Claim 7.8. *With probability at least* $1 - 108m^{-1/12}$ *we have* $Y \leq m/36$.

Proof. By Claim 7.6 we have $\mathbb{E}Y \leq m \cdot 3m^{-1/12} = 3m^{11/12}$. Hence, by Markov's inequality we have $Y \geq m/36$ with probability at most $108m^{-1/12}$. □

From the previous two claims, we conclude that if m is sufficiently large, then with probability at least $1 - 109m^{-1/12} \geq 1 - m^{-1/24}$ we have $X - Y \geq m/36$. But note that whenever $X - Y \geq m/36$, there are at least $m/36$ interesting $\ell \in \{1, \ldots, m\}$ such that $|P^*_\ell| \geq \sigma$, which implies $|P_\ell| \geq \sigma - (\sigma - \tau) = \tau = \lambda/(4n^4)$. This proves (7.1), and finishes the proof of Lemma 3.2.

8 Concluding remarks

We have proved that the permanent of a random symmetric ±1 matrix typically has magnitude $n^{n/2+o(n)}$. This encapsulates the upper bound in Proposition 1.2 and the lower bound in Theorem 1.3.

The upper bound and lower bound both permit some fairly immediate generalisations. For example, in the setting of Theorem 1.3 we actually have $\Pr(\operatorname{per} M_n = a) \leq \Pr(|\operatorname{per} M_n - a| \leq n^{n/2-\varepsilon n}) \leq n^{-c}$ for all a (not just $a = 0$). To see this, recall that the proof of Theorem 1.3 concludes by repeatedly applying Lemma 3.2, where the final application shows that, conditional on a typical outcome of $\operatorname{per} M_{n-1}$, it is very likely that $|\operatorname{per} M_n| \leq n^{n/2-\varepsilon n}$. In this final application we can instead apply a slight generalisation of Lemma 3.2 (proved in the same way), showing that for any a in fact it is very likely that $|\operatorname{per} M_n - a| \leq n^{n/2-\varepsilon n}$.

We can also permit the entries of our random matrix to take more general distributions. As defined in the introduction, consider any real probability distributions μ and ν, and let $M_n^{\mu,\nu}$ be the random symmetric matrix whose diagonal entries have distribution ν and whose off-diagonal entries have distribution μ (and whose entries on and above the diagonal are mutually independent). If μ and ν are fixed (not depending on n), ν has finite second moment and μ has vanishing first moment and finite fourth moment, then essentially the same proof as for Proposition 1.2 shows that $\Pr\big(|\operatorname{per} M_n^{\mu,\nu}| \geq n^{n/2+\varepsilon n}\big) \leq n^{-\varepsilon n}$ for n sufficiently large with respect to ε, μ and ν.

The conclusion of Theorem 1.3 can be even more freely generalised to any fixed distributions μ and ν such that μ is supported on at least two points (not requiring any moment assumptions at all). In the case where μ is supported on two points, any quadratic polynomial in independent μ-distributed random variables can be rewritten as a multilinear polynomial. One can then use essentially the same proof as for Theorem 1.3 (only changing the various constants in the lemma statements) to prove that there is a constant c (depending on μ but not ν) such that for all $\varepsilon > 0$ and all n sufficiently large with respect to ε, we have $\Pr(|\operatorname{per} M_n^{\mu,\nu}| \leq n^{n/2-\varepsilon n}) \leq n^{-c}$. Note that changing ν has no effect on our arguments, because we never actually use the randomness of the diagonal entries. One needs some slight generalisations of the anti-concentration lemmas in Section 4 for μ-distributed random variables, but these are indeed available; see [5, Theorem A.1] and [23, Theorem 1.8].

In the case where μ is supported on more than two points, then it is necessary to make slightly more involved changes to the proof of Theorem 1.3, but the same result does hold. To give a brief sketch: in this case, the main issue is that in the proof of Lemma 3.2 we are no longer able to assume that the relevant quadratic polynomials are multilinear, so we must treat the square terms x_i^2 in basically the same way we treat the linear terms x_i. To be specific, in the proof of Lemma 3.2, we must allow the polynomials Q_ℓ to contain square terms x_i^2 in addition to linear terms. There are several aspects to this. First, we need to generalise Fact 4.3 and Theorem 4.4 to quadratic polynomials of independent μ-distributed random variables. In particular, we need a generalisation of Theorem 4.4 for quadratic polynomials that are not necessarily multilinear, still giving a bound in terms of the graph matching number $\nu(G^{(r)}(f))$ (where we ignore the square terms in f for the purpose of constructing the graph $G^{(r)}(f)$). Suitable generalisations of Fact 4.3 and Theorem 4.4 can be proved with the methods in [23, Section 4] (in fact, in this section the authors prove [23, Theorem 1.8], which is essentially the required generalisation of Theorem 4.4 but with slightly different assumptions).

Second, we need a generalisation of Fact 4.2 for μ-distributed random variables (for which we can make fairly trivial changes to the proof of Fact 4.2). And lastly, we need a generalisation of the Erdős–Littlewood–Offord theorem (Theorem 4.1) for polynomials of independent μ-distributed random variables (when μ is supported on at least three values) which applies not just to linear polynomials, but also to quadratic polynomials consisting of both linear and square terms (having no multilinear degree-2 terms). Such polynomials can be interpreted as linear polynomials in independent (but not identically distributed) random variables, and therefore the appropriate generalisation of Theorem 4.1 follows from the Doeblin–Lévy–Kolmogorov–Rogozin inequality [27] (or alternatively the method in [23, Section 4]), together with a basic single-variable quadratic anti-concentration bound. Namely, we need the fact that for any real random variable x supported on at least 3 values, and any $r > 0$, there are $s > 0$ and $p > 0$ such that we have $\Pr(|f(x)| < s) < 1 - p$ for any one-variable quadratic polynomial f whose linear or quadratic coefficient has absolute value at least r.

So, for example, both Proposition 1.2 and Theorem 1.3, and therefore Theorem 1.1, can be generalised to the case where μ is a centred Gaussian distribution (as long as μ and ν do not depend on n, and ν has finite second moment). The Gaussian Orthogonal Ensemble (GOE) is an important special case. Another case that may be of particular interest is the case where the support of μ is $\{0, 1\}$, and ν is the trivial distribution always taking the value zero (Proposition 1.2 does not hold in this case, but Theorem 1.3 does). In this case $M_n^{\mu,\nu}$ can be interpreted as the adjacency matrix of a random graph. However, we note that in this case the statement in Theorem 1.3 can be proved in a much simpler way: we can take advantage of the fact that changing any off-diagonal entry in $M_n^{\mu,\nu}$ from 0 to 1 typically causes a large increase in the value of $\operatorname{per} M_n^{\mu,\nu}$, and apply a more general anti-concentration inequality for functions with this property [14, Theorem 1.2]. Actually, in this setting we suspect that it is possible to prove a limit law for $\operatorname{per} M_n^{\mu,\nu}$, using the ideas in [17].

Regarding further directions for research, it would be very interesting to prove stronger upper bounds for the concentration probabilities of the permanent, in both the i.i.d. and the symmetric case. It is currently only known that $\Pr(\operatorname{per} M_n = a)$ and $\Pr(\operatorname{per} A_n = a)$ are bounded by n^{-c} for some constant c. It would be very interesting to prove bounds of the form $n^{-\omega(1)}$, where $\omega(1)$ is any function going to infinity with n. Actually, Vu conjectured (see [40, Conjecture 6.12]) that $\Pr(\operatorname{per} A_n = 0)$ is of the form $\omega(1)^{-n}$ (in contrast to the situation for the determinant, where we have $\Pr(\det A_n = 0) = 2^{-n+o(n)}$). It seems reasonable to conjecture that in fact all probabilities of the form $\Pr(\operatorname{per} M_n = a)$ or $\Pr(\operatorname{per} A_n = a)$ are upper-bounded by n^{-cn}, for some constant c. However, essentially all known tools for studying permanents of random matrices also apply to determinants, so significant new ideas would be required to prove such strong results.

Acknowledgements. We would like to thank Asaf Ferber for insightful discussions. We are also grateful to the referee for their careful reading of the paper, and their useful comments and suggestions.

References

[1] S. Aaronson, *Anti-concentration bound for permanents of Gaussian matrices?*, MathOverflow post, `https://mathoverflow.net/q/45822`, 2010.

[2] S. Aaronson and A. Arkhipov, *The computational complexity of linear optics*, Theory Comput. **9** (2013), 143–252.

[3] N. Alon and J. H. Spencer, *The probabilistic method*, fourth ed., Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2016.

[4] P. Bourgade and K. Mody, *Gaussian fluctuations of the determinant of Wigner matrices*, Electron. J. Probab. **24** (2019), Paper No. 96.

[5] J. Bourgain, V. H. Vu, and P. M. Wood, *On the singularity probability of discrete random matrices*, J. Funct. Anal. **258** (2010), no. 2, 559–603.

[6] M. V. Budrevich and A. E. Guterman, *Kräuter conjecture on permanents is true*, J. Combin. Theory Ser. A **162** (2019), 306–343.

[7] M. V. Budrevich, A. E. Guterman, and K. A. Taranin, *On the divisibility of the permanent of* (± 1)*-matrices*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **439** (2015), no. Chislennye Metody i Voprosy Organizatsii Vychisleniĭ. XXVIII, 26–37.

[8] M. V. Budrevich, A. E. Guterman, and K. A. Taranin, *On the Kräuter-Seifter theorem on the divisibility of permanents*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **463** (2017), no. Chislennye Metody i Voprosy Organizatsii Vychisleniĭ. XXX, 25–35.

[9] M. Campos, L. Mattos, R. Morris, and N. Morrison, *On the singularity of random symmetric matrices*, Duke Math. J. **170** (2021), no. 5, 881–907.

[10] K. P. Costello, T. Tao, and V. Vu, *Random symmetric matrices are almost surely nonsingular*, Duke Math. J. **135** (2006), no. 2, 395–413.

[11] L. Eldar and S. Mehraban, *Approximating the permanent of a random matrix with vanishing mean*, 59th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2018, IEEE Computer Soc., Los Alamitos, CA, 2018, pp. 23–34.

[12] A. Ferber, *Singularity of random symmetric matrices – simple proof*, C. R. Math. Acad. Sci. Paris **359** (2021), no. 6, 743–747.

[13] A. Ferber and V. Jain, *Singularity of random symmetric matrices—a combinatorial approach to improved bounds*, Forum Math. Sigma **7** (2019), e22.

[14] J. Fox, M. Kwan, and L. Sauermann, *Combinatorial anti-concentration inequalities, with applications*, Math. Proc. Cambridge Philos. Soc. **171** (2021), no. 2, 227–248.

[15] V. L. Girko, *A central limit theorem for random determinants*, Teor. Veroyatnost. i Primenen. **24** (1979), no. 4, 728–740.

[16] V. L. Girko, *A refinement of the central limit theorem for random determinants*, Teor. Veroyatnost. i Primenen. **42** (1997), no. 1, 63–73.

[17] S. Janson, *The numbers of spanning trees, Hamilton cycles and perfect matchings in a random graph*, Combin. Probab. Comput. **3** (1994), no. 1, 97–126.

[18] J. Kahn, J. Komlós, and E. Szemerédi, *On the probability that a random* ± 1*-matrix is singular*, J. Amer. Math. Soc. **8** (1995), no. 1, 223–240.

[19] J. Komlós, *On the determinant of* $(0,1)$ *matrices*, Studia Sci. Math. Hungar. **2** (1967), 7–21.

[20] A. R. Kräuter and N. Seifter, *On some questions concerning permanents of* $(1,-1)$*-matrices*, Israel J. Math. **45** (1983), no. 1, 53–62.

[21] A. R. Kräuter and N. Seifter, *Some properties of the permanent of* $(1,-1)$*-matrices*, Linear and Multilinear Algebra **15** (1984), no. 3-4, 207–223.

[22] P. Lundow and K. Markström, *Efficient computation of permanents, with applications to boson sampling and random matrices*, arXiv preprint arXiv:1904.06229 (2019).

[23] R. Meka, O. Nguyen, and V. Vu, *Anti-concentration for polynomials of independent random variables*, Theory Comput. **12** (2016), Paper No. 11.

[24] H. H. Nguyen, *Inverse Littlewood-Offord problems and the singularity of random symmetric matrices*, Duke Math. J. **161** (2012), no. 4, 545–586.

[25] H. H. Nguyen, *On the least singular value of random symmetric matrices*, Electron. J. Probab. **17** (2012), Paper No. 53.

[26] H. H. Nguyen and V. Vu, *Random matrices: law of the determinant*, Ann. Probab. **42** (2014), no. 1, 146–167.

[27] B. A. Rogozin, *On the increase of dispersion of sums of independent random variables*, Teor. Verojatnost. i Primenen **6** (1961), 106–108.

[28] M. Rudelson and R. Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), no. 2, 600–633.

[29] N. Seifter, *Upper bounds for permanents of* $(1, -1)$*-matrices*, Israel J. Math. **48** (1984), no. 1, 69–78.

[30] R. Simion and F. W. Schmidt, *On* $(+1, -1)$*-matrices with vanishing permanent*, Discrete Math. **46** (1983), no. 1, 107–108.

[31] T. Tao and V. Vu, *On random* ± 1 *matrices: singularity and determinant*, Random Structures Algorithms **28** (2006), no. 1, 1–23.

[32] T. Tao and V. Vu, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc. **20** (2007), no. 3, 603–628.

[33] T. Tao and V. Vu, *On the permanent of random Bernoulli matrices*, Adv. Math. **220** (2009), no. 3, 657–669.

[34] T. Tao and V. Vu, *A central limit theorem for the determinant of a Wigner matrix*, Adv. Math. **231** (2012), no. 1, 74–101.

[35] T. Tao and V. H. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2010, Paperback edition.

[36] K. Tikhomirov, *Singularity of random Bernoulli matrices*, Ann. of Math. (2) **191** (2020), no. 2, 593–634.

[37] L. G. Valiant, *The complexity of computing the permanent*, Theoret. Comput. Sci. **8** (1979), no. 2, 189–201.

[38] R. Vershynin, *Invertibility of symmetric random matrices*, Random Structures Algorithms **44** (2014), no. 2, 135–182.

[39] V. Vu, *Random matrices: a survey*, slides for a presentation at an IPAM workshop, `http://campuspress.yale.edu/vanvu/files/2013/01/ipam09.pdf`, 2009.

[40] V. H. Vu, *Recent progress in combinatorial random matrix theory*, Probab. Surv. **18** (2021), 179–200.

[41] E. T.-H. Wang, *On permanents of* $(1, -1)$*-matrices*, Israel J. Math. **18** (1974), 353–361.

[42] I. M. Wanless, *Permanents of matrices of signed ones*, Linear Multilinear Algebra **53** (2005), no. 6, 427–433.

[43] E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math. (2) **62** (1955), 548–564.

[44] E. P. Wigner, *On the distribution of the roots of certain symmetric matrices*, Ann. of Math. (2) **67** (1958), 325–327.