Combinatorics, Probability and Computing (2000) 9, 149–166. Printed in the United Kingdom

The Expected Norm of Random Matrices

YOAV SEGINER

Check Point Software Technologies Ltd., 3A Jabotinsky St.,
Diamond Tower, Ramat Gan, 52520, Israel
(e-mail: yoav@checkpoint.com)

Received 2 March 1998; revised 23 April 1999

We compare the Euclidean operator norm of a random matrix with the Euclidean norm of its rows and columns. In the first part of this paper, we show that if A is a random matrix with i.i.d. zero mean entries, then $E\|A\|^h \leqslant K^h(E\max_i \|a_{i\bullet}\|^h + E\max_j \|a_{\bullet j}\|^h)$, where K is a constant which does not depend on the dimensions or distribution of A (h, however, does depend on the dimensions). In the second part we drop the assumption that the entries of A are i.i.d. We therefore consider the Euclidean operator norm of a random matrix, A, obtained from a (non-random) matrix by randomizing the signs of the matrix's entries. We show that in this case, the best inequality possible (up to a multiplicative constant) is $E\|A\|^h \leqslant (c\log^{1/4}\min\{m,n\})^h(E\max_i \|a_{i\bullet}\|^h + E\max_j \|a_{\bullet j}\|^h)$ (m,n the dimensions of the matrix and c a constant independent of m,n).

1. Introduction

We consider here the Euclidean operator norm (denoted by $\|\cdot\|$) of a matrix whose entries are independent random variables. This norm appears, for example, in different applications of linear algebra to data collected from measurements where the random matrix is the measurement error matrix and its norm is used to determine the significance of these errors (see the introduction to [5] and the papers cited there). In many other applications, it is rather $\|A\|^2 = \|AA^t\|$ and not the norm of the matrix which is of primary interest. Such is the case in multivariate statistics where AA^t is the covariance matrix and its norm is used in principle component analysis. Therefore, in this paper, we deal with $\|A\|^h$ where h is at least 1. This question and other similar questions have been previously investigated by several authors. In 1980, Geman [4] proved the following limit theorem for the norm of random matrices.

Theorem. *Suppose $\{v_{ij}\}$ $i = 1,2,\ldots, j = 1,2,\ldots$ are i.i.d. random variables with $Ev_{11} = 0$ and $E|v_{11}|^n \leqslant n^{\alpha n}$ for all $n \geqslant 2$ and some α. Let $\sigma^2 = Ev_{11}^2$ and $V_{pn} = \{v_{ij}\}_{1\leqslant i\leqslant p;1\leqslant j\leqslant n}$. If p_n is a sequence of integers such that $\lim_{n\to\infty}\frac{p_n}{n} = y$ for some $0 < y < \infty$ then $\lim_{n\to\infty}\frac{1}{n}\|V_{p_n n}V_{p_n n}^t\| = (1+y^{1/2})^2\sigma^2$ a.s.*

(For later versions of this theorem with slightly more relaxed conditions, see [8, p. 165]).

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

While in this paper we deal only with matrices with independent entries, it is worth mentioning that similar results were obtained by Füredi and Komlós [3] for symmetric random matrices (the entries on the diagonal and above it being independent random variables and the entries below the diagonal defined so as to make the matrix symmetric). Here the independent random variables were not required to be identically distributed, but only to be bounded from above by a common bound, have common expectation and variance for the off-diagonal entries and common expectation for the entries on the diagonal.

A completely different type of theorem related to this problem is Chevet's inequality ([1], [2] and [6, Theorem 3.20]), which concerns tensor products of Gaussian measures. As a special case of this theorem we get that, if $\{g_i\}_{1\leqslant i}$ is a sequence of i.i.d. $N(0,1)$ random variables and $G=(g_{ij})_{1\leqslant i\leqslant m;1\leqslant j\leqslant n}$ a matrix of i.i.d. $N(0,1)$ random variables, then

$$\max\{E\|(g_1,\dots,g_n)\|, E\|(g_1,\dots,g_m)\|\} \leqslant E\|G\| \leqslant E\|(g_1,\dots,g_n)\| + E\|(g_1,\dots,g_m)\|,$$

where all the norms are Euclidean.

Chevet's inequality and the limit theorems cited above are similar in that they both state that the Euclidean operator norm of a matrix is, on average, not much larger than the Euclidean norm of a row or column in the matrix (the larger of the two, depending on the dimensions of the matrix). In view of this, we may be tempted to hypothesize that an inequality of the form $E\|A\| \leqslant K(E\|a_{1\bullet}\| + E\|a_{\bullet 1}\|)$ should hold for every random matrix A with i.i.d., zero mean, random entries ($a_{1\bullet}$ and $a_{\bullet 1}$ being the first row and column of the matrix) with a constant K which does not depend on the dimensions or distribution of A. The following example shows that such an inequality cannot hold.

Example. Let $A_n = (a_{ij}^{(n)})_{1\leqslant i,j\leqslant n}$ be a sequence of random matrices with i.i.d. random entries such that $P\{a_{ij}^{(n)}=+1\} = P\{a_{ij}^{(n)}=-1\} = \frac{1}{2n}$ and $P\{a_{ij}^{(n)}=0\} = 1-\frac{1}{n}$. For every n, $E\|a_{\bullet j}^{(n)}\| = E\|a_{i\bullet}^{(n)}\| \leqslant \sqrt{E\|a_{i\bullet}^{(n)}\|^2} = 1$ but $E\|A_n\| \geqslant E\max_{1\leqslant i\leqslant n}\|a_{i\bullet}^{(n)}\|$ and, as $\|a_{i\bullet}^{(n)}\|^2$ are approximately Poisson(1), $E\|A_n\|$ is at least of order $\sqrt{\log n/\log\log n}$. Therefore, an inequality of the form $E\|A\| \leqslant K(E\|a_{1\bullet}\| + E\|a_{\bullet 1}\|)$ cannot hold.

In this example, the desired inequality could not hold because of the failure of the inequality $E\max_i\|a_{i\bullet}\| \leqslant C\cdot E\|a_{1\bullet}\|$ to hold with a constant C independent of the distribution and dimensions of the matrix A (as implied by the limit theorems, such an inequality does hold for a sufficiently large matrix, the size of which depends on the distribution of the matrix entries). It turns out that the inequality that does hold with a common constant for all random matrices A with i.i.d. zero mean entries is the inequality $E\|A\|^h \leqslant K^h(E\max_i\|a_{i\bullet}\|^h + E\max_j\|a_{\bullet j}\|^h)$. This is the main theorem of the first part of this paper. We will prove the theorem under the condition that the matrix entries are symmetric random variables (*i.e.*, X and $-X$ have the same distribution). The zero mean case is then an easy corollary of the theorem (Corollary 2.2 below).

Theorem 1.1. *There exists a constant K such that, for any n,m any $h\leqslant 2\log\max\{m,n\}$ and any $m\times n$ random matrix $A=(a_{ij})$ where a_{ij} are i.i.d. symmetric random variables, the*

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

following inequality holds:

$$\max\left\{E\max_{1\leqslant i\leqslant m}\|a_{i\bullet}\|^h, E\max_{1\leqslant j\leqslant n}\|a_{\bullet j}\|^h\right\}\leqslant E\|A\|^h$$

$$\leqslant K\left(E\max_{1\leqslant i\leqslant m}\|a_{i\bullet}\|^h+E\max_{1\leqslant i\leqslant n}\|a_{\bullet j}\|^h\right).$$

If any of the expectations appearing in the inequality is ∞, the inequality holds in the sense that the other side of the inequality is also ∞.

There are two factors which make the expected matrix norm small in Theorem 1.1: the first is the symmetry of the random variables and the second is the independence and identical distribution of the entries. Clearly, dropping the assumption that the entries are symmetric (or zero mean) random variables brings us back to the worst possible case for constant matrices (the matrix of identical constant entries). The symmetry of the random variables, however, is sufficient to ensure a smaller ratio between the expected operator norm of the matrix and the expectation of the maximum row or column norm, but this ratio is not as small as the ratio in Theorem 1.1. In the second part of this paper we show that if A is a random matrix obtained from a (non-random) matrix by randomizing the signs of the matrix entries, then the best inequality possible (up to a multiplicative constant) is $E\|A\|^h\leqslant(c\log^{1/4}\min\{m,n\})^h(E\max_i\|a_{i\bullet}\|^h+E\max_j\|a_{\bullet j}\|^h)$ (m,n the dimensions of the matrix, c a constant independent of m,n and h as in Theorem 1.1).

2. The expected norm of a matrix with i.i.d. entries

Our main aim in this section is to prove Theorem 1.1. In addition to the Euclidean norm (denoted by $\|\cdot\|$) we define the norm

$$|||A|||=\max_{\substack{1\leqslant i\leqslant m\\1\leqslant j\leqslant n}}\{\|a_{i\bullet}\|,\|a_{\bullet j}\|\}$$

for an $m\times n$ matrix A, where $a_{i\bullet}$ is the ith row of the matrix and $a_{\bullet j}$ its jth column.

To prove Theorem 1.1 we will need the following theorem.

Theorem 2.1. *There exists a constant K such that, for any m, n, for any $h\leqslant 2\log\max\{m,n\}$ and for any $m\times n$ matrix A,*

$$E_{\sigma,\varepsilon}\left\|\frac{S_{\sigma,\varepsilon}(A)}{|||S_{\sigma,\varepsilon}(A)|||}\right\|^h\leqslant K^h,$$

where $\sigma=(\sigma_1,\ldots,\sigma_m)$ is a vector of independent random permutations, each uniformly distributed over S_n, $\varepsilon=(\varepsilon_{ij}),1\leqslant i\leqslant m,1\leqslant j\leqslant n$ a matrix of i.i.d. random variables such that for every i,j $P\{\varepsilon_{ij}=+1\}=P\{\varepsilon_{ij}=-1\}=\frac{1}{2}$ and $S_{\sigma,\varepsilon}(A)=(\varepsilon_{ij}\cdot a_{i,\sigma_i(j)})_{i,j}$.

Given Theorem 2.1, the proof of Theorem 1.1 is simple.

Proof of Theorem 1.1. Similarly to $S_{\sigma,\varepsilon}(A)$, define $S_{\sigma^t,\varepsilon}(A)=(\varepsilon_{ij}\cdot a_{\sigma_j(i),j})_{i,j}$ where each σ_j is uniformly distributed over S_m.

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

Because the a_{ij} are i.i.d. symmetric random variables, $A, S_{\sigma,\varepsilon}(A)$ and $S_{\sigma^t,\varepsilon}(A)$ have the same distribution. Denoting the rows of $S_{\sigma,\varepsilon}(A)$ by $\tilde{a}_{i\bullet}$ and the columns of $S_{\sigma^t,\varepsilon}(A)$ by $\tilde{a}_{\bullet j}$ we have

$$E\|A\|^h = E|||A|||^h \left\| \frac{A}{|||A|||} \right\|^h \leqslant E \max_{1\leqslant i\leqslant m} \|a_{i\bullet}\|^h \left\| \frac{A}{|||A|||} \right\|^h + E \max_{1\leqslant j\leqslant n} \|a_{\bullet j}\|^h \left\| \frac{A}{|||A|||} \right\|^h$$

$$= E \max_{1\leqslant i\leqslant m} \|\tilde{a}_{i\bullet}\|^h \left\| \frac{S_{\sigma,\varepsilon}(A)}{|||S_{\sigma,\varepsilon}(A)|||} \right\|^h + E \max_{1\leqslant j\leqslant n} \|\tilde{a}_{\bullet j}\|^h \left\| \frac{S_{\sigma^t,\varepsilon}(A)}{|||S_{\sigma^t,\varepsilon}(A)|||} \right\|^h .$$

Because sign changes and permutations of vector entries do not change the Euclidean norm of a vector, max $\|\tilde{a}_{i\bullet}\|^h$ remains constant and equal to max $\|a_{i\bullet}\|^h$ while averaging over σ and ε (and similarly for max $\|\tilde{a}_{\bullet j}\|^h$). Therefore we get

$$E\|A\|^h \leqslant E_A \max_{1\leqslant i\leqslant m} \|a_{i\bullet}\|^h E_\sigma E_\varepsilon \left\| \frac{S_{\sigma,\varepsilon}(A)}{|||S_{\sigma,\varepsilon}(A)|||} \right\|^h$$

$$+ E_A \max_{1\leqslant j\leqslant n} \|a_{\bullet j}\|^h E_{\sigma^t} E_\varepsilon \left\| \frac{S_{\sigma^t,\varepsilon}(A)}{|||S_{\sigma^t,\varepsilon}(A)|||} \right\|^h$$

$$\leqslant K \cdot E \max_{1\leqslant i\leqslant m} \|a_{i\bullet}\|^h + K \cdot E \max_{1\leqslant j\leqslant n} \|a_{\bullet j}\|^h$$

Theorem 2.1 is used to justify the last inequality.

The left inequality of the theorem is trivial. □

In Theorem 1.1 it is possible to replace the condition that all entries are symmetric random variables by the condition that all entries are zero mean random variables. This follows as a corollary from Theorem 1.1.

Corollary 2.2. *There exists a constant K such that, for any n, m for any $h \leqslant 2 \log \max \{m, n\}$ and any $m \times n$ random matrix $A = (a_{ij})$, where a_{ij} are i.i.d. zero mean random variables, the following inequality holds:*

$$\max \left\{ E \max_{1\leqslant i\leqslant m} \|a_{i\bullet}\|^h, E \max_{1\leqslant j\leqslant n} \|a_{\bullet j}\|^h \right\} \leqslant E\|A\|^h$$

$$\leqslant (2K)^h \left(E \max_{1\leqslant i\leqslant m} \|a_{i\bullet}\|^h + E \max_{1\leqslant i\leqslant n} \|a_{\bullet j}\|^h \right).$$

Proof. Let $A = (a_{ij})$ be a matrix as in the conditions of the corollary and let $B = (b_{ij})$ be an i.i.d. copy of A. Let $\varepsilon = (\varepsilon_{ij}), 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n$ be a matrix of i.i.d. random variable such that, for every i, j, $P\{\varepsilon_{ij} = +1\} = P\{\varepsilon_{ij} = -1\} = \frac{1}{2}$.

As A and B are i.i.d., $\varepsilon \circ (A - B)$ and $A - B$ have the same distribution ($\circ$ being the Hadamard, *i.e.*, entry by entry, product for matrices). Therefore, by convexity, $2^h E\|\varepsilon \circ A\|^h = 2^{h-1} E\|\varepsilon \circ A\|^h + 2^{h-1} E\|\varepsilon \circ B\|^h \geqslant E\|\varepsilon \circ (A - B)\|^h = E\|A - B\|^h$. In addition, using $Eb_{ij} = 0$

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

in the last equality:

$$\begin{aligned} E_B\|A-B\| &= E_B \sup_{\{x\in\Re^m, y\in\Re^n:\|x\|=1,\|y\|=1\}} \left(\sum_{i,j} a_{ij}\, x_i\, y_j - \sum_{i,j} b_{ij}\, x_i\, y_j\right) \\ &\geqslant \sup_{\{x\in\Re^m, y\in\Re^n:\|x\|=1,\|y\|=1\}} E_B \left(\sum_{i,j} a_{ij}\, x_i\, y_j - \sum_{i,j} b_{ij}\, x_i\, y_j\right) = \|A\|. \end{aligned}$$

Putting the last two inequalities together we get $E\|A\|^h \leqslant E_A(E_B\|A-B\|)^h \leqslant E_A E_B\|A-B\|^h \leqslant 2^h E\|\varepsilon \circ A\|^h$.

The corollary now follows by applying Theorem 1.1 to the matrix $\varepsilon \circ A$. □

We now prove Theorem 2.1 (the lemmas used in this proof will be proved at the end of this section).

Proof of Theorem 2.1. Define $S_{\sigma,\varepsilon,\mu}(A) = (\varepsilon_{ij} \cdot a_{\mu(i),\sigma_i(j)})_{i,j}$ where σ and ε are as in the definition of $S_{\sigma,\varepsilon}(A)$ and μ is a random permutation uniformly distributed over S_m. As changing the order of the rows does not affect the Euclidean norm or the $|||\cdot|||$ norm of a matrix, it is equivalent to show that

$$E_{\sigma,\varepsilon,\mu}\left\|\frac{S_{\sigma,\varepsilon,\mu}(A)}{|||\, S_{\sigma,\varepsilon,\mu}(A)\, |||}\right\|^h \leqslant K^h.$$

In addition, as we can freely transpose the matrix, it is enough to prove the theorem for $h \leqslant 2\log n$.

Let $k = \lceil \log n \rceil$ and denote

$$B = (b_{ij})_{i,j} = B(A) = \frac{S_{\sigma,\varepsilon,\mu}(A)}{|||\, S_{\sigma,\varepsilon,\mu}(A)\, |||} \text{ and } P = (p_{ij}) = (b_{ij}^2).$$

Notice that P is a (random) doubly sub-stochastic matrix (*i.e.*, all entries are nonnegative and the sum of entries in any row or column is never larger than 1). Then,

$$\begin{aligned} E\left\|\frac{S_{\sigma,\varepsilon,\mu}(A)}{|||\, S_{\sigma,\varepsilon,\mu}(A)\, |||}\right\|^h &= E\|B\|^h \leqslant (E\|B\|^{2k})^{h/2k} \leqslant (E\operatorname{tr}(B^tB)^k)^{h/2k} \\ &= \left(\sum_{1\leqslant i_1,\dots,i_k\leqslant m}\ \sum_{1\leqslant j_1,\dots,j_k\leqslant n} E_\mu E_\sigma E_\varepsilon b_{i_1j_1}\, b_{i_1j_2}\, b_{i_2j_2}\dots b_{i_kj_k}\, b_{i_kj_1}\right)^{h/2k}. \qquad (2.1) \end{aligned}$$

Because sign changes of the matrix entries do not change the norm $|||\cdot|||$ of a matrix, we have, with $S_\sigma(A) = (a_{i,\sigma_i(j)})_{i,j}$,

$$b_{ij} = \varepsilon_{ij}\frac{a_{\mu(i),\sigma_i(j)}}{|||\, S_\sigma(A)\, |||}$$

and therefore $E_\varepsilon b_{i_1j_1}\, b_{i_1j_2}\, b_{i_2j_2}\dots b_{i_kj_k}\, b_{i_kj_1} = 0$ whenever there exists an entry b_{ij} which appears in the product an odd number of times. It suffices therefore to deal with products in which every b_{ij} appears an even number of times.

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

Let $G(U, V)$ be the complete bipartite graph over the vertex sets U and V where $|U| = m$ and $|V| = n$. We associate every edge $e_{ij} = (u_i, v_j)(u_i \in U, v_j \in V)$ in the graph G with the i, jth entry in the matrix B and assign to it the weight $p(e_{ij}) = p_{ij}$ according to the matrix P defined above.

Every product $b_{i_1j_1}\, b_{i_1j_2}\, b_{i_2j_2} \ldots b_{i_kj_k}\, b_{i_kj_1}$ is thus associated with a walk on the graph G that begins and ends at v_{j_1}.

Let $\gamma = (x_1(\gamma), \ldots, x_{2k}(\gamma))$ be a walk of length $2k$ on the graph G (where $x_1(\gamma), \ldots, x_{2k}(\gamma)$ are the edges and $v_0(\gamma), v_1(\gamma), \ldots, v_{2k}(\gamma)$ the vertices of the walk). In addition, let $\gamma(i) = (x_1(\gamma), \ldots, x_i(\gamma))$ be the prefix of length i of the walk γ.

As we are interested only in those walks which pass through every edge an even number of times, we can, instead of calculating products of b_{ij}s, calculate products of p_{ij}s where the weight p_{ij} appears in the product only every second pass of the walk through the edge e_{ij}. Therefore we shall distinguish between 'odd' and 'even' passes of a walk through an edge. We say that $x_i(\gamma)$ is an odd step in γ if the edge traversed by this step has been traversed in total an odd number of times in $\gamma(i)$. We similarly define an even step. Taking the product of the weights of the odd steps

$$\pi_i(\gamma) = \prod_{\{1 \leqslant j \leqslant i\, :\, x_j(\gamma) \text{ is an odd step}\}} p(x_j(\gamma))$$

we have that, for the walks which interest us, $Eb_{i_1j_1}\, b_{i_1j_2}\, b_{i_2j_2} \ldots b_{i_kj_k}\, b_{i_kj_1} = E\pi_{2k}(\gamma)$.

Step 1. For every walk γ, we define the odd/even step sequence of the walk as $T(\gamma) = (t_1(\gamma), \ldots, t_{2k}(\gamma))$ where $t_i(\gamma) = +1$ if $x_i(\gamma)$ is an odd step and $t_i(\gamma) = -1$ otherwise. Clearly $\sum_{j=1}^{i} t_j(\gamma) \geqslant 0$ for every i and $\sum_{j=1}^{2k} t_j(\gamma) = 0$. Denote the set of all ± 1 sequences of length $2k$ for which the first condition holds by $\mathbf{T}$ and the subset of sequences for which both conditions hold by $\mathbf{T}_0$. We begin by fixing a starting point $(v_0 \in V)$ and an odd/even step sequence $(T \in \mathbf{T}_0)$ for the walk. Denote all the walks with the given starting point and odd/even step sequence by $\Gamma_{v_0}(T) = \{\gamma : v_0(\gamma) = v_0, T(\gamma) = T\}$. As there are n vertices to choose the starting point from and it is well known (see, for example, [7], exercise 1.33) that the number of elements in $\mathbf{T}_0$ is

$$\frac{1}{k+1}\binom{2k}{k},$$

the number of different subsets $\Gamma_{v_0}(T)$ is at most

$$\frac{n}{k+1}\binom{2k}{k} \leqslant n\, 2^{2k} k^{-1}.$$

Step 2. We say that an edge is in an 'odd state' ('even state') in $\gamma(i)$ if the last step in $\gamma(i)$ which traversed it was an odd (even) step. Let $O_i(\gamma)$ be the graph whose edges are the edges in an 'odd state' in $\gamma(i)$. Denote by $M_i(\gamma)$ the degree (maximal number of edges incident with a single vertex) of the graph $O_i(\gamma)$. For $i_1 \leqslant i_2$ denote $M_{i_1,i_2}(\gamma) = \max_{i_1 \leqslant j \leqslant i_2} M_i(\gamma)$.

Having fixed $v_0 \in V$ and $T \in \mathbf{T}_0$ we now partition the set of walks γ in $\Gamma_{v_0}(T)$ into subsets according to the maximal value of $M_i(\gamma)$ attained along the walk (*i.e.*, $M_{1,2k}(\gamma)$) and the first step at which this value is attained. As $1 \leqslant M_{1,2k}(\gamma) \leqslant k$ and this value is first attained on an odd step, there are at most k^2 subsets in this partition.

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

Step 3. We now fix a set in the above partition (*i.e.*, all the walks in $\Gamma_{v_0}(T)$ for which $M_{1,2k}(\gamma) = \lambda$ and the maximum is first attained on the ith step). Instead of directly finding an upper bound for the sum of products $\pi_{2k}(\gamma)$ for these walks we shall find an upper bound for products of the form $\Pi_{x \in O_i(\gamma)}\, p(x)$. These products have the advantage that every edge appears in them at most once. The following lemma allows us to make this transition.

Lemma 2.3. *For any* $1 \leqslant \lambda$ *and* $1 \leqslant i \leqslant r$,

$$\sum_{\{\gamma(r):\gamma\in\Gamma_{v_0}(T), M_{1,i-1}(\gamma)<\lambda, M_i(\gamma)=\lambda, M_{i+1,r}(\gamma)\leqslant\lambda\}} \pi_r(\gamma) \leqslant \lambda^{\lambda_i(T)+\mu_{i+1,r}(T)} \sum_{H\in H_{i,\lambda}(T,v_0)} \prod_{x\in H} p(x),$$

where $\lambda_i(T)$ *is the number of odd steps in* $\gamma(i), \mu_{i_1,i_2}(T)$ *the number of even steps from step* i_1 *to step* i_2 *(inclusive) and* $H_{i,\lambda}(T, v_0) = \{O_i(\gamma) : \gamma \in \Gamma_{v_0}(T), M_{1,i-1}(\gamma) < \lambda, M_i(\gamma) = \lambda\}$ *(the 'odd edge graphs' of the given walks).*

Setting $r = 2k$, this lemma gives an upper bound of

$$\sum_{\{\gamma\in\Gamma_{v_0}(T):M_{1,2k}(\gamma)=\lambda,\ \text{the maximum is first attained at step } i\}} \pi_{2k}(\gamma) \leqslant \lambda^{2k} \sum_{H\in H_{i,\lambda}(T,v_0)} \prod_{x\in H} p(x)$$

for the walks in the subset we have been considering.

Step 4. The upper bound in the previous step contains a summation over the set $H_{i,\lambda}$. The next lemma gives an upper bound for the number of graphs in this set (in fact, a somewhat larger set: $\{O_i(\gamma) : M_i(\gamma) = \lambda\}$).

Lemma 2.4. *For every* $v_0 \in V, T \in \mathbf{T}, 1 \leqslant \lambda, d \leqslant r \leqslant n$,

$$|\{H \in H_\lambda(T, v_0) : |V(H) \cap U| = d, |H| = r\}| \leqslant 2^{\lfloor \frac{r}{2} \rfloor} e^d d^{-1/2} m^d n^{r-\lceil \frac{\lambda-2}{2} \rceil - d+1},$$

where $H_\lambda(T, v_0) = \{O_i(\gamma) : \gamma \in \Gamma_{v_0}(T), M_i(\gamma) = \lambda, 1 \leqslant i \leqslant 2k\}$ *and* $V(H)$ *is the set of vertices incident with at least one edge in* H.

Step 5. After we have found a bound for the number of graphs in the sum, it remains to bound the expressions $E\Pi_{x\in H}\, p(x)$. Here for the first time we make use of the row and column permutations. The next lemma gives the desired upper bound.

Lemma 2.5. *Let* H *be a subgraph of* $G(U, V)$ *and let* r *be the number of edges in* H *and* d *the number of vertices in* U *incident with at least one edge in* H; *then,*

$$E_\mu E_\sigma \prod_{x\in H} p(x) \leqslant \left(\frac{1}{n}\right)^r \min\left\{1, \left(\frac{n}{m}\right)^d\right\}.$$

We now substitute the bounds from Steps 4 and 5 into the inequality of Step 3. Notice that while the bound in Step 3 increases with λ, the bound in Step 4 decreases with λ. Comparing the corresponding elements in the bounds, a simple calculation shows that

$$\lambda^{2k} n^{-\lceil \frac{\lambda-2}{2} \rceil} \leqslant k^2 (4e^{-1/2})^{2k}.$$

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

Using this last inequality and the fact that r and d above are at most k, we get

$$\sum_{\{\gamma\in\Gamma_{v_0}(T)\,:\,M_{1,2k}(\gamma)=\lambda,\text{ the maximum is first attained at step } i\}} \pi_{2k}(\gamma) \leqslant k^2 2^{k/2} e^k n k^2 (4e^{-1/2})^{2k}.$$

This bound is true for any of the $n\,2^{2k}\,k$ subsets of walks produced by the first two steps of the proof. Therefore we conclude that

$$E\|B\|^{2k} \leqslant n\,2^{2k} k k^2 2^{k/2} e^k n k^2 (4e^{-1/2})^{2k} \leqslant k^5 (2\cdot 2^{1/4} 4)^{2k} n^2 \leqslant (8\cdot 2^{1/4} e^2)^{2k}.$$

For any $h \leqslant 2k$ we get the desired inequality: $E\|B\|^h \leqslant (8\cdot 2^{1/4}e^2)^h$. □

Proofs of the lemmas

Proof of Lemma 2.3. The proof is by induction on the steps of the walk. This is divided into two parts: the steps up to the step at which the maximal $M_i(\gamma)$ is first attained, and the steps from this point on.

For the steps of the first group we prove the following claim.

Claim. For any $v_0 \in U\cup V$, $T \in \mathbf{T}, v \in U\cup V, H \subseteq E(G), i \geqslant 1, \lambda \geqslant 1$,

$$\sum_{\{\gamma(i)\,:\,\gamma\in\Gamma_{v_0}(T), O_i(\gamma)=H, v_i(\gamma)=v, M_{1,i}(\gamma)\leqslant\lambda\}} \pi_i(\gamma) \leqslant \lambda^{\lambda_i(T)-1} \prod_{x\in H} p(x) \tag{2.2}$$

($\Pi_{x\in H}\ p(x)$ is defined to be 1 if H is empty).

Proof. Without loss of generality we assume $v_0 \in V$. The proof is by induction on i. It is easy to verify that the lemma holds for $i = 1$. We now assume the lemma holds for $i-1$ and prove it for i.

Without loss of generality we assume that $v \in V$. Also, in what follows all walks are in $\Gamma_{v_0}(T)$.

If $t_i = -1$, $x_i(\gamma)$ is an even step and therefore $\pi_{i-1}(\gamma) = \pi_i(\gamma)$ and $\lambda_i(T) = \lambda_{i-1}(T)$. Therefore, by the induction hypothesis and because $\sum_{\{y=(u,v):u\in U, y\notin H\}} p(y) \leqslant 1$, we get

$$\begin{aligned}
&\sum_{\{\gamma(i)\,:\,O_i(\gamma)=H, v_i(\gamma)=v, M_{1,i}(\gamma)\leqslant\lambda\}} \pi_i(\gamma)\\
&\quad= \sum_{\{y=(u,v)\,:\,u\in U, y\notin H\}} \ \sum_{\{\gamma(i-1)\,:\,O_{i-1}(\gamma)=H\cup\{y\}, v_{i-1}(\gamma)=u, M_{1,i-1}(\gamma)\leqslant\lambda\}} \pi_{i-1}(\gamma)\\
&\quad\leqslant \sum_{\{y=(u,v)\,:\,u\in U, y\notin H\}} \lambda^{\lambda_{i-1}(T)-1} p(y) \prod_{x\in H} p(x) = \lambda^{\lambda_i(T)-1} \prod_{x\in H} p(x)\\
&\qquad\times \sum_{\{y=(u,v)\,:\,u\in U, y\notin H\}} p(y) \leqslant \lambda^{\lambda_i(T)-1} \prod_{x\in H} p(x).
\end{aligned}$$

If $t_i = +1$, denote by $E(H, v)$ the set of edges in H incident with v. We may assume that $|E(H, v)| \leqslant \lambda$. Therefore, using the induction hypothesis, we have

$$\sum_{\{\gamma(i)\,:\,O_i(\gamma)=H, v_i(\gamma)=v, M_{1,i}(\gamma)\leqslant\lambda\}} \pi_i(\gamma)$$

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

$$= \sum_{(u,v)=y\in E(H,v)} \sum_{\{\gamma(i-1)\,:\,O_{i-1}(\gamma)=H\setminus\{y\},v_{i-1}(\gamma)=u,M_{1,i-1}(\gamma)\leqslant\lambda\}} \pi_{i-1}(\gamma)p(y)$$

$$\leqslant \sum_{(u,v)=y\in E(H,v)} \lambda^{\lambda_{i-1}(T)-1}p(y) \prod_{x\in H\setminus\{y\}} p(x) = |E(H,v)|\lambda^{\lambda_{i-1}(T)-1}$$

$$\times \prod_{x\in H} p(x) \leqslant \lambda^{\lambda_i(T)-1} \prod_{x\in H} p(x). \qquad \square$$

We now use the above claim to prove the lemma for $i = r$.

We have to sum inequality (2.2) over all $H \in H_{i,\lambda}(T, v_0)$ and over all vertices

$$V_{i,\lambda}(H) = \{v \,:\, \exists\gamma \in \Gamma_{v_0}(T)H = O_i(\gamma), v_i(\gamma) = v, M_{1,i-1}(\gamma) < \lambda, M_i(\gamma) = \lambda\},$$

which are the vertices that may be reached on the ith step by a walk which generates H in $H_{i,\lambda}(T, v_0)$. Because $M_{1,i-1}(\gamma) < \lambda$ but $M_i(\gamma) = \lambda$, $O_i(\gamma)$ contains at most two vertices incident with λ edges δ (these are the two vertices incident with the last edge added). As i determines the side of the graph which the walk reaches in the ith step, given i and $O_i(\gamma)$ there are at most λ vertices in $V_{i,\lambda}(H)$. Therefore, summing over inequality (2.2),

$$\sum_{\{\gamma(i)\,:\,\gamma\in\Gamma_{v_0}(T),M_{1,i-1}(\gamma)<\lambda,M_i(\gamma)=\lambda\}} \pi_i(\gamma)$$

$$= \sum_{H\in H_{i,\lambda}(T,v_0)} \sum_{v\in V_{i,\lambda(H)}} \sum_{\{\gamma(i)\,:\,\gamma\in\Gamma_{v_0}(T),O_i(\gamma)=H,v_i(\gamma)=v,M_{1,i}(\gamma)\leqslant\lambda\}} \pi_i(\gamma)$$

$$\leqslant \sum_{H\in H_{i,\lambda}(T,v_0)} |V_{i,\lambda}(H)| \cdot \lambda^{\lambda_i(T)-1} \prod_{x\in H} p(x) \leqslant \sum_{H\in H_{i,\lambda}(T,v_0)} \lambda^{\lambda_i(T)} \prod_{x\in H} p(x),$$

which proves the lemma for $i = r$.

To complete the proof of the lemma, we proceed by induction on r. The proof is very similar to the proof of the claim above, only now it is the even steps that add an extra λ factor to the bound. We omit the details of the calculation. □

The next two lemmas characterize the structure of the graphs in the sets $H_{i,\lambda}$ and enable us to calculate an upper bound (in Lemma 2.4) for the number of graphs in these sets.

Lemma 2.6. *For any walk γ on the graph G and for any i, $O_i(\gamma) = \bigcup_{j=1}^{h} C_j \cup L$ where*
(a) *$C_1, \ldots, C_h, L$ are edgewise disjoint,*
(b) *every C_j is a cycle,*
(c) *L is a path* (*perhaps a cycle*) *which begins at $v_0(\gamma)$ and ends at $v_i(\gamma)$.*

Proof. First, if L is a trail (possibly closed) containing cycles, we can write L as an edgewise disjoint union $L = \bigcup \tilde{C}_j \cup \tilde{L}$ where $\tilde{C}_j$ are cycles and $\tilde{L}$ is a path (possibly a cycle) which begins and ends at the same vertices as L.

It is therefore sufficient to prove the lemma without the requirement that L be a path.

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

For $i = 1$ the lemma is trivial. We now proceed by induction from $i-1$ to i: $O_{i-1}(\gamma) = \bigcup_{j=1}^{h} C_j \cup L$.

We write $L = (v_0, \dots, v_r)$, where $v_0 = v_0(\gamma)$ and $v_r = v_{i-1}(\gamma)$. If $t_i = +1$, $x_i(\gamma) = (v_r, v_i(\gamma)) \notin O_{i-1}(\gamma)$ and therefore $\tilde{L} = (v_0, \dots, v_r, v_i(\gamma))$ is a trail that begins at $v_0(\gamma)$ and ends at $v_i(\gamma)$. Obviously, $O_i(\gamma) = \bigcup_{j=1}^{h} C_j \cup L$ is an edgewise disjoint union.

If $t_i = -1$ the edge being removed may be either part of the path L or of one of the cycles. If $x_i(\gamma) \in L$ then what remains of L after the edge is removed is the desired path (the cycles remain unchanged). If $x_i(\gamma) \in C_{j_0}$ then $C_{j_0} = (\tilde{v}_0, \dots, \tilde{v}_s)$ with $\tilde{v}_0 = \tilde{v}_s = v_{i-1}(\gamma) = v_r$ and $\tilde{v}_{s-1} = v_i(\gamma)$. We can then take $\tilde{L} = (v_0, \dots, v_{r-1}, \tilde{v}_0, \dots, \tilde{v}_{s-1})$ and

$$O_i(\gamma) = \bigcup_{\substack{j=1 \\ j \neq j_0}}^{h} C_j \cup \tilde{L}$$

is the required decomposition. □

Lemma 2.7. *For any walk γ on the graph G and for any i, if $\lambda = \deg_{O_i(\gamma)}(v)$ then $O_i(\gamma)$ contains at least $\lceil \frac{\lambda-2}{2} \rceil$ cycles that pass through v.*

Proof. There exist λ edges $e_1, \dots, e_\lambda \in O_i(\gamma)$ incident with the vertex v. By Lemma 4, $O_i(\gamma) = \bigcup_{j=1}^{h} C_j \cup L$. The lemma follows directly from this decomposition and the observation that each C_j contains 0 or 2 of these edges while L contains at most 1 or 2 of these edges (depending on the total number of edges). □

Proof of Lemma 2.4. As $M(H) = \lambda$, there exists a vertex v such that $\deg_H(v) = \lambda$. By Lemma 2.6 and Lemma 2.7,

$$H = \bigcup_{j=1}^{\lceil \frac{\lambda-2}{2} \rceil} C_j \bigcup_{j=1}^{h} \tilde{C}_j \cup L$$

where C_j are cycles passing through v, $\tilde{C}_j$ are cycles and L is a path, beginning at v_0. Let $s_j, \tilde{s}_j, s$ be the number of edges in $C_j, \tilde{C}_j, L$ (respectively).

(a) Because

$$\sum_{j=1}^{\lceil \frac{\lambda-2}{2} \rceil} s_j + \sum_{j=1}^{h} \tilde{s}_j + s = |H| = r,$$

and because the s_js and $\tilde{s}_j$s are even numbers (as the lengths of cycles on a bipartite graph), there are at most $s^{\lfloor \frac{r}{2} \rfloor}$ ways to choose h and $s_1, \dots, s_{\lceil \frac{\lambda-2}{2} \rceil}, \tilde{s}_1, \dots, \tilde{s}_h, s$ (the partial sums $s_1, s_1 + s_2, \dots,$ $s_1, \dots, s_{\lceil \frac{\lambda-2}{2} \rceil}, \tilde{s}_1 + \dots + \tilde{s}_h$, which uniquely determine the sequence, are a subset of the $\lfloor \frac{r}{2} \rfloor$ positive even numbers which are no larger than r).

(b) There are $\binom{m}{d}$ ways to choose the d vertices in U that are incident with the graph H.

(c) Given h, $s_1, \dots, s_{\lceil \frac{\lambda-2}{2} \rceil}, \tilde{s}_1, \dots, \tilde{s}_h, s$ and the d vertices in $V(H) \cap U$, there are at most $2d^d n^{r - \lceil \frac{\lambda-2}{2} \rceil - d + 1}$ ways to choose H.

We distinguish between the case in which v (the vertex common to the cycles C_j) is in

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

U and the case in which v is in V. We will show that in each of these two cases there are at most $d^d n^{r-\lceil\frac{\lambda-2}{2}\rceil-d+1}$ ways to choose the graph H. Assume we have already selected the side of the graph (U or V) from which the vertex v is chosen. To choose H, we choose the vertices of the cycles and the path. Every cycle C_j passes through v, and we may assume it starts at v. Therefore we choose the vertex v once, and then for every cycle C_j have to choose $s_j - 1$ more vertices. In order to choose the cycle $\tilde{C}_j$ we have to choose $\tilde{s}_j$ vertices, and in order to choose the path L we have to choose s vertices (because the starting point of the path is given). In total we have to choose a sequence of

$$1 + \sum_{j=1}^{\lceil\frac{\lambda-2}{2}\rceil}(s_j - 1) + \sum_{j=1}^{h}\tilde{s}_j + s = r - \left\lceil\frac{\lambda-2}{2}\right\rceil + 1$$

vertices. We can assume the cycles $\tilde{C}_j$ begin in U and therefore, as the starting point of the cycles C_j is v and the starting point of L is given, the side of the graph from which each vertex in the sequence is chosen in predetermined. This means that there are at most n ways to choose each vertex (d cannot be larger than n). In addition there are at least d vertices which have to be chosen from side U (as $|V(H) \cap U| = d$ and by assumption v_0 is from side V). This gives the bound $d^d n^{r-\lceil\frac{\lambda-2}{2}\rceil-d+1}$. To conclude the proof we multiply the bounds calculated in (a), (b) and (c) and use Stirling's inequality $(\frac{d}{e})^d\sqrt{2\pi d} \leqslant d!$, which implies $\binom{m}{d} d^d \leqslant \frac{1}{2}m^d e^d d^{-1/2}$. □

The next lemma is the basic inequality that underlies Lemma 2.5.

Lemma 2.8. *Let $(\alpha_1, \ldots, \alpha_n)$ be a vector such that $\alpha_i \geqslant 0$ for all i.*

Let $X = (X_1, \ldots, X_n) = (\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(n)})$ be a random vector where σ is a random permutation uniformly distributed over S_n. Then, for any $I \subseteq \{1, \ldots, n\}$ and $j \notin I$,

$$E\prod_{i\in I} X_i \leqslant \prod_{i\in I} EX_i = \left(\frac{1}{n}\sum_{i=1}^{n}\alpha_i\right)^{|I|}.$$

Proof. We first show that $EX_j \prod_{i\in I} X_i \leqslant EX_j \cdot E\prod_{i\in I} X_i$.

For $I \subseteq \{1, \ldots, n\}$ denote $\alpha_I = \Pi_{i\in I}\ \alpha_i$. Fix i and j.

$$(\alpha_i - \alpha_j) \sum_{\substack{i\notin I,|I|=k\\ j\notin J,|J|=k}}(\alpha_I - \alpha_J) = (\alpha_i - \alpha_j) \sum_{i,j\notin K,|K|=k}(\alpha_K - \alpha_K) + (\alpha_i - \alpha_j)$$

$$\times \sum_{i,j\notin K,|K|=k-1}(\alpha_{K\cup\{j\}} - \alpha_{K\cup\{i\}})$$

$$= 2\alpha_i\alpha_j \sum_{i,j\notin K,|K|=k-1}\alpha_K - \alpha_i^2 \sum_{i,j\notin K,|K|=k-1}\alpha_K - \alpha_j^2 \sum_{i,j\notin K,|K|=k-1}\alpha_K$$

$$= -(\alpha_i - \alpha_j)^2 \sum_{i,j\notin K,|K|=k-1}\alpha_K \leqslant 0.$$

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

Let Y be an i.i.d. copy of X. It follows from the above inequality that

$$0 \geqslant E_{X,Y}\left(\prod_{j\in I} X_i - \prod_{i\in I} Y_i\right)(X_j - Y_j),$$

and therefore

$$0 \geqslant EX_j\prod_{i\in I} X_i + EY_j\prod_{i\in I} Y_i - EY_j\prod_{i\in I} X_i - EX_j\prod_{i\in I} Y_i = 2EX_j\prod_{i\in I} X_i - 2EX_j\cdot E\prod_{i\in I} X_i.$$

This shows that $EX_j\Pi_{i\in I}X_i \leqslant EX_j\cdot E\Pi_{i\in I}X_i$. The lemma is proved by repeatedly applying this inequality. □

Proof of Lemma 2.5. Let

$$a^2 = \max\left\{\|a_{i\bullet}\|^2, \frac{1}{n}\sum_{\substack{l\leqslant i\leqslant m\\ l\leqslant j\leqslant n}} a_{i,j}^2\right\}.$$

For every point in the sample space we have $\|| S_{\sigma,\varepsilon,\mu}(A)\||^2 \geqslant a^2$. Therefore, using Lemma 2.8 for the second inequality and

$$\frac{\sum_{1\leqslant j\leqslant n} a_{\mu(i),j}^2}{a^2} \leqslant 1$$

for the last inequality,

$$\begin{aligned}
E_\mu E_\sigma\prod_{x\in H} p(x) &\leqslant E_\mu E_\sigma\prod_{(i,j)\in H}\frac{a_{\mu(i),\sigma_i(j)}^2}{a^2} = E_\mu\prod_i E_{\sigma_i}\prod_{(i,j)\in H}\frac{a_{\mu(i),\sigma_i(j)}^2}{a^2}\\
&\leqslant E_\mu\prod_{1\leqslant i\leqslant m}\left(\frac{1}{n}\frac{\sum_{1\leqslant j\leqslant n} a_{\mu(i),j}^2}{a^2}\right)^{|\{j\,:\,(i,j)\in H\}|}\\
&\leqslant\left(\frac{1}{n}\right)^r E_\mu\prod_{\{i\,:\,\exists j(i,j)\in H\}}\left(\frac{\sum_{1\leqslant j\leqslant n} a_{\mu(i),j}^2}{a^2}\right).
\end{aligned}$$

If $m \leqslant n$ then, because

$$\frac{\sum_{1\leqslant j\leqslant n} a_{\mu(i),j}^2}{a^2} \leqslant 1,$$

this proves the lemma. Otherwise, we use Lemma 2.8 again, together with the definition of a^2, which implies that

$$\sum_{i=1}^m\frac{\sum_{j=1}^n a_{ij}^2}{a^2} \leqslant n,$$

and get

$$E_\mu\prod_{\{i\,:\,\exists j(i,j)\in H\}}\left(\frac{\sum_{1\leqslant j\leqslant n} a_{\mu(i),j}^2}{a^2}\right) \leqslant \left(\frac{1}{m}\sum_{i=1}^m\frac{\sum_{j=1}^n a_{ij}^2}{a^2}\right)^{|\{i\,:\,\exists j(i,j)\in H\}|} \leqslant \left(\frac{n}{m}\right)^d,$$

which completes the proof. □

3. The expected norm of a matrix of entries with random signs

Let $M_{m,n}$ be the set of real $m \times n$ martices. Define

$$C(m,n,h) = \left(\sup_{A \in M_{m,n}} \frac{E\|\varepsilon \circ A\|^h}{|||A|||^h} \right)^{1/h},$$

where $\circ$ is the Hadamard product for matrices $(A \circ B = (a_{ij}b_{ij}))$ and $\varepsilon = (\varepsilon_{ij})1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n$ is a matrix of i.i.d. random variables such that, for every $i, j, P\{\varepsilon_{ij} = +1\} = P\{\varepsilon_{ij} = -1\} = \frac{1}{2}$. We will show that $C(m,n,h) \sim \log^{1/4} \min\{m,n\}$ (for $h \leqslant 2 \log \max\{m,n\}$).

First we prove an upper bound for $C(m,n,h)$.

Theorem 3.1. *There exists a constant c such that for every m, n and $h \leqslant 2 \log \max\{m,n\}, C(m,n,h) \leqslant c(\log \min\{m,n\})^{1/4}$.*

Proof. We may assume, without loss of generality, that $n \leqslant m$ and prove that $C(m,n,h) \leqslant c(\log n)^{1/4}$.

Let $A \in M_{m,n}$. We begin by defining $B, P, G(U,V), \gamma$ and $O_i(\gamma)$ as in the proof of Theorem 2.1, only now we do not average over permutations, that is,

$$B = \frac{\varepsilon \circ A}{|||\varepsilon \circ A|||} = \frac{\varepsilon \circ A}{|||A|||}.$$

We also let $k = \lceil \log n \rceil$ and begin with the inequality

$$E\|B\|^h \leqslant \left(\sum_{1 \leqslant i_1,\ldots,i_k \leqslant n} \sum_{1 \leqslant j_1,\ldots,j_k \leqslant m} E_\varepsilon b_{i_1 j_1} b_{i_1 j_2} b_{i_2 j_2} \ldots b_{i_k j_k} b_{i_k j_1} \right)^{h/2k}.$$

The entries of the matrix B are constant up to a sign and their signs are independent, each with equal probability for $+1$ and -1. Therefore (again as in the previous proof) $E_\varepsilon b_{i_1 j_1} b_{i_1 j_2} b_{i_2 j_2} \ldots b_{i_k j_k} b_{i_k j_1} = 0$ whenever some entry b_{ij} appears an odd number of times. We are therefore interested only in walks on the graph G which traverse every edge an even number of times. Unless otherwise specified, all walks in the sequel will be such walks.

Basically, as in the proof of Theorem 2.1, we want to replace the above products with products of p_{ij}s on the odd steps. It is then easy to see that the odd steps add nothing to the sum while every even step adds at most a factor of k to the total sum. To prove the theorem, however, we can allow an extra k factor only for every two even steps. It is again easy to see that two consecutive even steps do not contribute together more than a factor of k. The problem is that even steps do not always come in pairs. The main idea of the proof is to slightly redistribute the p_{ij}s so that steps on which these weights are not counted appear consecutively in pairs.

Step 1. Let γ be a walk of length $2k$ on the graph G that traverses every edge an even number of times. We give every step in γ a code in the following manner.

(a) We mark a step with a '+' if it is an 'odd' step and otherwise with a '–'.

(b) In addition, we mark a '–' step with an 'r' if between this step and the previous step

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

which passes through the same edge (and is necessarily a '+' step) there are only '−' steps. We mark all other '−' steps with an '*'.

(c) In addition to the previous marks, we also mark each step with either 'p' or '1' as follows.

First, we mark all '+' steps with a 'p' and all '−' steps with a '1'. Now we go consecutively over the steps of the walk from first to last. Assume we reached the ith step, and that this step is currently marked by (+, p). If this step is separated from the next '+' step by an odd number of '−' steps, we change the ith step to (+, 1). In this case we mark the step in which we first return to the same edge with a 'p'. (Such a step exists, as the walk passes through every edge an even number of times. Up to now this step was marked either $(-, r, 1)$ or $(-, *, 1)$ and will now be marked $(-, r, p)$ or $(-, *, p)$.) Note that this change may have an effect on changes that will be made further on along the sequence. In any other case we do not change the way the steps are marked.

Three properties hold.

(1) These changes do not change the total number of steps marked by 'p' and the total number of steps marked by '1'. Therefore at the end of the process we have k edges of each type.

(2) The changes associated with the ith step do not change the number of steps from among the immediately following '−' steps (*i.e.*, up to the next '+' step) which are marked by (−, *, 1). This is because if the first return to the edge of the ith step is in one of these steps, this step must be marked by (−, r, 1) (and after the change by (−, r, p)).

(3) For every edge in the graph, exactly half the steps in the walk which pass through it are marked by 'p' and the other half are marked by '1' (in total always an even number of steps).

In this way we have assigned each walk γ a sequence of length $2k$ with elements taken from the set $\{+\} \times \{p, 1\} \cup \{-\} \times \{r, *\} \times \{p, 1\}$, with an equal number of $+$s and $-$s in each sequence and an equal number of ps and 1s in each sequence. The possible number of such sequences is no more than $\binom{2k}{k}\binom{2k}{k}2^k$ and therefore bounded from above by $(4\sqrt{2})^{2k}$.

Denote by $\mathbf{T}$ the set of all such possible sequences (only sequences which are produced by some walk γ are considered here).

For a given walk γ, denote

$$\pi_i(\gamma) = \prod_{\substack{1 \leqslant j \leqslant i \\ t_j(\gamma) = `p'}} p(x_j(\gamma))$$

where $t_j(\gamma)$ is the code with which the jth step of the walk is marked (we will always write only that part of the code which is of interest to us). By Remark 3 above, we have $b_{i_1 j_1} b_{i_1 j_2} b_{i_2 j_2} \ldots b_{i_k j_k} b_{i_k j_1} = \pi_{2k}(\gamma)$ for every product in which every entry appears an even number of times.

For $T \in \mathbf{T}$ and $v_0 \in V$ denote $\Gamma_{v_0}(T) = \{\gamma : v_0(\gamma) = v_0, T(\gamma) = T\}$. We will show that,

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

for every $T \in \mathbf{T}$ and $v_0 \in V$,

$$\sum_{\gamma \in \Gamma_{v_0}(T)} \pi_{2k}(\gamma) \leqslant k^{k/2}. \tag{3.1}$$

Hence the theorem follows directly:

$$E\|B\|^h \leqslant \left(\sum_{v_0 \in V} \sum_{T \in \mathbf{T}} \sum_{\gamma \in \Gamma_{v_0}(T)} \pi_{2k}(\gamma) \right)^{h/2k}$$

$$\leqslant \left(n (4\sqrt{2})^{2k} k^{k/2} \right)^{h/2k} \leqslant (\sqrt{e}\, 4\sqrt{2} k^{1/4})^h \leqslant (c(\log n)^{1/4})^h.$$

Step 2. It remains to prove (3.1). Fix $v_0 \in V$ and $T \in \mathbf{T}$. We will call a sequence of steps $i_1, i_1+1, \ldots, i_2$ a '$-$block' (or simply a 'block') if all the steps i_1 through i_2 are '$-$' steps and this block is maximal (that is, not contained in any larger block of '$-$' steps). If, in addition, the $i_1 - 1$ step is marked by $(+,\, 1)$, we will call $i_1 - 1, i_1, i_1 + 1, \ldots, i_2$ an 'extended block'. Otherwise, by definition, the extended block is the same as the block. By the construction of the sequence T, every extended block contains an even number of steps marked by '1' but not by 'r' (possibly zero such steps). We divide these steps into pairs (in the obvious way, by their order in the sequence). In what follows we will simply refer to each such pair as a 'pair'. A step which appears between the first step and second step of the same pair will be said to be 'inside' the pair (this does not include the steps of the pair itself).

Let $1 = i_1 < i_2 < \cdots < i_\lambda = 2k$ be the sequence of all steps in T which are not a first step in a pair or a step appearing 'inside' a pair. In other words, the steps appearing in the sequence are of one of the following types.

(a) A step which is not part of any extended block. This is possible if and only if this is a $(+,\, p)$ step.
(b) A $(-,\, *,\, p)$ step which does not appear inside a pair.
(c) A $(-, r)$ step (either $(-, r, p)$ or $(-, r, 1)$) which does not appear inside a pair.
(d) A second step in a pair (this must be a $(-,\, *,\, 1)$ step).

We will show that, for every step in the sequence $1 = i_1 < i_2 < \cdots < i_\lambda = 2k$,

$$\sum_{\{\gamma(i_j)\,:\,\gamma \in \Gamma_{v_0}(T)\}} \pi_{i_j}(\gamma) \leqslant k^{\lambda_{i_j}(T)}, \tag{3.2}$$

where $\lambda_{i_j}(T)$ is the number of pairs up to the i_jth step. Because every step in a pair is marked by a '1', $\lambda_{i_\lambda}(T) = \lambda_{2k}(T) \leqslant \frac{k}{2}$. Therefore, for $j = \lambda$, (3.1) follows from (3.2).

Step 3. To complete the proof we prove (3.2) by induction on the sequence $1 = i_1 < i_2 < \cdots < i_\lambda = 2k$.

The first step in T must be a $(+,\, p)$ step. Therefore, as the weight matrix P is doubly sub-stochastic, it is easy to see that (3.2) holds for $i_1 = 1$.

We now proceed by induction on j and prove (3.2) for i_j, as follows.

(1) Step i_j is of type (a) or (b).

Because this step does not appear inside a pair and cannot be a first or second step in a pair, $i_{j-1} = i_j - 1$. The i_j step is a 'p' step, and therefore, by the induction hypothesis:

$$\begin{aligned}\sum_{\{\gamma(i_j)\,:\,\gamma\in\Gamma_{v_0}(T)\}} \pi_{i_j}(\gamma) &\leqslant \sum_{\{\gamma(i_{j-1})\,:\,\gamma\in\Gamma_{v_0}(T)\}} \pi_{i_{j-1}}(\gamma) \sum_{v\in V_{i_j}} p((v_{i_{j-1}}(\gamma), v)) \\ &\leqslant \sum_{\{\gamma(i_{j-1})\,:\,\gamma\in\Gamma_{v0}(T)\}} \pi_{i_{j-1}}(\gamma) \leqslant k^{\lambda_{i_{j-1}}(T)} = k^{\lambda_{i_j}(T)},\end{aligned}$$

where $V_{i_j} = U$ or V.

(2) Step i_j is of type (c).
As in the previous cases, $i_{j-1} = i_j - 1$. The step is a $(-, r)$ step and therefore $\gamma(i_j)$ is uniquely determined by $\gamma(i_{j-1})$. It follows directly from the induction hypothesis that

$$\sum_{\{\gamma(i_j)\,:\,\gamma\in\Gamma_{v_0}(T)\}} \pi_{i_j}(\gamma) \leqslant \sum_{\{\gamma(i_{j-1})\,:\,\gamma\in\Gamma_{v_0}(T)\}} \pi_{i_{j-1}}(\gamma) \leqslant k^{\lambda_{i_{j-1}}(T)} = k^{\lambda_{i_j}(T)}.$$

(3) Step i_j is of type (d).

The step i_j is the second step in a pair. The first step in the same pair is $i_{j-1}+1$. The step i_j must be a (–, *, 1) step while the step $i_{j-1}+1$ may be either a (–, *, 1) step or a (+, 1) step. Between the step $i_{j-1}+1$ and the step i_j only (–, *, p), (–, r, p) and (–, r, 1) steps may appear. The step i_j must traverse one of the edges in the graph $O_{i_{j-1}}(\gamma)$. This is true because the only '+' step possible between the step $i_{j-1}+1$ and the step i_j is $i_{j-1}+1$ itself (which may possibly be a (+, 1) step at the beginning of an extended block) but if the step i_j passes the same edge as the step $i_{j-1}+1$, it must be marked by an 'r', and such steps cannot be part of a pair. Therefore, given $\gamma(i_{j-1})$ there are at most k choices for the step i_j (the number of edges in $O_{i_{j-1}}(\gamma)$ cannot exceed k).

Let us assume first that all the steps $i_{j-1}+2, \ldots, i_j - 1$ (possibly zero such steps) are of type 'p':

$$\begin{aligned}\sum_{\{\gamma(i_j)\,:\,\gamma\in\Gamma_{v_0}(T)\}} \pi_{i_j}(\gamma) &\leqslant \sum_{\{\gamma(i_j)\,:\,\gamma\in\Gamma_{v_0}(T)\}} \pi_{i_{j-1}}(\gamma) \prod_{\lambda=i_{j-1}+2}^{i_j-1} p((v_{\lambda-1}(\gamma), v_\lambda(\gamma))) \\ &\leqslant \sum_{\{\gamma(i_{j-1})\,:\,\gamma\in\Gamma_{v_0}(T)\}} \pi_{i_{j-1}}(\gamma) \sum_{(v_{i_j-1}, v_{i_j})\in O_{i_{j-1}}(\gamma)} \sum_{v_{i_j-2}\in V_{i_j-2}} p((v_{i_j-2}, v_{i_j-1}))\cdots \\ &\quad\times \sum_{v_{i_{j-1}+1}\in V_{i_{j-1}+1}} p((v_{i_{j-1}+1}, v_{i_{j-1}+2})) \\ &\leqslant \sum_{\{\gamma(i_{j-1}):\gamma\in\Gamma_{v_0}(T)\}} \pi_{i_{j-1}}(\gamma) \sum_{(v_{i_j-1}, v_{i_j})\in O_{1_{j-1}}(\gamma)} 1 \leqslant k \sum_{\{\gamma(i_{j-1}):\gamma\in\Gamma_{v_0}(T)\}} \pi_{i_{j-1}}(\gamma) \leqslant k\cdot k^{\lambda_{i_{j-1}}(T)} = k^{\lambda_{i_j}(T)}.\end{aligned}$$

The only possible case in which not all of the steps $i_{j-1}+2, \ldots, i_j-1$ are of type 'p' is when one of these steps is a $(-, r, 1)$ step. In this case, the first step in the pair (the step $i_{j-1}+1$) cannot be a (+, 1) step (because then the 'r' step must be a $(-, r, p)$ step. Therefore, the 'r' step is uniquely determined by $\gamma(i_{j-1})$. Assume this 'r' step is step number i', then the vertices $v_{i'-1}(\gamma)$ and $v_{i'}(\gamma)$ are uniquely determined by $\gamma(i_{j-1})$ and therefore we may repeat

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

the above calculation only replacing the expression

$$\sum_{v_{i'-1} \in V_{i'-1}} p((v_{i'-1}, v_{i'})),$$

which appears in the original product, by 1. The result of the calculation remains valid after this change and therefore (3.2) holds in this last case too.

This completes the proof of (3.2) and of the whole theorem. □

We now show that this upper bound is best possible (up to a multiplicative constant). For any $n \leqslant m, C(n,n,h) \leqslant C(n,m,h)$. In addition, as $\|\| A \|\|$ is constant, $C(n,n,1) \leqslant C(n,n,h)$. Therefore it is enough to prove the following theorem.

Theorem 3.2. *For every $0 < \delta < 1$ there exists N such that for every $N < n$ there exists an $n \times n$ matrix such that $E_\varepsilon \|\varepsilon \circ A\| \geqslant \delta (\log n)^{1/4} \|\| A \|\|$.*

Proof. Let $k = \lfloor \sqrt{\log n} \rfloor$ and $\lambda = \lfloor \frac{n}{k} \rfloor$. We define A to be the block matrix

$$A = \begin{pmatrix} A_1 & & & & \\ & A_2 & & 0 & \\ & & \cdot & & \\ & & & \cdot & \\ & 0 & & & \cdot \\ & & & & A_\lambda \end{pmatrix},$$

where every A_h is a $k \times k$ matrix of identical entries all equal to $k^{-1/2}$. Denote by ε_h the sub-matrix of ε which corresponds to A_h.

The probability that all the entries in ε_h are $+1$ is 2^{-k^2}, and therefore

$$P(\|\varepsilon_h \circ A_h\| = \sqrt{k}) \geqslant 2^{-k^2}.$$

Then,

$$\begin{aligned} E\|\varepsilon \circ A\| &= E \max_h \|\varepsilon_h \circ A_h\| \geqslant \sqrt{k} P\{\max_h \|\varepsilon_h \circ A_h\| \geqslant \sqrt{k}\} \\ &= \sqrt{k}\left(1 - P\{\|\varepsilon_1 \circ A_1\| < \sqrt{k}\}^{\lfloor \frac{n}{k} \rfloor}\right) \\ &\geqslant \sqrt{k}\left(1 - (1 - 2^{-k^2})^{\lfloor \frac{n}{k} \rfloor}\right) \\ &\geqslant \sqrt{\lfloor \sqrt{\log n} \rfloor}\left(1 - (1 - 2^{-\log n})^{\frac{n}{\sqrt{\log n}} - 1}\right) \\ &= \sqrt{\lfloor \sqrt{\log n} \rfloor} = \left(1 - (1 - n^{-\log 2})^{\frac{n}{\sqrt{\log n}} - 1}\right). \end{aligned}$$

Fix $0 < \delta < 1$ and let v be such that $\sqrt{\delta} < 1 - e^{-v}$.

For a sufficiently large n, $n^{1 - \log 2} > (v+1)\sqrt{\log n}$, from which it follows that

$$\frac{n}{\sqrt{\log n}} - 1 = \frac{n - \sqrt{\log n}}{\sqrt{\log n}} > v n^{\log 2}$$

https://doi.org/10.1017/S096354830000420X Published online by Cambridge University Press

and therefore

$$(1 - n^{-\log 2})^{\frac{n}{\sqrt{\log n}}-1} < e^{-\nu}.$$

Substituting this inequality together with $\sqrt{\lfloor\sqrt{\log n}\rfloor} \geqslant \sqrt{\delta}(\log n)^{1/4}$ into the previous inequality, we get that $E\|\varepsilon \circ A\| \geqslant \sqrt{\delta}(1 - e^{-\nu})(\log n)^{1/4} \geqslant \delta(\log n)^{1/4}$.

This, together with the fact that $|||A||| = 1$, completes the proof. □

References

[1] Chevet, S. (1978) Séries de variables aléatoires gaussiennes à valeurs dans $E \hat{\otimes}_\varepsilon F$: Applications aux produits d'espaces de Wiener abstraits. *Séminaire sur la Géométrie des Espaces de Banach 1977–78*, Exp XIX, Ecole Polytechnique, Paris.

[2] Chevet, S. (1978) Quelques nouveaux résultats sur les mesures cylandriques. In Vol. 644 of *Lecture Notes in Mathematics*, Springer, Berlin, pp. 125–158.

[3] Füredi, Z. and Komlós, J. (1981) The eigenvalues of random symmetric matrices. *Combinatorica* **1** 233–241.

[4] Geman, S. (1980) A limit theorem for the norm of random matrices. *Ann Probab.* **8** 252–261.

[5] Hansen, P. S. (1988) The 2-norm of random matrices. *J. Comput. Appl. Math.* **23** 117–120.

[6] Ledoux, M. and Talagrand, M. (1991) *Probability in Banach Spaces*, Springer, Berlin.

[7] Lovász, L. (1979) *Combinatorial Problems and Exercises*, Akademiai Kiado, North Holland, Budapest/Amsterdam.

[8] Yin, Y. Q., Bai, Z. D. (1986) Spectra for large dimensional random matrices. In *Random Matrices and their Applications* (J. E. Cohen, H. Kesten, C. N. Newman, eds), Vol. 50 of *Contemporary Mathematics*, American Mathematical Society, Providence, RI.