

A GENERAL LAW OF LARGE PERMANENT

JÓZSEF BALOGH AND HOI NGUYEN

ABSTRACT. In this short note we establish a law of large permanent for matrices with entries from an \mathbf{N}^2 -indexed stochastic process. This answers a question by Bochi, Iommi and Ponce in [4].

1. INTRODUCTION

Let $M_n = (m_{ij})_{1 \leq i, j \leq n}$ be a square matrix of size n of real-valued entries. The permanent of M_n is defined as

$$\mathbf{perm}(M_n) := \sum_{\pi \in S_n} m_{1\pi(1)} \cdots m_{n\pi(n)}.$$

Let Ω_n denote the set of doubly stochastic matrices $M_n = (m_{ij})_{1 \leq i, j \leq n}$ of size n , that is $0 \leq m_{ij} \leq 1$ and $\sum_{j=1}^n m_{ij} = 1$ for every $1 \leq i \leq n$ and $\sum_{i=1}^n m_{ij} = 1$ for every $1 \leq j \leq n$. It is well-known that if $M_n \in \Omega_n$ then

$$\frac{n!}{n^n} \leq \mathbf{perm}(M_n) \leq 1. \tag{1}$$

The upper bound of (1) is elementary, which is attained at permutation matrices; one can also obtain a stability-type result in this direction: for instance [1, Theorem A.1] shows that if $\mathbf{perm}(M_n) \geq n^{-O(1)}$, then all but at most $O(\log n)$ rows (and columns) of M_n contain an entry that is at least 0.9.

The lower bound of (1) was conjectured by van der Waerden in 1926 [29] and proved in 1981 independently by Egorychev [9] and Falikman [10]. Moreover, the minimum of the permanent on Ω_n is attained at the matrix J_n of entries $1/n$.

Following [4], we denote the *permanental mean* of a matrix M_n by

2010 *Mathematics Subject Classification.* 15A15, 26E60, 60B20.

Key words and phrases. Permanent, law of large number.

J. Balogh is partially supported by NSF grant DMS-1500121, Arnold O. Beckman Research Award (UIUC Campus Research Board 15006) and Langan Professional Scholarship. H. Nguyen is partially supported by NSF grant DMS-1600782.

$$\mathbf{pm}(M_n) := \left(\frac{\mathbf{perm}(M_n)}{n!} \right)^{1/n}.$$

Thus for doubly stochastic matrices $M_n \in \Omega_n$, by (1)

$$1 \leq \lim_{n \rightarrow \infty} \mathbf{pm}(nM_n) \leq e. \quad (2)$$

Note that by definition, if $M_n = D_n G_n E_n$ where $D_n = \text{diag}(d_i)_{1 \leq i \leq n}$, $E_n = \text{diag}(e_i)_{1 \leq i \leq n}$ with $d_i, e_i > 0$ (in other words, $m_{ij} = d_i g_{ij} e_j$), then

$$\mathbf{pm}(M_n) = \mathbf{gm}(\mathbf{d}) \mathbf{pm}(G_n) \mathbf{gm}(\mathbf{e}), \quad (3)$$

where $\mathbf{d} = (d_1, \dots, d_n)$ and $\mathbf{e} = (e_1, \dots, e_n)$, and $\mathbf{gm}(\cdot)$ is the geometric mean

$$\mathbf{gm}(\mathbf{d}) := \left(\prod_i d_i \right)^{1/n}, \text{ and } \mathbf{gm}(\mathbf{e}) := \left(\prod_i e_i \right)^{1/n}.$$

In other words, (3) says that the permanental mean $\mathbf{pm}(\cdot)$ is homogeneous under matrix scaling.

In this note we will try to relate the permanental mean for matrices of non-negative entries to the so-called *scaling mean* $\mathbf{sm}(M_n)$, which is in turn defined as

$$\mathbf{sm}(M_n) := \frac{1}{n^2} \inf_{\mathbf{x}, \mathbf{y} \in \mathbf{R}_{>0}^n} \frac{\mathbf{x}^T M_n \mathbf{y}}{\mathbf{gm}(\mathbf{x}) \mathbf{gm}(\mathbf{y})},$$

where $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$.

Thus the scaling mean can be obtained via an optimization problem. An extremely nice property of the scaling mean for doubly stochastic matrices M_n is that (by using AM-GM inequality)

$$\mathbf{sm}(nM_n) = 1. \quad (4)$$

Another important property, which is again not hard to show, is that $\mathbf{sm}(\cdot)$ is also homogeneous under matrix scaling. In other words, if $M_n = D_n G_n E_n$, with $D_n = \text{diag}(d_i)_{1 \leq i \leq n}$, $E_n = \text{diag}(e_i)_{1 \leq i \leq n}$ and $d_i, e_i > 0$, then

$$\mathbf{sm}(M_n) = \mathbf{gm}(\mathbf{d}) \mathbf{sm}(G_n) \mathbf{gm}(\mathbf{e}). \quad (5)$$

The forms $M_n = D_n G_n E_n$ in (3) and (5), with doubly stochastic G_n , is called *Sinkhorn decomposition*. Not every matrix M_n of non-negative entries can be scaled back to a doubly stochastic matrix through Sinkhorn decompositions. However, the following beautiful theorem provides a necessary and sufficient condition.

Theorem 1.1. [7, 21, 27] *A matrix $M_n = (a_{ij})_{1 \leq i, j \leq n}$ with non-negative entries has a Sinkhorn decomposition $M_n = D_n G_n E_n$ if and only if for each positive element $m_{ij} > 0$ there exists a permutation $\pi \in S_n$ such that $\pi(i) = j$ and $m_{1\pi(1)}, \dots, m_{n\pi(n)}$ are all positive. Moreover, the doubly stochastic matrix G_n is unique and the map $M_n \rightarrow G_n$ is continuous.*

We refer the reader to [4, 12, 14, 22, 26] for the history and further developments of Theorem 1.1, and also to [20, 30] for algorithmic aspects of this fundamental decomposition.

We put here together the relations between (2),(3), (4) and (5).

Claim 1.2. *Assume that $\{M_n\}_{1 \leq n < \infty}$ is a sequence of matrices of non-negative entries with a Sinkhorn decomposition $M_n = D_n G_n E_n$. Then*

$$1 \leq \lim_{n \rightarrow \infty} \frac{\mathbf{pm}(M_n)}{\mathbf{sm}(M_n)} \leq e.$$

1.3. Main result. One of the main goals of our note, and also of the mentioned paper [4] by Bochi, Iommi and Ponce, is to show that the limit in Claim 1.2 is exactly one in the general context of ergodic theory.

To prepare for the main statement, we still need to introduce the scaling limit for functions. Fix a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. Let $\mathcal{G}(\mathbf{P})$ denote the set of positive measurable functions $\varphi : \Omega \rightarrow \mathbf{R}_{>0}$ such that $\log \varphi \in L_1(\mathbf{P})$. The *geometric mean* of φ is defined as

$$\mathbf{gm}(\varphi) := \exp\left(\int \log \varphi d\mathbf{P}\right).$$

Let us also fix a pair of σ -algebras $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$. For $i = 1, 2$, we define

$$\mathcal{G}_i := \left\{ \varphi : \Omega \rightarrow \mathbf{R}_{>0} : \varphi \text{ is } \mathcal{A}_i\text{-measurable and } \log \varphi \in L^1(\mathbf{P}) \right\}.$$

The *scaling limit* of a non-negative measurable function $f : \Omega \rightarrow \mathbf{R}$ with respect to $\mathcal{A}_1, \mathcal{A}_2$ is then defined as

$$\mathbf{sm}_{\mathcal{A}_1, \mathcal{A}_2}(f) := \inf_{g_i \in \mathcal{G}_i} \frac{1}{\mathbf{gm}(g_1)\mathbf{gm}(g_2)} \int g_1 f g_2 d\mathbf{P}.$$

Now we are ready to introduce the beautiful result by Bochi, Iommi and Ponce from [4].

Assume that (X, \mathcal{X}, μ) and (Y, \mathcal{Y}, ν) are Lebesgue probability spaces, and $S : X \rightarrow X$, $T : Y \rightarrow Y$ are measure preserving transformations. Given a function $f : X \times Y \rightarrow \mathbf{R}^+$, for each $(x, y) \in X \times Y$ and for each integer n we define the matrix $\square_n f(x, y)$ to be

$$\square_n f(x, y) := (f(S^i x, T^j y))_{0 \leq i, j \leq n-1}.$$

Let \mathcal{A}_1 and \mathcal{A}_2 be the sub- σ -algebras formed by the S -invariant and the T -invariant sets respectively. Let $\mathcal{B}(\mu \times \nu)$ denote the set of positive measurable functions on $X \times Y$ which are essentially bounded away from zero and infinity.

Theorem 1.4 (Law of large permanent). [4, Theorem 4.1] *If S and T are ergodic and $f \in \mathcal{B}(\mu \times \nu)$ then for $\mu \times \nu$ -almost every $(x, y) \in X \times Y$*

$$\lim_{n \rightarrow \infty} \frac{\mathbf{pm}(\square_n f(x, y))}{\mathbf{sm}_{\mathcal{A}_1, \mathcal{A}_2}(f)} = 1.$$

In particular, Theorem 1.4 not only shows the $(\mu \times \nu)$ -a. e. existence of the limit of the permanental mean of $\square_n f(x, y)$, but it also indicates that this limit is precisely the scaling mean of f , therefore the result connects the limit to an optimization problem.

We also invite the reader to [4, Section 5] for applications to Muirhead means as well as to a classical result of Halász and Székely [18].

We now introduce a generalization of Theorem 1.4. Suppose that T is an ergodic measure-preserving action on the semigroup \mathbf{N}^2 on a Lebesgue probability space $(\Omega, \mathcal{A}, \mathbf{P})$. Given a function $f : \Omega \rightarrow \mathbf{R}^+$, define the matrix $\square_n f(\omega)$ to be

$$\square_n f(\omega) := (f(T^{(i,j)}(\omega)))_{0 \leq i, j \leq n-1}.$$

Let \mathcal{A}_1 and \mathcal{A}_2 be the sub- σ -algebras formed by the $T^{(1,0)}$ -invariant and the $T^{(0,1)}$ -invariant sets respectively. The following was conjectured in 1.4.

Conjecture 1.5 (Law of large permanent, another version). *If $\log f \in L_\infty(\mathbf{P})$ then for \mathbf{P} -almost every ω*

$$\lim_{n \rightarrow \infty} \frac{\mathbf{pm}(\square_n f(\omega))}{\mathbf{sm}_{\mathcal{A}_1, \mathcal{A}_2}(f)} = 1.$$

Note that this conjecture would imply Theorem 1.4 with $S = T^{(1,0)}$ and $T = T^{(0,1)}$. Our result confirms this conjecture.

Theorem 1.6 (Main result). *Conjecture 1.5 holds.*

The rest of the note is devoted to prove Theorem 1.6. Although we will use some important ingredients from the paper [4] by Bochi, Iommi and Ponce, our key approach is quite

different from theirs. Roughly speaking, the proof consists of three steps: (1) reduction to doubly stochastic functions, (2) passing to doubly stochastic matrices of bounded entries, (3) establishing upper bound for permanents of such matrices.

2. PROOF OF THEOREM 1.6: PASSING TO DOUBLY STOCHASTIC FUNCTIONS

We record here other key properties of the scaling mean, these are functional analogues of the results introduced in Section 1.

Theorem 2.1. *Fix a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ and a pair of σ -algebras $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$. The following holds.*

- (Homogeneity) *If $\varphi \in \mathcal{G}_1$ and $\psi \in \mathcal{G}_2$ then*

$$\mathbf{sm}_{\mathcal{A}_1, \mathcal{A}_2}(\varphi g \psi) = \mathbf{gm}(\varphi) \mathbf{sm}_{\mathcal{A}_1, \mathcal{A}_2}(g) \mathbf{gm}(\psi). \quad (6)$$

- (Restriction to L_∞) *Let*

$$\mathcal{B}_i := \left\{ h : \Omega \rightarrow \mathbf{R}^+ : h \text{ is } \mathcal{A}_i\text{-measurable and } \log h \in L_\infty(\mathbf{P}) \right\}.$$

Then

$$\mathbf{sm}_{\mathcal{A}_1, \mathcal{A}_2}(f) = \inf_{g_i \in \mathcal{B}_i} \frac{1}{\mathbf{gm}(g_1) \mathbf{gm}(g_2)} \int g_1 f g_2 d\mathbf{P}. \quad (7)$$

- (Doubly stochastic) *If an integrable non-negative function $g, g : \Omega \rightarrow \mathbf{R}_{\geq 0}$, is doubly stochastic with respect to \mathcal{A}_1 and \mathcal{A}_2 , that is*

$$\mathbf{E}(g|\mathcal{A}_1) = \mathbf{E}(g|\mathcal{A}_2) = 1, \quad \mathbf{P} - \text{almost everywhere,}$$

then we have the following analogue of (4)

$$\mathbf{sm}_{\mathcal{A}_1, \mathcal{A}_2}(g) = 1. \quad (8)$$

We refer the reader to Propositions 3.2 and 3.3 of [4] for proofs of these results.

Next, we will also need a functional version of Theorem 1.1 regarding the Sinkhorn decomposition.

Theorem 2.2. [4, Theorem 3.6] *Every $f : \Omega \rightarrow \mathbf{R}^+$ such that $\log f \in L_\infty(\mathbf{P})$ has a Sinkhorn decomposition, that is there exist functions $\varphi \in \mathcal{B}_1, \psi \in \mathcal{B}_2$ and g doubly stochastic with respect to \mathcal{A}_1 and \mathcal{A}_2 such that for \mathbf{P} -almost every ω*

$$f(\omega) = \varphi(\omega)g(\omega)\psi(\omega). \quad (9)$$

By Theorem 2.2, for \mathbf{P} -almost every ω we can write $f(\omega) = \varphi(\omega)g(\omega)\psi(\omega)$ for some $\varphi \in \mathcal{B}_1, \psi \in \mathcal{B}_2$, and g doubly stochastic, and so

$$f(T^{(i,j)}(\omega)) = \varphi(T^{(i,j)}(\omega))g(T^{(i,j)}(\omega))\psi(T^{(i,j)}(\omega)).$$

Now, as φ is \mathcal{A}_1 -measurable, for \mathbf{P} -almost every ω the following holds for any fixed j and for every $1 \leq i \leq n$

$$\varphi(T^{(i,j)}(\omega)) = \varphi(T^{(i,0)}(T^{(0,j)}(\omega))) = \varphi(T^{(1,0)}(T^{(0,j)}(\omega))) = \varphi(T^{(1,j)}(\omega)).$$

Similarly, because ψ is \mathcal{A}_2 -measurable and $T^{(1,0)}$ and $T^{(0,1)}$ commute, for any fixed i and for every $1 \leq j \leq n$

$$\psi(T^{(i,j)}(\omega)) = \psi(T^{(i,0)}(T^{(0,j)}(\omega))) = \psi(T^{(i,0)}(T^{(0,1)}(\omega))) = \psi(T^{(i,1)}(\omega)).$$

As a consequence, for \mathbf{P} -almost every ω

$$\mathbf{pm}(\square_n f(\omega)) = \left(\prod_{i=0}^{n-1} \varphi(T^{(i,1)}(\omega)) \right)^{1/n} \left(\prod_{i=0}^{n-1} \psi(T^{(1,i)}(\omega)) \right)^{1/n} \mathbf{pm}(\square_n g(\omega)).$$

Now by the ergodic theorem

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{0 \leq i \leq n-1} \log \varphi(T^{(i,1)}(\omega)) = \log \mathbf{gm}(\varphi),$$

and similarly

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{0 \leq j \leq n-1} \log \psi(T^{(1,j)}(\omega)) = \log \mathbf{gm}(\psi), \text{ for } \mathbf{P} - \text{almost every } \omega.$$

Thus by (6) of Theorem 2.1, it suffices to establish Theorem 1.6 for doubly stochastic function g . In other words, by (8) of Theorem 2.1 we will need to show the following.

Theorem 2.3. *Assume that $\log g \in L_\infty(\mathbf{P})$ and g is doubly stochastic with respect to \mathcal{A}_1 and \mathcal{A}_2 . Then*

$$\lim_{n \rightarrow \infty} \mathbf{pm}(\square_n g(\omega)) = 1 \text{ for } \mathbf{P} - \text{almost every } \omega. \quad (10)$$

From now on we assume that there exists $\lambda > 1$ such that for \mathbf{P} -almost every ω

$$\lambda^{-1} \leq g(\omega) \leq \lambda.$$

All of the implied constants below are allowed to depend on λ . For Theorem 2.3, by a limiting argument, it suffices to show the following asymptotic analog.

Theorem 2.4. *For any $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon, \lambda)$ such that for any $n \geq n_0$, there exists a measurable set \mathcal{E}_n of measure at most ε such that for all $\omega \notin \mathcal{E}_n$ we have*

$$1 - C\varepsilon \leq \mathbf{pm}(\square_n g(\omega)) \leq 1 + C\varepsilon,$$

where C is a constant depending on λ .

3. PROOF OF THEOREM 1.6: APPROXIMATION BY DOUBLY STOCHASTIC MATRICES OF BOUNDED ENTRIES

We next show that most of the row sums and column sums of the matrix $\frac{1}{n}\square_n g(\omega)$ are asymptotically the same.

Lemma 3.1. *With an exception of at most εn rows and columns, the following holds for the rows i and columns j of the matrix $\square_n g(\omega)$*

$$(1 - \varepsilon)n \leq \sum_{k=0}^{n-1} g(T^{(i,k)}(\omega)), \quad \sum_{k=0}^{n-1} g(T^{(k,j)}(\omega)) \leq (1 + \varepsilon)n.$$

Proof. (of Lemma 3.1) Set

$$g_n^{(1)}(\omega) := \frac{1}{n} \sum_{i=0}^{n-1} g(T^{(i,0)}(\omega)) \quad \text{and} \quad g_n^{(2)}(\omega) := \frac{1}{n} \sum_{i=0}^{n-1} g(T^{(0,i)}(\omega)).$$

By Birkhoff's ergodic theorem in L_1 (see for instance [19, Theorem 2.1.5]),

$$\lim_{n \rightarrow \infty} \frac{1}{n} g_n^{(1)} \rightarrow \mathbf{E}(g|\mathcal{A}_1) = 1 \text{ in } L_1(\mathbf{P}),$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{n} g_n^{(2)} \rightarrow \mathbf{E}(g|\mathcal{A}_2) = 1 \text{ in } L_1(\mathbf{P}),$$

where we used the fact that g is doubly stochastic with respect to \mathcal{A}_1 and \mathcal{A}_2 .

Thus for any $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon)$ such that for $n \geq n_0$ we have

$$\int_{\Omega} |g_n^{(1)}(\omega) - 1| d\mathbf{P}(\omega) \leq \varepsilon^4 \text{ and } \int_{\Omega} |g_n^{(2)}(\omega) - 1| d\mathbf{P}(\omega) \leq \varepsilon^4. \quad (11)$$

Next, define

$$\mathcal{E}_n^{(1)} := \left\{ \omega : \frac{1}{n} \sum_{k=0}^{n-1} |g_n(T^{(0,k)}(\omega)) - 1| > \varepsilon^2 \right\}$$

as well as

$$\mathcal{E}_n^{(2)} := \left\{ \omega : \frac{1}{n} \sum_{k=0}^{n-1} |g_n(T^{(k,0)}(\omega)) - 1| > \varepsilon^2 \right\}.$$

By Markov's bound,

$$\begin{aligned} \mathbf{P}(\mathcal{E}_n^{(1)}) &\leq \varepsilon^{-2} \int_{\Omega} \frac{1}{n} \sum_{k=0}^{n-1} |g_n^{(1)}(T^{(0,k)}(\omega)) - 1| d\mathbf{P}(\omega) \\ &\leq \varepsilon^{-2} \frac{1}{n} \sum_{k=0}^{n-1} \int_{\Omega} |g_n^{(1)}(T^{(0,k)}(\omega)) - 1| d\mathbf{P}(\omega) \\ &= \varepsilon^{-2} \int_{\Omega} |g_n^{(1)}(\omega) - 1| d\mathbf{P}(\omega) \leq \varepsilon^2, \end{aligned}$$

where we just used the fact that $T^{(0,1)}$ is measure preserving together with the bound (11) on the L_1 -norm of $g_n^{(1)} - 1$.

Similarly, we also have

$$\mathbf{P}(\mathcal{E}_n^{(2)}) \leq \varepsilon^2.$$

Let $\omega \in \Omega \setminus (\mathcal{E}_n^{(1)} \cup \mathcal{E}_n^{(2)})$. By definition

$$\frac{1}{n} \sum_{k=0}^{n-1} |g_n^{(1)}(T^{(0,k)}(\omega)) - 1| < \varepsilon^2.$$

Thus by averaging, for all but at most εn indices $k \in \{0, \dots, n-1\}$, $|g_n^{(1)}(T^{(0,k)}(\omega)) - 1| < \varepsilon$. In other words, by the definition of g_n

$$\left| \sum_{i=0}^{n-1} g(T^{(i,k)}(\omega)) - n \right| \leq \varepsilon n.$$

Similarly, for all but at most εn indices $k \in \{0, \dots, n-1\}$,

$$\left| \sum_{i=0}^{n-1} g(T^{(k,i)}(\omega)) - n \right| \leq \varepsilon n.$$

□

As we have seen from Lemma 3.1, most of the row sums and column sums of the matrix $\square_n(\omega)$ are asymptotically $(1 + o(1))n$. In the next lemma we show that this matrix can be approximated by a genuine doubly stochastic matrix in L_1 -norm.

Lemma 3.2. *Let $0 < \varepsilon < 1 < \lambda$ be given positive constants, where ε is sufficiently small depending on λ . Suppose that X_n is a matrix with the following properties*

- $\lambda^{-1} \leq x_{ij} \leq \lambda$;
- all but at most εn rows and columns of X_n have sum belonging to the range $[(1 - \varepsilon)n, (1 + \varepsilon)n]$.

Then there exists $X'_n \in n \cdot \Omega_n$ such that $(2\lambda)^{-1} \leq x'_{ij} \leq 2\lambda$ for all i, j and

$$\sum_{1 \leq i, j \leq n} |x_{ij} - x'_{ij}| \leq 16\varepsilon\lambda^2 n^2.$$

Proof. (of Lemma 3.2) We first completely truncate the same number, assuming the worse case $\lceil \varepsilon n \rceil$, of rows and columns whose sums were not in the range $[(1 - \varepsilon)n, (1 + \varepsilon)n]$. The obtained square matrix has size $m = n - \lceil \varepsilon n \rceil$ with row and column sums belonging to $[(1 - \varepsilon)n - \lambda \lceil \varepsilon n \rceil, (1 + \varepsilon)n] \subset [(1 - 2\lambda\varepsilon)n, (1 + \varepsilon)n]$. We next multiply each row of the obtained matrix by an appropriate factor from the interval $[1 - 2\varepsilon, 1 + 2\lambda\varepsilon]$ to make the row sum exactly m . Let the new matrix be $Y_m = (y_{ij})_{1 \leq i, j \leq m}$, whose properties are summarized below when ε was chosen sufficiently small:

- every row \mathbf{r}_i for $1 \leq i \leq m$ has sum $s(\mathbf{r}_i)$ exactly m ;
- every column sum for $s(\mathbf{c}_j), 1 \leq j \leq m$ belongs to the range $[(1 - 4\lambda\varepsilon)m, (1 + 4\lambda\varepsilon)m]$,
- for all i, j we have $(2\lambda)^{-1} \leq y_{ij} \leq 2\lambda$.

We now approximate Y_m by matrices from $m \cdot \Omega_m$.

Claim 3.3. *There exists a matrix $Y'_m \in m \cdot \Omega_m$ such that*

$$\sum_{1 \leq i, j \leq m} |y_{ij} - y'_{ij}| \leq 4\varepsilon\lambda^2 m^2.$$

It is clear that by the construction of Y_m , after gluing nI_{n-m} to $\frac{n}{m}Y'_m$ one creates a matrix $X'_n \in n \cdot \Omega_n$ which approximates X_n as desired.

It remains to prove Claim 3.3. We are going to modify the column vectors of Y_m so that they all have sum m . Let \mathbf{c}_i and \mathbf{c}_j be two columns with

$$s(\mathbf{c}_i) < m < s(\mathbf{c}_j).$$

Case 1. Assume for now that $s(\mathbf{c}_i) + s(\mathbf{c}_j) \geq 2m$. We are going to modify the entries of $\mathbf{c}_i, \mathbf{c}_j$ as follows: for $k = 1$ to m , consider the pair (y_{ki}, y_{kj}) . Increase y_{ki} by a largest possible amount $y_k \geq 0$ (and decrease y_{kj} by y_k accordingly to preserve the row sum) so that $s(\mathbf{c}_i)$ is still below m and then entries are still within $(2\lambda)^{-1} \leq y'_{ki}, y'_{kj} \leq 2\lambda$.

We claim that $s(\mathbf{c}_i) = m$ after modifying all $(y_{ki}, y_{kj}), 1 \leq k \leq m$. Assume otherwise, then the reason we are not able to increase $s(\mathbf{c}_i)$ furthermore is that for any $1 \leq k \leq m$, either $y_{ki} = 2\lambda$ or $y_{kj} = (2\lambda)^{-1}$. But in either case, $y_{ki} \geq y_{kj}$, and so $s(\mathbf{c}_i)$ must be at least $s(\mathbf{c}_j)$, a contradiction to our assumption that $s(\mathbf{c}_i) + s(\mathbf{c}_j) \geq 2m$.

Case 2. For the case $s(\mathbf{c}_i) + s(\mathbf{c}_j) < 2m$, for $k = 1$ to m we decrease y_{kj} by a largest possible amount $y_k \geq 0$ (and increase y_{ki} by y_k accordingly) so that $s(\mathbf{c}_j)$ is still at least m , and $(2\lambda)^{-1} \leq y'_{ki}, y'_{kj} \leq 2\lambda$. Again it is not hard to check that after modifying all $(y_{ki}, y_{kj}), 1 \leq k \leq m$, we will obtain $s(\mathbf{c}_j) = m$.

Note that because of the nature of our process (as we always either increase or decrease all entries of one column), the L_1 -distance of the new matrix B'_m and the original matrix Y_m is bounded by

$$\sum_{i,j} |y_{ij}(Y_m) - y_{ij}(Y'_m)| \leq 4\varepsilon\lambda^2 m.$$

Now consider the new matrix, if it still does not belong to $m \cdot \Omega_m$ then choose any column pair $(\mathbf{c}_{i'}, \mathbf{c}_{j'})$ with $s(\mathbf{c}_{i'}) < m < s(\mathbf{c}_{j'})$ and continue the modifying process as above.

After at most m such iterations, our matrix must belong to $m \cdot \Omega_m$ because the number of columns of sum m increase by at least one after each iteration. \square

It is plausible to obtain X'_n by considering the Sinkhorn decomposition of X_n but we have not followed this approach. However, in our simple proof above we did use the trick of scaling the rows appropriately.

For convenience, we gather here an immediate consequence of Lemma 3.1 and Lemma 3.2.

Corollary 3.4. *There exists a matrix $A_n(\omega) = (a_{ij}(\omega))_{1 \leq i, j \leq n} \in n \cdot \Omega_n$ such that*

$$(2\lambda)^{-1} \leq a_{ij}(\omega) \leq 2\lambda, \text{ and } \frac{1}{n^2} \sum_{1 \leq i, j \leq n} |g(T^{(i,j)}(\omega)) - a_{ij}(\omega)| \leq 16\varepsilon\lambda^2.$$

We next apply the following nice result by Bochi, Iommi and Ponce [4, Lemma 4.4].

Lemma 3.5. *If X_n and Y_n are matrices with $\lambda^{-1} \leq x_{ij}, y_{ij} \leq \lambda$, then*

$$\left| \log \frac{\mathbf{pm}(X_n)}{\mathbf{pm}(Y_n)} \right| \leq \frac{\lambda^5}{n^2} \sum_{i,j} |x_{ij} - y_{ij}|.$$

By Corollary 3.4 and Lemma 3.5 above, to prove Theorem 2.4 we just need to show that, with A_n as in Corollary 3.4,

$$\mathbf{pm}(A_n) = (1 + o(1)).$$

This is the content of another question posed by Bochi, Iommi and Ponce.

Theorem 3.6. [4, Conjecture 6.2] *Assume that $\{A_n = (a_{ij})\}_{1 \leq n < \infty}$ is a sequence of matrices of increasing size n such that $A_n \in n \cdot \Omega_n$. Assume furthermore that there exists $\lambda > 1$ so that $\lambda^{-1} \leq a_{ij} \leq \lambda, 1 \leq i, j \leq n$. Then*

$$\lim_{n \rightarrow \infty} \mathbf{pm}(A_n) = 1.$$

Clearly, by the van der Waerden bound (1), for Theorem 3.6 the main task is to bound $\mathbf{pm}(A_n)$ from above. In the next section we will give a simple proof of it, and hence concluding the proof of Theorem 1.6.

4. PROOF OF THEOREM 1.6: PERMANENT OF DOUBLY STOCHASTIC MATRICES OF BOUNDED ENTRIES

Let \mathcal{S}_n denote the set of (row) stochastic matrices $M_n = (m_{ij})_{1 \leq i, j \leq n}$ of size n , that is $0 \leq m_{ij} \leq 1$ and $\sum_{j=1}^n m_{ij} = 1$ for every $1 \leq i \leq n$. The following result immediately implies Theorem 3.6.

Theorem 4.1. *Suppose that $\{A_n\}_{1 \leq n < \infty}$ is a sequence of matrices of increasing size n such that $A_n \in n \cdot \mathcal{S}_n$. Suppose furthermore that there exists $\lambda > 1$ so that $a_{ij} \leq \lambda$ for all i, j . Then*

$$\mathbf{perm}(A_n) \leq e^{2\lambda} n^{(\lambda-1)/2} n!.$$

It remains to prove Theorem 4.1. Denote \mathcal{Ext}_n the collection of matrices $A_n = (a_{ij})$ where $A_n \in n \cdot \mathcal{S}_n, a_{ij} \leq \lambda$ and $\mathbf{perm}(A_n)$ is maximum. Note that \mathcal{Ext}_n is non-empty because of the compactness.

Lemma 4.2. *The family \mathcal{Ext}_n contains of a matrix $A_n = (a_{ij})_{1 \leq i, j \leq n}$ with the following properties: for every row $1 \leq i \leq n$, all but at most one entry take values either zero or λ , and the (possible) remaining entry is strictly between zero and λ .*

Proof. (of Lemma 4.2) It suffices to work with the first row. We show that for any $A_n \in \mathcal{Ext}_n$, there is a way to force all but at most one of the entries a_{11}, \dots, a_{1n} of its first row to be either zero or λ , while keeping $\mathbf{perm}(A_n)$ to be optimal.

In what follows we freeze all a_{ij} with $2 \leq i \leq n, 1 \leq j \leq n$. For $1 \leq i \leq n$, let M_i be to $(n-1) \times (n-1)$ minor obtained from A_n by deleting its first row and i -th column. By the Laplace expansion

$$\mathbf{perm}(A_n) = \sum_{i=1}^n a_{1i} \mathbf{perm}(M_i).$$

Assume that there are two entries among the a_{11}, \dots, a_{1n} that are not either zero or λ . Without loss of generality, assume that these are a_{11} and a_{12} . Assume furthermore that

$$0 \leq \mathbf{perm}(M_1) \leq \mathbf{perm}(M_2).$$

Given the constraint $a_{11} + a_{12} = n - \sum_{3 \leq i \leq n} a_{1i}$ and $0 \leq a_{11}, a_{12} \leq \lambda$ (with temporarily fixed $a_{1i}, 3 \leq i \leq n$), it is easy to see that the sum $a_{11} \mathbf{perm}(M_1) + a_{12} \mathbf{perm}(M_2)$ is bounded from above by

$$a_{11} \mathbf{perm}(M_1) + a_{12} \mathbf{perm}(M_2) \leq \begin{cases} 0 \times \mathbf{perm}(M_1) + (a_{11} + a_{12}) \mathbf{perm}(M_2), & a_{11} + a_{12} \leq \lambda \\ (a_{11} + a_{12} - \lambda) \mathbf{perm}(M_1) + \lambda \mathbf{perm}(M_2), & \lambda < a_{11} + a_{12} < 2\lambda. \end{cases}$$

In other words, we do not decrease $\mathbf{perm}(A_n)$ by shifting $(a_{11}, a_{12}, a_{13}, \dots, a_{1n})$ to either $(0, a_{11} + a_{12}, a_{13}, \dots, a_{1n})$ or $(a_{11} + a_{12} - \lambda, \lambda, a_{13}, \dots, a_{1n})$. Remark that in either case we increase the number of entries taking values zero or λ in the first row of A_n . \square

Let $A_n \in \mathcal{Ext}_n$ be a matrix obtained by applying Lemma 4.2. After replacing in each row the exceptional entry by λ (if needed), we obtain a matrix A'_n whose each row contains $m = \lceil \frac{n}{\lambda} \rceil$ entries of value λ and $n - m$ entries of value zero. Also $\mathbf{perm}(A_n) \leq \mathbf{perm}(A'_n)$.

After scaling down the entries of A'_n by a factor of λ (and so the permanent is scaled down by a factor of λ^n), we obtain a $\{0, 1\}$ -matrix B_n where the number of ones in each row is exactly m . We next apply Bregman-Minc inequality for permanent of $\{0, 1\}$ matrices (see for instance [6, 23] or [2, (2.1)])

Theorem 4.3. *Let $M_n = (m_{ij})_{1 \leq i, j \leq n}$ be a $\{0, 1\}$ -matrix. Denote $r_i = \sum_{j=1}^n m_{ij}, 1 \leq i \leq n$. Then we have*

$$\mathbf{perm}(M_n) \leq \prod_{i=1}^n (r_i!)^{\frac{1}{r_i}},$$

where equality holds if and only if up to permutation of rows and columns M_n is a block diagonal matrix where each block is a square matrix of all ones. In other words, equality holds only when M_n is an adjacency matrix of disjoint union of bipartite K_{r_i, r_i} graphs.

Applying Theorem 4.3 to B_n we obtain

$$\mathbf{perm}(B_n) \leq (m!)^{n/m}, \tag{12}$$

with equality holds only when B_n is an adjacency matrix of disjoint union of $K_{m,m}$ graphs.

Proof. (of Theorem 4.1) We have

$$\mathbf{perm}(A_n) \leq \mathbf{perm}(A'_n) = \lambda^n \mathbf{perm}(B_n) \leq (\lambda)^n (m!)^{n/m}.$$

Note that by Stirling's approximation $\sqrt{2\pi k} \left(\frac{k}{e}\right)^k \leq k! \leq e\sqrt{k} \left(\frac{k}{e}\right)^k$. Thus,

$$\begin{aligned} \mathbf{perm}(A_n) &\leq \lambda^n (m!)^{n/m} \leq \lambda^n \left(e\sqrt{m} \left(\frac{m}{e}\right)^m \right)^{n/m} \leq (e\sqrt{m})^\lambda \left(\frac{n+\lambda}{e}\right)^n \\ &\leq (e\sqrt{m})^\lambda e^\lambda \left(\frac{n}{e}\right)^n \leq e^{2\lambda} n^{(\lambda-1)/2} n!. \end{aligned}$$

□

Added to proof. After the proof of Theorem 4.1 was written, we were informed that the result follows from [28]. However, as our proof looks short and direct, we decided to keep it here for completeness.

5. FURTHER REMARKS

A crucial problem is to calculate the scaling mean $\mathbf{sm}_{\mathcal{A}_1, \mathcal{A}_2}(f)$ for various natural candidates of T and f . We refer the reader to [4] and [5] for many illuminating examples as well as for a fast and simple iterative process regarding this issue.

Theorem 1.4 and Theorem 1.6 can be considered as law of large number. It remains an interesting problem to extend these results to central limit theorem for the logarithmic permanents. We hope to address this challenging issue in the near future. In what follows we gather two small applications of our main result.

5.1. Determinant of gaussian matrices with different variances. By taking advantage of the explicit approximation $n!$ of $\mathbf{perm}(A_n)$ in Theorem 4.1, we deduce that the logarithmic of square determinant of random gaussian matrices with variance profile A_n is concentrated around $\log n!$. This is by no means fundamental, but we have not found similar results in the literature.

Let $A_n = (a_{ij})_{1 \leq i, j \leq n}$ be a deterministic matrix with non-zero entries. Assume that $X_n = (x_{ij})_{1 \leq i, j \leq n}$ is a random matrix where the entries x_{ij} are of the form $x_{ij} = \sqrt{a_{ij}} g_{ij}$, with g_{ij} being independent and identically distributed standard gaussian random variable. Theorem 4.1 then yields the following statement.

Corollary 5.2. *Let $\varepsilon > 0$ be a constant, and let $\lambda < n$ be a parameter that might depend on n . Let the matrix $A_n = (a_{ij})_{1 \leq i, j \leq n} \in n \cdot \Omega_n$ be such that $\varepsilon \leq a_{ij} \leq \lambda$ for every $1 \leq i, j \leq n$. Then with high probability $\log \mathbf{det}(X_n)^2$ is concentrated around $\log n!$. More precisely, there exist constants $c_0 = c_0(\varepsilon)$, $c_1 = c_1(\varepsilon)$ and $c_2 = c_2(\varepsilon)$ such that*

$$\mathbf{P}\left(|\log \mathbf{det}(X_n)^2 - \log n!| \geq c_0 \sqrt{\lambda n} \log^{c_1} n\right) \leq \exp(-c_2 \log^4 n).$$

Note that when $a_{ij} = 1$ (i.e. Ginibre ensemble), Corollary 5.2 (or its stronger form) can be deduced directly from the observation of Goodman [16] that $\mathbf{det}(X_n)^2 = \prod_{i=1}^n d_i^2$ (where d_i is the distance from the i -th row to the subspace generated by the last $n - i$ rows) and that the d_i^2 are independent χ^2 of parameter i thanks to the invariance property of gaussian vectors. However, when the a_{ij} 's are not necessarily the same as in Corollary 5.2, this invariance property totally breaks down.

On the other hand, one can still establish concentration for $\log \mathbf{det}(X_n)^2$ by spectral mean, namely by using the result of Guionnet and Zeitouni [17] on the concentration of linear statistics $\sum_i f(s_i)$, where $s_1 \geq \dots \geq s_n$ are the singular values of X_n . Although in our case $f(x) = \log x$ blows up at zero and infinity, one can still remove the singularity by truncation and by showing that the last few singular values are bounded away from zero with high probability (treatment for the soft edge is more standard). Such an approach can be found for instance in the work of Friedland, Rider and Zeitouni [13]; we also refer the reader to [3, 8, 25] and the references therein. By this concentration phenomenon, and by the fact that $\mathbf{Edet}(X_n)^2 = \mathbf{perm}(A_n)$, one can establish (for a wide range of A_n) that with high probability $\log \mathbf{det}(X_n)^2$ is concentrated around $\log \mathbf{perm}(A_n)$. This was indeed the motivation for the Barvinok-Godsil-Gutman estimator [3, 15]. In this spirit, allow us to cite here a special version of [25, Corollary 1.5] by Ruelle and Zeitouni, a result directly relevant to our simple goal above.

Lemma 5.3. *With A_n as in Corollary 5.2,*

$$\mathbf{P}\left(|\log \mathbf{det}(X_n)^2 - \log \mathbf{perm}(A_n)| \geq c_0 \sqrt{\lambda n} \log^{c_1} n\right) \leq \exp(-c_2 \log^4 n).$$

Corollary 5.2 then follows from this result and Theorem 4.1.

5.4. Perfect matchings in random bipartite graphs in random environment. This subsection is motivated by another paper of Bochi, Iommi and Ponce [5].

Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a Lebesgue probability space, and let $f : \Omega \rightarrow (0, 1]$ be a function with $\log f \in L_\infty(\mathbf{P})$. Suppose that T is an ergodic measure-preserving action on the semi-group \mathbf{N}^2 on $(\Omega, \mathcal{A}, \mathbf{P})$ as in Theorem 1.6, with \mathcal{A}_1 and \mathcal{A}_2 being the sub- σ -algebras formed by the $T^{(1,0)}$ -invariant and the $T^{(0,1)}$ -invariant sets respectively.

For each *environment* ω of the space Ω , for each $n \geq 1$ we define a random bipartite graph $G_n(\omega)$ on the vertex sets $W_n = \{w_1, \dots, w_n\}$ and $M = \{m_1, \dots, m_n\}$ according to the following law $\mathbf{P}_n(\omega)$: for $1 \leq i, j \leq n$, each edge $w_i m_j$ is chosen independently at random with probability

$$a_{ij}(\omega) := f(T^{(i,j)}(\omega)).$$

Note that our random graph is *inhomogeneous* as the a_{ij} can be totally different. It is clear that the number N of perfect matchings in this random bipartite graph is

$$N = \sum_{\pi \in S_n} \mathbf{1}_{w_1 m_{\pi(1)} \text{ is an edge}} \cdots \mathbf{1}_{w_n m_{\pi(n)} \text{ is an edge}}.$$

Thus the expected number of perfect matchings in $G_n(\omega)$ with respect to the law $\mathbf{P}_n(\omega)$ is

$$\begin{aligned} N_n(\omega) &:= \mathbf{E}_{\mathbf{P}_n(\omega)} N = \mathbf{E}_{\mathbf{P}_n(\omega)} \sum_{\pi \in S_n} \mathbf{1}_{w_1 m_{\pi(1)} \text{ is an edge}} \cdots \mathbf{1}_{w_n m_{\pi(n)} \text{ is an edge}} \\ &= \sum_{\pi \in S_n} a_{w_1 m_{\pi(1)}} \cdots a_{w_n m_{\pi(n)}} \\ &= \mathbf{perm}(\square_n(f(\omega))). \end{aligned}$$

By using Theorem 1.6, we obtain the following variant of law of large number for the number of perfect matchings of random bipartite graphs in random environment.

Theorem 5.5. *For \mathcal{P} -almost every environment $\omega \in \Omega$,*

$$\lim_{n \rightarrow \infty} \left(\frac{N_n}{n!} \right)^{1/n} = \mathbf{sm}_{\mathcal{A}_1, \mathcal{A}_2}(f).$$

Acknowledgements. The authors are thankful to S. Leibman, G. Iommi and to the anonymous referee for their very helpful comments.

REFERENCES

- [1] S. Aaronson and H. Nguyen, *Near invariance of the hypercube*, Israel Journal of Mathematics, 212, (2016), 385–417.
- [2] N. Alon and S. Friedland, *The maximum number of perfect matchings in graphs with a given degree sequence*, Electron. J. Combin. 15 (2008), no. 1, Note 13, 2 pp.
- [3] A. Barvinok, *Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor*, Random Structures and Algorithms 14 (1999), 29–61.
- [4] J. Bochi, G. Iommi, and M. Ponce, *The scaling mean and a law of large permanents*, Adv. Math. 292 (2016), 374–409.
- [5] J. Bochi, G. Iommi, and M. Ponce, *Perfect matching in inhomogeneous random bipartite graphs in random environment*, submitted, arxiv.org/pdf/1605.06137v2.pdf.
- [6] L.M. Bregman, *Some properties of non-negative matrices and their permanents*, Soviet Math. Dokl. 14 (1973), 945–949.
- [7] R. Brualdi, S. Parter, and H. Schneider, *The diagonal equivalence of a non-negative matrix to a stochastic matrix*, J. Math. Anal. Appl. 16 (1966), 31–50.
- [8] K. P. Costello and V. Vu, *Concentration of random determinants and permanent estimators*, SIAM J. Discrete Math. 23 (2009), 1356–1371.
- [9] G. Egorychev, *The solution of van der Waerdens problem for permanents*, Adv. in Math. 42 (1981), no. 3, 299–305.
- [10] D. Falikman, *Proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix*, Mat. Zametki 29 (1981), no. 6, 931–938, 957.
- [11] S. Friedland, *Positive diagonal scaling of a nonnegative tensor to one with prescribed slice sums*, Linear Algebra Appl. 434 (7) (2011) 1615–1619.
- [12] S. Friedland and S. Karlin, *Some inequalities for the spectral radius of non-negative matrices and applications*, Duke Math. J. 42 (1975), no. 3, 459–490.
- [13] S. Friedland, B. Rider, and O. Zeitouni, *Concentration of permanent estimators for certain large matrices*, Annals Appl. Probab 14 (2004), 1359–1576.
- [14] S. Friedland, S. Low, and C. Tan, *Nonnegative matrix inequalities and their application to non-convex power control optimization*, SIAM J. Matrix Anal. Appl. 32 (3) (2011), no. 3, 1030–1055.
- [15] C. D. Godsil and I. Gutman, *On the matching polynomial of a graph*, in Algebraic methods in graph theory I-II (L. Lovász and V. T. Sós, eds.), North-Holland, Amsterdam (1981), pp. 67–83.
- [16] N. R. Goodman, *Distribution of the determinant of a complex Wishart distributed matrix*, Annals Stat. 34 (1963), 178–180.

- [17] A. Guionnet and O. Zeitouni, *Concentration of the spectral measure for large matrices*, Elec. Comm. Probab. 5 (2000), 119–136.
- [18] G. Halász and G. Székely, *On the elementary symmetric polynomials of independent random variables*, Acta Math. Acad. Sci. Hungar. 28 (1976), no. 3–4, 397–400.
- [19] G. Keller, *Equilibrium States in Ergodic Theory*, London Math. Soc. Stud. Texts, vol. 42, Cambridge University Press, Cambridge, 1998.
- [20] N. Linial, A. Samorodnitsky, and A. Wigderson, *A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents*, Combinatorica 20 (2000), 545–568.
- [21] A. Marshall and I. Olkin, *Scaling of matrices to achieve specified row and column sums*, Numer. Math. 12 (1968), 83–90.
- [22] M. Menon, *Matrix links, an extremization problem, and the reduction of a non-negative matrix to one with prescribed row and column sums*, Canad. J. Math. 20 (1968), 225–232.
- [23] H. Minc, *Upper bounds for permanents of $(0,1)$ -matrices*, Bull. Amer. Math. Soc. 69 (1963), 789–791.
- [24] G. Rempala and J. Wesolowski, *Symmetric Functionals on Random Matrices and Random Matchings Problems*, Springer, New York, NY, 2008.
- [25] M. Rudelson and O. Zeitouni, *Singular values of Gaussian matrices and permanent estimators*, Random Structures and Algorithms 48 (2016), no. 1, 183–212.
- [26] R. Sinkhorn, *Continuous dependence on A in the D_1AD_2 theorems*, Proc. Amer. Math. Soc. 32 (1972) 395–398.
- [27] R. Sinkhorn and P. Knopp, *Concerning nonnegative matrices and doubly stochastic matrices*, Pacific J. Math. 21 (1967), 343–348.
- [28] G. Soules, *New permanental upper bounds for nonnegative matrices*, Linear Multilinear Algebra 51 (2003), no. 4, 319–337.
- [29] B. van der Waerden, *Aufgabe 45*. Jber. Deutsch. Math. Verein. 35 (1926), 117.
- [30] A. Wigderson, *Matrix and Operator scaling and their many applications*, <http://www.math.ias.edu/avi/talks>.

DEPARTMENT OF MATHEMATICAL SCIENCES, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, ILLINOIS 61801

E-mail address: jobal@math.uiuc.edu

DEPARTMENT OF MATHEMATICS, THE OHIO STATE UNIVERSITY, COLUMBUS, OHIO 43210

E-mail address: nguyen.1261@math.osu.edu