Concentration of norms and eigenvalues of random matrices

Mark W. Meckes[*],[1]

Department of Mathematics, Case Western Reserve University, Cleveland, OH 44106, USA

Abstract

We prove concentration results for ℓ_p^n operator norms of rectangular random matrices and eigenvalues of self-adjoint random matrices. The random matrices we consider have bounded entries which are independent, up to a possible self-adjointness constraint. Our results are based on an isoperimetric inequality for product spaces due to Talagrand.
© 2003 Elsevier Inc. All rights reserved.

Keywords: Random matrices; Eigenvalues; Concentration of measure; Large deviations; Talagrand's inequality

1. Introduction

In this paper we prove concentration results for norms of rectangular random matrices acting as operators between ℓ_p^n spaces, and eigenvalues of self-adjoint random matrices. Except for the self-adjointness condition when we consider eigenvalues, the only assumptions on the distribution of the matrix entries are independence and boundedness. Our approach is based on a powerful isoperimetric inequality for product probability spaces due to Talagrand [20].

Throughout this paper $X = X_{m,n}$ will stand for an $m \times n$ random matrix with real or complex entries x_{jk}. (Specific technical conditions on the x_{jk}'s will be introduced as needed for each result below.) If $1 \leqslant p, q \leqslant \infty$ and A is an $m \times n$ matrix, we denote by

---

*Corresponding author. Department of Mathematics, Stanford University, Stanford, CA 94305, USA.
E-mail address: mark@math.stanford.edu.

$||A||_{p \to q}$ the operator norm of $A : \ell_p^n \to \ell_q^m$. We denote by $p' = p/(p-1)$ the conjugate exponent of p. For a real random variable Y we denote by $\mathbb{E}Y$ the expected value and by $\mathbb{M}Y$ any median of Y. Our first main result is the following.

Theorem 1. *Let* $1 < p \leqslant 2 \leqslant q < \infty$. *Suppose the entries* x_{jk} *of* X *are independent complex random variables, each supported in a set of diameter at most D. Then*

$$\mathbb{P}[| \, ||X||_{p \to q} - \mathbb{M}||X||_{p \to q}| \geqslant t] \leqslant 4\exp\left[-\frac{1}{4}\left(\frac{t}{D}\right)^r\right] \tag{1}$$

for all $t > 0$, *where* $r = \min\{p', q\}$.

To prove Theorem 1 we show that Talagrand's isoperimetric inequality, which at first appears adapted primarily to prove normal concentration for functions which are Lipschitz with respect to a Euclidean norm, actually implies sometimes stronger concentration for functions which are Lipschitz with respect to more general norms. In particular, as we show in Corollary 4 below, one obtains concentration of the kind in (1) for convex functions which are Lipschitz with respect to ℓ_r norms for $r \geqslant 2$. Since such functions are automatically Lipschitz with respect to the Euclidean norm, one can apply the known $r = 2$ case of this fact directly, but would then obtain the upper bound with r replaced by 2 in the r.h.s. of (1). Since the conclusion of Theorem 1 is trivial when $t/D \leqslant 1$, estimate (1) is stronger than the estimate one would obtain this way.

To put Theorem 1 in perspective, we consider the particular case in which $p = q'$, $m = n$, and $\mathbb{P}[x_{jk} = 1] = \mathbb{P}[x_{jk} = -1] = 1/2$ for all j, k. In this situation,

$$n^{1/q} \leqslant \mathbb{M}||X||_{q' \to q} \leqslant Cn^{1/q},$$

where $C > 0$ is a universal constant. Theorem 1 implies that, while $||X||_{q' \to q}$ achieves values as large as $n^{2/q}$, it is comparable to its median except on a set whose probability decays exponentially quickly as $n \to \infty$. Furthermore, in this situation the estimate in (1) is sharp as long as $n^{-1/q}t$ is sufficiently large and $n^{-2/q}t$ is sufficiently small. These observations apply more generally; see the remarks in Section 3 following the proof of Theorem 1.

If A is a self-adjoint $n \times n$ matrix, we denote by $\lambda_1(A) \geqslant \lambda_2(A) \geqslant \cdots \geqslant \lambda_n(A)$ the eigenvalues of A, counted with multiplicity. Our second main result is the following.

Theorem 2. *Suppose* $m = n$ *and the entries* x_{jk}, $1 \leqslant j \leqslant k \leqslant n$, *of* X *are independent complex random variables such that*:

(i) *for* $1 \leqslant j \leqslant n$, x_{jj} *is real and is supported on an interval of length at most* $\sqrt{2}D$; *and*
(ii) *for* $1 \leqslant j < k \leqslant n$, *either* x_{jk} *is supported on a set of diameter at most D; or* $x_{jk} = w_{jk}(\alpha_{jk} + i\beta_{jk})$, *where* $w_{jk} \in \mathbb{C}$ *is a constant with* $|w_{jk}| \leqslant 1$, *and* α_{jk}, β_{jk} *are independent real random variables each supported in intervals of length at most D*;

and that $x_{jk} = \overline{x_{kj}}$ for $k < j$. Then

$$\mathbb{P}[|\lambda_1(X) - \mathbb{M}\lambda_1(X)| \geqslant t] \leqslant 4e^{-t^2/8D^2} \tag{2}$$

for all $t > 0$, and the same holds if $\lambda_1(X)$ is replaced by $\lambda_n(X)$. Furthermore, for each $2 \leqslant k \leqslant n - 1$, there exists an $M_k \in \mathbb{R}$ such that

$$\mathbb{P}[|\lambda_k(X) - M_k| \geqslant t] \leqslant 8 \exp\left[-\left(\frac{t}{2\sqrt{2}(\sqrt{k} + \sqrt{k-1})D}\right)^2\right]$$

$$< 8 \exp\left[-\frac{t^2}{32kD^2}\right] \tag{3}$$

for all $t > 0$, and the same upper bound holds if $\lambda_k(X)$ is replaced by $\lambda_{n-k+1}(X)$ and M_k by M_{n-k+1}.

Note that Theorem 2 applies in particular to the case of real symmetric random matrices with off-diagonal entries supported in intervals of length D and diagonal entries supported in intervals of length $\sqrt{2}D$.

The proof of Theorem 2 is also based on Talagrand's theorem, in this case applying it only to functions which are Lipschitz with respect to a Euclidean norm. Theorem 2 is (up to numerical constants) a sharpening and generalization of a result of Alon et al. [1]. Their proof is also based on Talagrand's theorem, although they apply it in a very different way. For perspective, we note that in the particular case in which $\mathbb{P}[x_{jk} = 1] = \mathbb{P}[x_{jk} = -1] = 1/2$, $\mathbb{M}\lambda_1(X)$ is of the order $\sqrt{n}$, while $\lambda_1(X)$ can achieve values as large as n. Furthermore, in this situation the estimate in (2) is sharp when $n^{-1/2}t$ is sufficiently large and $n^{-1}t$ is sufficiently small. (See [1] for a discussion of this point when X is the adjacency matrix of the random graph $G(n, 1/2)$.) However, the estimate in (3) is probably not sharp in its dependence on k. See the remarks in Section 3 following the proof of Theorem 2 for details.

We emphasize that our results are of interest as bounds for large deviations. Beginning with the work of Tracy and Widom [22,23], which has been refined and extended in [2,11,17], it is known that, at least in some situations, the kind of result contained in (2), while nontrivial, is not sharp when $t = o(\sqrt{n})$. More precisely, it has been established in [2] that for certain self-adjoint Gaussian random matrices (normalized in such a way as to be comparable to X in our Theorem 2), one has concentration of the largest eigenvalue of the form

$$\mathbb{P}[\lambda_1(X) - \mathbb{E}\lambda_1(X) \geqslant t] \leqslant C \exp[-\max\{c_1 t^2, c_2(n^{1/6}t)^{3/2}\}], \tag{4}$$

where $C, c_1, c_2 > 0$ are constants. The results of [17], although they do not imply an estimate of the type in (4) for fixed n, show that such an estimate holds asymptotically (as $n \to \infty$) also for more general random matrices; the result is probably also true for fixed n. A related result for the largest singular value of a rectangular Gaussian random matrix is proved in [9].

We also emphasize that aside from the uniform boundedness assumption, the distributions of the independent entries of X in Theorems 1 and 2 are completely arbitrary. In particular, there is no assumption of identical distribution of independent entries, and no assumption about the values of their means. For example, our approach can be used to study small random perturbations of an arbitrary fixed matrix.

Talagrand's theorem was first applied in the context of random matrices by Guionnet and Zeitouni [8], who used it to prove a concentration result for the spectral measure of self-adjoint random matrices, and who also remarked that the same methods give concentration results for other functionals of self-adjoint matrices. For general discussions of applications of concentration of measure phenomena to random matrices, see the survey [6] by Davidson and Szarek and Section 8.5 of the book [13] by Ledoux.

In Section 2, we show how to obtain concentration for Lipschitz functions on ℓ_q sum spaces (and more general sums of normed spaces) from Talagrand's isoperimetric inequality. In Section 3, we prove Theorems 1 and 2, and give an infinite-dimensional version of Theorem 1 and a version of Theorem 2 for singular values of rectangular matrices. We also compare the results obtained by our methods with the corresponding results for Gaussian random matrices obtained from the Gaussian isoperimetric inequality.

2. General concentration results

We first need some notation. Let $(\Omega_1, \Sigma_1, \mu_1), \ldots, (\Omega_N, \Sigma_N, \mu_N)$ be probability spaces, $\Omega = \Omega_1 \times \cdots \times \Omega_N$, $\mathbb{P} = \mu_1 \otimes \cdots \otimes \mu_N$. For $x = (x_1, \ldots, x_N) \in \Omega$, $y = (y_1, \ldots, y_N) \in \Omega$, $h(x, y) \in \mathbb{R}^N$ is defined by

$$h(x, y)_j = \begin{cases} 0 & \text{if } x_j = y_j, \\ 1 & \text{if } x_j \neq y_j. \end{cases}$$

For $A \subseteq \Omega$, $x \in \Omega$, $U_A(x) = \{h(x, y) : y \in A\} \subset \mathbb{R}^N$. Finally, we define the convex hull distance from x to A by

$$f_c(A, x) = \inf\{|z| : z \in \text{conv } U_A(x)\}, \tag{5}$$

where $|\cdot|$ is the standard Euclidean norm and conv denotes the convex hull. (In fact, $U_A(x)$ is usually defined to be the larger set

$$\{h \in \{0, 1\}^N : \exists y \in A \text{ such that } h_j = 0 \Rightarrow x_j = y_j\},$$

but this difference does not affect the value of $f_c(A, x)$.) Talagrand's isoperimetric inequality from [20] is the following.

Theorem 3 (Talagrand). *Let $(\Omega, \mathbb{P})$ be a product probability space as above. For any $A \subseteq \Omega$,*

$$\int_{\Omega} \exp\left[\frac{1}{4}f_c^2(A,x)\right] d\mathbb{P}(x) \leqslant \frac{1}{\mathbb{P}(A)},$$

which by Markov's inequality implies

$$\mathbb{P}(\{x : f_c(A,x) \geqslant t\}) \leqslant \frac{1}{\mathbb{P}(A)} e^{-t^2/4}$$

for all $t > 0$.

As in [20], we have ignored measurability issues in the statement of Theorem 3. To be strictly correct, the integrals and probabilities which appear must be replaced by upper integrals and outer probabilities; however, this issue is irrelevant in applications, since one typically applies such a result to estimate expressions in which all the functions and sets which appear are measurable.

Let $E = (\mathbb{R}^N, ||\cdot||_E)$ be such that the standard basis of $\mathbb{R}^N$ is 1-unconditional for $||\cdot||_E$. (We will refer to such E as a 1-unconditional space.) For normed vector spaces $(V_j, ||\cdot||_{V_j})$, $j = 1, \ldots, N$, we denote by

$$V_E = \left(\bigoplus_{j=1}^{N} V_j\right)_E$$

the direct sum of vector spaces with the norm

$$||(v_1, \ldots, v_N)||_{V_E} = ||(||v_1||_{V_1}, \ldots, ||v_N||_{V_N})||_E.$$

Theorems 1 and 2 will be proved using the following consequence of Theorem 3.

Corollary 4. *Let $(V_j, ||\cdot||_{V_j})$, $j = 1, \ldots, N$ be normed vector spaces and $V = V_{\ell_q^N}$ for $q \geqslant 2$. For $j = 1, \ldots, N$, let μ_j be a probability measure on V_j which is supported on a compact set of diameter at most 1. Let $\mathbb{P} = \mu_1 \otimes \cdots \otimes \mu_N$. Suppose $F : V \to \mathbb{R}$ is 1-Lipschitz and quasiconvex, that is, $F^{-1}((-\infty, a])$ is convex for all $a \in \mathbb{R}$. Then*

$$\mathbb{P}[|F - \mathbb{M}F| \geqslant t] \leqslant 4e^{-t^q/4} \tag{6}$$

for all $t > 0$.

We will postpone the proof of Corollary 4 until after some remarks. The $q = 2$ case of Corollary 4 has been widely noted and applied in various degrees of generality already; see [13,16,20,21] and their references. As observed in the introduction, if F is 1-Lipschitz with respect to the ℓ_q sum norm on V, then F is also 1-Lipschitz with respect to the ℓ_2 sum norm, so the $q = 2$ case of Corollary 4 applies, but yields directly only the weaker upper bound $4e^{-t^2/4}$ in inequality (6).

Corollary 4 can be applied in the case that F is L-Lipschitz and each μ_i is supported on a set of diameter at most D, by replacing t with t/LD in the r.h.s. of (6). This fact is used implicitly in the proofs in Section 3. Although it is not immediately

obvious from Theorem 3, the statement of Corollary 4 allows the formal replacement of F by $-F$; thus Corollary 4 also applies when F is quasiconcave, that is, when $F^{-1}([a, \infty))$ is convex for all $a \in \mathbb{R}$. In particular, Corollary 4 applies to both convex and concave Lipschitz functions. Talagrand gives an example which shows that some form of convexity assumption in Corollary 4 cannot be removed in general in [19], in which the special case of Theorem 3 for the uniform measure on the discrete cube was first proved.

The conclusion of Corollary 4 does not hold in general for functions which are only Lipschitz with respect to an ℓ_p sum norm for $1 \leqslant p < 2$ without the introduction of dimension-dependent constants, even if the bound $4e^{-t^p/4}$ is replaced by any other dimension-independent function which approaches 0 at infinity. To see this, let $\{Y_j : j \in \mathbb{N}\}$ be independent random variables with $\mathbb{P}[Y_j = 0] = \mathbb{P}[Y_j = 1] = 1/2$ for each j, and $S_n = \sum_{j=1}^n Y_j$. Then $S_n^{1/p} = ||(Y_1, \ldots, Y_n)||_p$, and $(n/2)^{1/p}$ is a median for $S_n^{1/p}$. Suppose we have a concentration result which implies there exists a function f with $\lim_{t \to \infty} f(t) = 0$ such that for all n and for all $t > 0$,

$$\mathbb{P}[S_n^{1/p} - \mathbb{M}S_n^{1/p} \geqslant t] \leqslant f(t). \tag{7}$$

Then by Taylor's theorem,

$$
\begin{aligned}
\mathbb{P}\left[\frac{S_n - n/2}{\sqrt{n}/2} \geqslant t\right] &= \mathbb{P}\left[S_n^{1/p} \geqslant \left(\frac{n}{2} + \frac{\sqrt{n}}{2}t\right)^{1/p}\right] \\
&= \mathbb{P}\left[S_n^{1/p} - \left(\frac{n}{2}\right)^{1/p} \geqslant \frac{1}{2^{1/p}p}n^{\frac{1}{p} - \frac{1}{2}}t + O(n^{\frac{1}{p} - 1})t^2\right] \\
&\leqslant f\left(\frac{1}{2^{1/p}p}n^{\frac{1}{p} - \frac{1}{2}}t + O(n^{\frac{1}{p} - 1})t^2\right),
\end{aligned}
$$

which implies that for any $t > 0$,

$$\lim_{n \to \infty} \mathbb{P}\left[\frac{S_n - n/2}{\sqrt{n}/2} \geqslant t\right] = 0,$$

which contradicts the central limit theorem.

It is also not difficult to see that for the examples of $S_n^{1/q}$ with $q \geqslant 2$, the concentration result of Corollary 4 is sharp, up to the values of numerical constants, when $c = c(q) \leqslant n^{-1/q}t \leqslant 1 - 2^{-1/q}$. Moreover, by gluing together copies of $S_n^{1/q}$ for different values of $1 \leqslant n \leqslant N$, one obtains an example of a Lipschitz function on $\mathbb{R}^N$ for which Corollary 4 is sharp for the entire nontrivial range of t.

It is more typical to state results of the type in Corollary 4 in terms of deviations of a random variable from the mean rather than the median. This difference is inessential, since this level of concentration implies that the median and mean cannot be too far apart. For example, in the situation of Corollary 4, we have

$$|\mathbb{E}F - \mathbb{M}F| \leqslant \mathbb{E}|F - \mathbb{M}F| = \int_0^\infty \mathbb{P}[|F - \mathbb{M}F| \geqslant t]\,dt$$

$$\leqslant 4 \int_0^\infty e^{-t^q/4}\,dt = 4^{1+\frac{1}{q}}\Gamma\left(1 + \frac{1}{q}\right).$$

We now turn to the proof of Corollary 4. Rather than proving Corollary 4 directly from Theorem 3, we will deduce it as a special case of Proposition 5 below, which uses Theorem 3 to derive concentration for functions which are Lipschitz with respect to an arbitrary 1-unconditional norm, in terms of a modulus function for the norm. Theorem 3 bounds the size of the set of points which are far from a set A in terms of the convex hull distance $f_c(A, \cdot)$ from A. Thus it provides concentration for functions which satisfy a Lipschitz type condition with respect to the convex hull distance. However, since in general this distance is not induced by a metric, some care is needed in its application.

Let $E = (\mathbb{R}^N, ||\cdot||_E)$ be a 1-unconditional space. We define

$$K_E(t) = \inf\{|x| : ||x||_E \geqslant t, ||x||_\infty \leqslant 1\}$$

for $t > 0$, where we use the convention that $\inf \emptyset = \infty$.

Proposition 5. *Let V_E be as described before Corollary 4 and let K_E be as above. For $j = 1, \ldots, N$, let μ_j be a probability measure on V_j which is supported on a compact set of diameter at most 1. Let $\mathbb{P} = \mu_1 \otimes \cdots \otimes \mu_N$. Suppose $F : V_E \to \mathbb{R}$ is 1-Lipschitz and quasiconvex. Then*

$$\mathbb{P}[|F - \mathbb{M}F| \geqslant t] \leqslant 4\exp\left[-\frac{1}{4}(K_E(t))^2\right] \tag{8}$$

for all $t > 0$.

It is easy to verify that $K_{\ell_q^N}(t) \geqslant t^{q/2}$ for $q \geqslant 2$ and any $N \in \mathbb{N}$ (cf. Lemma 6 below), so that Corollary 4 follows immediately from Proposition 5.

Proof of Proposition 5. First we show that

$$K_E(\mathrm{dist}(x, \mathrm{conv}\,A)) \leqslant f_c(A, x) \tag{9}$$

for $x = (x_1, \ldots, x_N) \in \mathrm{supp}(\mathbb{P})$ and $\emptyset \neq A \subseteq V_E$, where dist is the distance in the normed space V_E. Let $y^k = (y_1^k, \ldots, y_N^k) \in A \cap \mathrm{supp}(\mathbb{P})$ and $0 \leqslant \theta_k \leqslant 1$ for $k = 1, \ldots, n$ such that $\sum_{k=1}^n \theta_k = 1$. Then for each $j = 1, \ldots, N$,

$$\left\|x_j - \sum_{k=1}^n \theta_k y_j^k\right\|_{V_j} \leqslant \sum_{k=1}^n \theta_k ||(x_j - y_j^k)||_{V_j} \leqslant \sum_{k=1}^n \theta_k h(x, y^k)_j$$

since $x_j, y_j^k \in \mathrm{supp}(\mu_j)$ for each j, k. Then by unconditionality,

$$\mathrm{dist}(x, \mathrm{conv}\, A) \leqslant \left\| x - \sum_{k=1}^{n} \theta_k y^k \right\|_{V_E} \leqslant \left\| \sum_{k=1}^{n} \theta_k h(x, y^k) \right\|_E,$$

and so

$$K_E(\mathrm{dist}(x, \mathrm{conv}\, A)) \leqslant \left| \sum_{k=1}^{n} \theta_k h(x, y^k) \right|,$$

since K_E is nondecreasing. Inequality (9) now follows since the r.h.s. of (5) is precisely the infimum of this last expression over all such finite sequences y^k, θ_k, $k = 1, \ldots, n$. Therefore, by Theorem 3, for any $A \subseteq V_E$,

$$\mathbb{P}(A) \mathbb{P}(\{x: K_E(\mathrm{dist}(x, \mathrm{conv}\, A)) \geqslant t\}) \leqslant e^{-t^2/4}$$

for all $t > 0$. Thus if F is quasiconvex and 1-Lipschitz on V_E, we have that for any $a \in \mathbb{R}$, $t > 0$,

$$\mathbb{P}[F \geqslant a + t] \leqslant \mathbb{P}(\{x: K_E(\mathrm{dist}(x, F^{-1}((-\infty, a]))) \geqslant K_E(t)\})$$

$$\leqslant \frac{1}{\mathbb{P}[F \leqslant a]} \exp\left[-\frac{1}{4}(K_E(t))^2 \right].$$

The proposition follows by applying this in turn with $a = \mathbb{M}F$ and $a = \mathbb{M}F - t$.   □

In order to apply Proposition 5, one needs to estimate the function K_E. This is of most interest if one can bound K_{E_j} uniformly for some family of spaces E_j for which $\sup_j \dim(E_j) = \infty$. This is not difficult to do for certain classes of spaces. For a nonincreasing sequence $w = (w_1, w_2, \ldots)$ of positive numbers and $p \geqslant 1$, the N-dimensional Lorentz space $\ell_{w,p}^N$ is $\mathbb{R}^N$ with the norm

$$\|x\|_{w,p} = \left(\sum_{j=1}^{N} w_j a_j^p \right)^{1/p},$$

where $\{a_j: 1 \leqslant j \leqslant N\}$ is the nonincreasing rearrangement of $\{|x_j|: 1 \leqslant j \leqslant N\}$. For an Orlicz function ψ, that is, a convex nondecreasing function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\psi(0) = 0$ and $\lim_{t \to \infty} \psi(t) = \infty$, the N-dimensional Orlicz space ℓ_ψ^N is $\mathbb{R}^N$ with the norm

$$\|x\|_\psi = \inf\left\{ \rho > 0 : \sum_{j=1}^{N} \psi\left(\frac{|x_j|}{\rho} \right) \leqslant 1 \right\}.$$

Observe that $\ell_p^N = \ell_{w,p}^N$ if $w_j = 1$ for $j = 1, \ldots, N$, and $\ell_p^N = \ell_\psi^N$ if $\psi(t) = t^p$, $p \geqslant 1$. For these two classes of spaces, we have the following elementary estimates, which we state without proof.

Lemma 6. *If $p \geqslant 1$ and $w \in \ell_r$ for some r such that $\max\{1, 2/p\} \leqslant r' < \infty$, then*

$$K_{\ell_{w,p}^N}(t) \geqslant \|w\|_r^{-r'/2} t^{pr'/2}.$$

If ψ is any Orlicz function, then

$$K_{\ell_\psi^N}(t) \geqslant \inf_{0 < u \leqslant 1} \frac{u}{\sqrt{\psi(u/t)}}.$$

In particular, $K_{\ell_q^N}(t) \geqslant t^{q/2}$ for $q \geqslant 2$.

Note that the estimates in Lemma 6 may be trivial and are not necessarily optimal, but when they are nontrivial, they are valid in all dimensions. By considering vectors $x \in \{0, 1\}^N$, one can see that the estimate $K_{\ell_q^N}(t) \geqslant t^{q/2}$ for $q \geqslant 2$ is sharp for $t = k^{1/q}$, $k = 1, \ldots, N$.

Observe that the proof of Proposition 5 actually gives separate tail estimates for deviations of F above and below its median; the same is therefore true of Corollary 4 as well. The full generality of Proposition 5 can in fact be derived with some amount of argument from the (known) $q = 2$ case of Corollary 4, using these bounds separately; however, we find it simpler to argue directly from the isoperimetric inequality of Theorem 3 as above. One could alternatively prove Corollary 4 by proving an ℓ_q version of Theorem 3, by defining an ℓ_q convex hull distance $f_q(A, x) = \inf\{\|z\|_q : z \in \mathrm{conv}\, U_A(x)\}$ and mimicking the proof of Theorem 3; or as a corollary to the more general and abstract Theorem 4.2.4 in [20]. However, this approach would result only in a slight sharpening of the constant $1/4$ which appears in the exponent.

We remark that to use Proposition 5 to full advantage for non-Euclidean norms, one must use a nonlinear lower bound on K_E and make use of the restriction $\|x\|_\infty \leqslant 1$. If, for example, one uses only the fact that $\|x\|_q \leqslant |x|$ for all x when $q \geqslant 2$, then one is using no more than the fact that a function which is 1-Lipschitz with respect to the ℓ_q norm is 1-Lipschitz with respect to the ℓ_2 norm, which, as we have observed already in the introduction, leads to a weaker concentration result.

It is instructive to compare the general concentration results above and the applications in the next section with the corresponding results for Gaussian measures. We begin by recalling the functional form of the Gaussian isoperimetric inequality, due independently to Borell [4] and Sudakov and Tsirel'son [18]. Let γ_N be the standard Gaussian measure on $\mathbb{R}^N$ defined by $d\gamma_N(x) = (2\pi)^{-N/2} e^{-|x|^2/2}\, dx$, where $|\cdot|$ is again the standard Euclidean norm.

Theorem 7 (Borell, Sudakov-Tsirel'son). *Let $F: \mathbb{R}^N \to \mathbb{R}$ be 1-Lipschitz with respect to the Euclidean norm on $\mathbb{R}^N$. Then*

$$\gamma_N(\{x:\ F(x) \geqslant \mathbb{M}F + t\}) \leqslant 1 - \gamma_1((-\infty, t]) < \frac{1}{2} e^{-t^2/2}$$

for all $t > 0$.

Observe that by composing F with an affine contraction, one obtains the same conclusion in Theorem 7 if the standard Gaussian measure γ_N is replaced by the product of one-dimensional Gaussian measures with arbitrary means and variances at most 1. Thus the $q = 2$ case of Corollary 4 provides a level of concentration for quasiconvex Lipschitz functions of independent bounded random variables comparable to the concentration of Lipschitz functions of independent Gaussian random variables with bounded variances.

Recall that a similar normal concentration principle is obeyed by any probability measure which satisfies a logarithmic Sobolev inequality, and that products of such measures also satisfy a logarithmic Sobolev inequality (see [12]). In particular, whenever we state concentration results below for random matrices with Gaussian entries, similar results hold under the weaker assumption of entries with uniformly bounded logarithmic Sobolev constants. We remark that Guionnet and Zeitouni [8] also proved a concentration result for the spectral measure in the case that the matrix entries satisfy a logarithmic Sobolev inequality.

3. Norms and eigenvalues of random matrices

Since any norm is a convex function, Proposition 5 can be applied directly to obtain concentration of norms of a random matrix X; all that is necessary is to estimate the function K_E, or the Lipschitz constant of the given norm with respect to one for which a bound on K_E is known. Note that by the triangle inequality, in order to estimate the Lipschitz constant of one norm with respect to another norm, it suffices to estimate the appropriate equivalence constant.

Proof of Theorem 1. For an $m \times n$ matrix A, let $A_j \in \mathbb{C}^n$ denote the jth row of A. Then Hölder's inequality implies

$$\|A\|_{p \to q} \leqslant \|(\|A_1\|_{p'}, \ldots, \|A_m\|_{p'})\|_q \leqslant \|(a_{jk})\|_r, \tag{10}$$

where (a_{jk}) represents the matrix A thought of as an element of $\mathbb{C}^{mn}$, and we recall that $r = \min\{p', q\}$. The claim follows by using this estimate and taking $V_j = \mathbb{C}$ for each j in Corollary 4. Alternatively, inequality (10) and Lemma 6 imply that

$$K_{L(\ell_p^n, \ell_q^m)}(t) \geqslant t^{r/2},$$

where $L(\ell_p^n, \ell_q^m)$ is identified with $\mathbb{C}^{mn}$, so that the claim follows from Proposition 5. $\quad \square$

Observe that for p, q in other ranges, one can derive concentration for $||X||_{p \to q}$ by comparing the $\ell_{p'}^n$ or ℓ_q^m norm to the ℓ_2 norm of the appropriate dimension. In this case one will obtain normal concentration on a scale which depends on m or n.

We remark that Theorem 1 can be extended to more general norms on $\mathfrak{M}_{m,n}(\mathbb{C})$ by using Proposition 5 together with estimates on the corresponding function K_E. In particular, as long as one has the appropriate Lipschitz estimates, the underlying normed spaces need not be unconditional, nor must the norm on matrices even be an operator norm.

Now for comparison, we let $G = G_{m,n}$ be an $m \times n$ random matrix whose entries are independent Gaussian random matrices with arbitrary means and variances at most 1. For $1 \leqslant p \leqslant 2 \leqslant q \leqslant \infty$, $||A||_{p \to q} \leqslant ||A||_2$ for any $m \times n$ matrix A, where $||A||_2$ is the Hilbert–Schmidt norm of A. Then Theorem 7 implies that

$$\mathbb{P}[|\,||G||_{p \to q} - \mathbb{M}||G||_{p \to q}| \geqslant t] < e^{-t^2/2}$$

for all $t > 0$. Observe that this is comparable to what one would obtain in the cases of independent bounded entries by using only the $q = 2$ case of Corollary 4.

Theorem 1 implies that the order of fluctuations of $||X_{m,n}||$ about its median is $O(1)$, independent of m and n. In typical situations, the median itself grows without bound as m or n does. Suppose for example that $\mathbb{E}|x_{jk}| \geqslant c > 0$ for all j, k. (In the situation of Theorem 1, this will be the case for example if each x_{jk} is real, $|x_{jk}| \leqslant 1$, $\mathbb{E}x_{jk} = 0$, and x_{jk} has variance at least c.) Then

$$\mathbb{E}||X||_{p \to q} \geqslant \mathbb{E}||Xe_1||_q \geqslant m^{1/q-1} \mathbb{E}||Xe_1||_1 = m^{1/q-1} \sum_{j=1}^{m} \mathbb{E}|x_{j1}| \geqslant cm^{1/q}.$$

Since $||X||_{p \to q} = ||X^*||_{q' \to p'}$, we obtain $\mathbb{E}||X||_{p \to q} \geqslant c \max\{m^{1/q}, n^{1/p'}\}$. As remarked earlier, $\mathbb{M}||X||_{p \to q}$ will also have at least this order when the hypotheses of Theorem 1 are satisfied.

A similar upper estimate is possible in the case $p = q'$. Suppose that each x_{jk} is a symmetric real random variable such that $|x_{jk}| \leqslant 1$. We note first that by the Riesz convexity theorem,

$$||X||_{q' \to q} \leqslant ||X||_{2 \to 2}^{\frac{2}{q}} ||X||_{1 \to \infty}^{1-\frac{2}{q}} \leqslant ||X||_{2 \to 2}^{\frac{2}{q}}.$$

By the contraction principle (see [14, Theorem 4.4]),

$$\mathbb{E}||X||_{2 \to 2} \leqslant \mathbb{E}||\widetilde{X}||_{2 \to 2},$$

where $\widetilde{X} = \widetilde{X}_{m,n}$ is an $m \times n$ matrix whose entries are independent Rademacher random variables; that is, $\mathbb{P}[\widetilde{x}_{jk} = 1] = \mathbb{P}[\widetilde{x}_{jk} = -1] = 1/2$ for all j, k. By standard

comparisons between Rademacher and Gaussian averages and Chevet's inequality [5] (see also [14]),

$$\mathbb{E}||\widetilde{X}||_{2\to2}\leqslant C(m^{1/2}+n^{1/2}),$$

where $C>0$ is an absolute numerical constant. Combining these estimates with Jensen's inequality, we obtain

$$\mathbb{E}||X||_{q'\to q}\leqslant 2C\max\{m^{1/q},n^{1/q}\}.$$

(The argument above is entirely standard and the estimate is probably known, although we could not find a reference in the literature.)

The example of $\widetilde{X}$ above can be used to show that the estimate in Theorem 1 is sharp for large enough values of t up to numerical constants in the case that $p=q'$. For $1\leqslant a\leqslant m$, $1\leqslant b\leqslant n$,

$$\mathbb{P}[||\widetilde{X}||_{q'\to q}\geqslant(ab)^{1/q}]\geqslant\mathbb{P}[\widetilde{X}\text{ has an }a\times b\text{ all-1 submatrix}]\geqslant 2^{-ab},$$

so that

$$\mathbb{P}[||\widetilde{X}||_{q'\to q}\geqslant t]\geqslant 2^{-t^q}$$

for $t=(ab)^{1/q}$, $a=1,\ldots,m$, $b=1,\ldots,n$. Together with the above upper bound on $\mathbb{E}||\widetilde{X}||_{q'\to q}$, this implies that in this situation, the concentration result of Theorem 1 is sharp when $(\max\{m,n\})^{-1/q}t$ is sufficiently large, up to the values of numerical constants.

Since the conclusion of Theorem 1 is independent of dimension, one can derive the following infinite-dimensional version for kernel operators from ℓ_p to ℓ_q.

Corollary 8. *Let $1<p\leqslant2\leqslant q<\infty$, and let $c_{jk}\geqslant0$, $j,k\in\mathbb{N}$, be constants such that*

$$\left(\sum_{j=1}^{\infty}\left(\sum_{k=1}^{\infty}c_{jk}^{p'}\right)^{q/p'}\right)^{1/q}<\infty. \tag{11}$$

Suppose that x_{jk}, $j,k\in\mathbb{N}$ are independent complex random variables each supported in a set of diameter at most D, such that $|x_{jk}|\leqslant c_{jk}$ for all j,k. Define the random operator $X:\ell_p\to\ell_q$ by setting $X(e_j)=\sum_{k=1}^{\infty}x_{jk}e_k$. Then

$$\mathbb{P}[|\,||X||-\mathbb{M}||X||\,|\geqslant t]\leqslant4\exp\left[-\frac{1}{4}\left(\frac{t}{D}\right)^r\right]$$

for all $t>0$, where $||X||$ is the operator norm of X and $r=\min\{p',q\}$.

We remark that when $p=q'$, the l.h.s. of (11) was shown by Persson [15] to coincide with both the q-summing norm $\pi_q(T)$ and the q-nuclear norm $\nu_q(T)$

of the kernel operator $T : \ell_{q'} \to \ell_q$ given by $T(e_j) = \sum_{k=1}^{\infty} c_{jk} e_k$. The same method of proof in this case gives the same conclusion in Corollary 8 if $||X||$ is replaced by $\pi_q(X) = v_q(X)$.

Proof of Corollary 8. The fact that $|x_{jk}| \leqslant c_{jk}$ implies that $||X|| < \infty$ always. Apply Theorem 1 to the $n \times n$ upper-left corner of the infinite matrix (x_{jk}), and use (11) and the estimate $|x_{jk}| \leqslant c_{jk}$ to pass to the limit $n \to \infty$. $\square$

Note that by taking $c_{jk} = 0$ when $j > m$ or $k > n$ in Corollary 8, we recover Theorem 1, so that these two statements are formally equivalent.

We now specialize to the case in which $m = n$ and consider X as an operator on ℓ_2^n, so that we use only the $q = 2$ case of Corollary 4. Guionnet and Zeitouni [8] noted that this concentration theorem implies normal concentration for any function on (self-adjoint) matrices which is convex and Lipschitz with respect to the Hilbert–Schmidt norm. For example, we have the following. Let the entries x_{jk} of X all be independent, and satisfying condition (ii) in the statement of Theorem 2, and for simplicity let $D = 1$. For $1 \leqslant p \leqslant \infty$, we denote by $||A||_p$ the Schatten p-norm of an $n \times n$ matrix A (see, e.g., [3]). Then for all $t > 0$,

$$\mathbb{P}[|\,||X||_p - \mathbb{M}||X||_p\,| \geqslant t] \leqslant 4e^{-t^2/4}$$

for $2 \leqslant p \leqslant \infty$, and

$$\mathbb{P}[|\,||X||_p - \mathbb{M}||X||_p\,| \geqslant t] \leqslant 4 \exp\left[-\frac{t^2}{4n^{\frac{2}{p}-1}}\right]$$

for $1 \leqslant p < 2$. (In particular, we observe that when $p = q = 2$, the conclusion of Theorem 1 holds when the matrix entries x_{jk} satisfy condition (ii) in the statement of Theorem 2.) Furthermore, since $|||A||| \leqslant ||A||_1 \leqslant \sqrt{n} ||A||_2$ for any unitarily invariant norm $||| \cdot |||$ on $\mathfrak{M}_n(\mathbb{C})$ satisfying $|||E_{11}||| = 1$, it follows that

$$\mathbb{P}[|\,|||X||| - \mathbb{M}|||X|||\,| \geqslant t] \leqslant 4e^{-t^2/4n}$$

for all $t > 0$ for any such norm. Each of these observations is in fact a special case of the tail inequalities for norms of sums of independent vector-valued random variables which were the original motivation for Talagrand's development of Theorem 3 and related concentration theorems.

We now consider eigenvalues of a self-adjoint random matrix. Although these are not (except in the extreme cases) quasiconvex or quasiconcave functions, Corollary 4 can still be used to derive concentration.

Proof of Theorem 2. For simplicity, we assume $D = 1$. First observe that

$$||X||_2 = \left(\sum_{j,k=1}^{n} |x_{jk}|^2 \right)^{1/2} = \sqrt{2} \left(\sum_{j=1}^{n} \left| \frac{x_{jj}}{\sqrt{2}} \right|^2 + \sum_{1 \leqslant j < k \leqslant n} |x_{jk}|^2 \right)^{1/2}.$$

We suppose for simplicity that each of the upper-diagonal entries x_{jk} for $j < k$ is supported in a set of diameter at most 1. Note that $\frac{x_{jj}}{\sqrt{2}}$, $j = 1, \ldots, n$, and x_{jk}, $1 \leqslant j < k \leqslant n$, are independent random variables in $\mathbb{R}$ or $\mathbb{C}$, each supported in a set of diameter at most 1. $||X||_2$ is $\sqrt{2}$ times the ℓ_2 sum norm of the direct sum of n copies of $\mathbb{R}$ and $n(n-1)/2$ copies of $\mathbb{C}$ spanned by these variables.

Recall also that

$$||X||_2 = \left(\sum_{k=1}^{n} \lambda_k(X)^2 \right)^{1/2},$$

which implies that each $\lambda_k(X)$ is a 1-Lipschitz function of X with respect to $||X||_2$. The first claim now follows directly from Corollary 4 with $V_j = \mathbb{C}$ or $V_j = \mathbb{R}$ for each j, since λ_1 is a convex function, and λ_n is concave.

To prove the second claim, we introduce the following functions for a self-adjoint matrix A. For $k = 1, \ldots, n$, let

$$F_k(A) = \sum_{j=1}^{k} \lambda_j(A),$$

$$G_k(A) = \sum_{j=1}^{k} \lambda_{n-j+1}(A) = \operatorname{Tr} A - F_k(A).$$

Then F_k and G_k are $\sqrt{k}$-Lipschitz with respect to the Hilbert–Schmidt norm. Moreover, F_k is convex and G_k is concave for each k; this follows from Ky Fan's maximum principle (see, e.g., [3]) or Davis's characterization [7] of all convex unitarily invariant functions of a self-adjoint matrix. Let $M_k = \mathbb{M}F_k - \mathbb{M}F_{k-1}$. Then by Corollary 4, for any $0 \leqslant \theta \leqslant 1$,

$$\mathbb{P}[|\lambda_k(X) - M_k| \geqslant t] = \mathbb{P}[|(F_k(X) - \mathbb{M}F_k(X)) - (F_{k-1}(X) - \mathbb{M}F_{k-1}(X))| \geqslant t]$$

$$\leqslant \mathbb{P}[|F_k(X) - \mathbb{M}F_k(X)| \geqslant \theta t]$$

$$+ \mathbb{P}[|F_{k-1}(X) - \mathbb{M}F_{k-1}(X)| \geqslant (1-\theta)t]$$

$$\leqslant 4 \exp\left[-\left(\frac{\theta t}{2\sqrt{2k}} \right)^2 \right] + 4 \exp\left[-\left(\frac{(1-\theta)t}{2\sqrt{2(k-1)}} \right)^2 \right].$$

Estimate (3) now follows by letting $\theta = \sqrt{k}/(\sqrt{k} + \sqrt{k-1})$. (This is not the optimal value of θ, but optimizing at this point would only result in a slight sharpening of the constants, and not of the dependence on t or k.) The claim for $\lambda_{n-k+1}(X)$ follows similarly, using $G_k(X)$ in place of $F_k(X)$, or as a formal consequence by replacing X with $-X$. $\quad\square$

Now, for comparison, we let H_n be an $n \times n$ random matrix with entries h_{jk}, $1 \leqslant j, k \leqslant n$, such that:

(i) the entries h_{jk}, $1 \leqslant j \leqslant k \leqslant n$ are independent Gaussian random variables,
(ii) the variance of h_{jk} for $1 \leqslant j < k \leqslant n$ is at most 1,
(iii) the variance of h_{jj} is at most $\sqrt{2}$ for $1 \leqslant j \leqslant n$, and
(iv) $h_{jk} = h_{kj}$ for $k < j$.

Then for each $1 \leqslant k \leqslant n$, Theorem 7 implies that

$$\mathbb{P}[|\lambda_k(H_n) - \mathbb{M}\lambda_k(H_n)| \geqslant t] < e^{-t^2/4}$$

for all $t > 0$. This is comparable to the result of Theorem 2 for $\lambda_1(X)$ and $\lambda_n(X)$, but the same level of concentration holds for eigenvalues in the bulk of the spectrum, which is not the case in Theorem 2.

The result of Theorem 2 for $\lambda_1(X)$ and $\lambda_n(X)$ (stated in less generality) was shown by Krivelevich and Vu [10]. After a preliminary version of this paper was written, we learned that Alon et al. [1] showed that for $1 \leqslant k \leqslant n$,

$$\mathbb{P}[|\lambda_k(X) - \mathbb{M}\lambda_k(X)| \geqslant t] \leqslant 4 \exp\left[-\frac{t^2}{8k^2 D^2}\right]$$

for all $t > 0$, and that the same holds if $\lambda_k(X)$ is replaced by $\lambda_{n-k+1}(X)$. The approach in [1] handles the lack of convexity of λ_k not by using the $q = 2$ case of Corollary 4, but instead applying Theorem 3 by directly estimating the convex hull distances involved. Our Theorem 2 improves the order of fluctuations of $\lambda_k(X)$ from $O(k)$ (as in [1]) to $O(\sqrt{k})$. It is also conjectured in [1] that $\lambda_k(X)$ should be concentrated at least as strongly as $\lambda_1(X)$, as one obtains from Theorem 7 in the Gaussian case. We emphasize again that we are dealing only with large deviations here. As we have already indicated in the introduction, the tail estimate (2) for the extreme eigenvalues is probably not sharp for $t = o(\sqrt{n})$; furthermore, it is likely that concentration is even tighter for eigenvalues in the bulk of the spectrum.

It follows as in the discussion following Corollary 4 that Theorem 2 implies that $\mathbb{E}\lambda_k(X)$ differs by at most $O(\sqrt{k})$ from the number M_k which appears in the statement of the theorem. One can also show that the number M_k which appears in the statement of the theorem differs by at most $O(\sqrt{k})$ from $\mathbb{M}\lambda_k(X)$. Specifically, by using the separate bounds for deviations above and below the median in the

situation of Corollary 4, we have

$$|M_k - \mathbb{M}\lambda_k(X)| \leqslant 2\sqrt{6\log 2}(\sqrt{k} + \sqrt{k-1})D.$$

We can also obtain a similar result to Theorem 2 for singular values in the rectangular case. Let $l = \min\{m, n\}$. For an $m \times n$ matrix A, we denote by $s_1(A) \geqslant s_2(A) \geqslant \cdots \geqslant s_l(A) \geqslant 0$ the singular values of A, counted with multiplicity; that is, $s_k(A) = \lambda_k((A^*A)^{1/2})$.

Theorem 9. *Suppose the entries x_{jk} of X are independent complex random variables, each satisfying condition* (ii) *in the statement of Theorem 2. Then*

$$\mathbb{P}[|s_1(X) - \mathbb{M}s_1(X)| \geqslant t] \leqslant 4e^{-t^2/4D^2}$$

for all $t > 0$. Furthermore, for each $2 \leqslant k \leqslant \min\{m, n\}$, there exists an $M_k \in \mathbb{R}$ such that

$$\mathbb{P}[|s_k(X) - M_k| \geqslant t] \leqslant 8\exp\left[-\left(\frac{t}{2(\sqrt{k}+\sqrt{k-1})D}\right)^2\right] < 8\exp\left[-\frac{t^2}{16kD^2}\right]$$

for all $t > 0$.

The proof is similar to the proof of Theorem 2, using in place of the functions F_k the Ky Fan k-norms, defined by

$$\|A\|_{(k)} = \sum_{j=1}^{k} s_j(A)$$

for $1 \leqslant k \leqslant \min\{m, n\}$.

Acknowledgments

References

[1] N. Alon, M. Krivelevich, V.H. Vu, On the concentration of eigenvalues of random symmetric matrices, Israel J. Math. 131 (2002) 259–267.
[2] G. Aubrun, A sharp small deviation inequality for the largest eigenvalue of a random matrix, preprint available at http://www.eleves.ens.fr/home/aubrun/, 2002.
[3] R. Bhatia, Matrix Analysis, Springer, New York, 1997.
[4] C. Borell, The Brunn–Minkowski inequality in Gauss space, Invent. Math. 30 (2) (1975) 207–216.

[5] S. Chevet, Séries de variables aléatoires gaussiennes à valeurs dans $E \otimes_\varepsilon F$. Application aux produits d'espaces de Wiener abstraits, in Séminaire sur la Géométrie des Espaces de Banach (1977–1978), Exp. No. 19, 15. École Polytech., Palaiseau, 1978.

[6] K.R. Davidson, S.J. Szarek, Local operator theory, random matrices and Banach spaces, in: W.B. Johnson and J. Lindenstrauss (Eds.), Handbook of the Geometry of Banach Spaces, Vol. I, North-Holland, Amsterdam, 2001, pp. 317–366.

[7] C. Davis, All convex invariant functions of hermitian matrices, Arch. Math. 8 (1957) 276–278.

[8] A. Guionnet, O. Zeitouni, Concentration of the spectral measure for large matrices, Electron. Comm. Probab. 5 (2000) 119–136 (electronic).

[9] I.M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Ann. Statist. 29 (2) (2001) 295–327.

[10] M. Krivelevich, V.H. Vu, Approximating the independence number and the chromatic number in expected polynomial time, J. Combin. Optim. 6 (2) (2002) 143–155.

[11] M. Ledoux, A remark on hypercontractivity and tail inequalities for the largest eigenvalues of random matrices, in: Séminaire de Probabilités XXXVII, Lecture Notes in Mathematics, Springer, Berlin, to appear.

[12] M. Ledoux, Concentration of measure and logarithmic Sobolev inequalities, in: Séminaire de Probabilités, XXXIII, Lecture Notes in Mathematics, Vol. 1709, Springer, Berlin, 1999, pp. 120–216.

[13] M. Ledoux, The Concentration of Measure Phenomenon, American Mathematical Society, Providence, RI, 2001.

[14] M. Ledoux, M. Talagrand, Probability in Banach Spaces, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), Vol. 23, Springer, Berlin, 1991.

[15] A. Persson, On some properties of p-nuclear and p-integral operators, Studia Math. 33 (1969) 213–222.

[16] G. Schechtman, Concentration, results and applications, in: W.B. Johnson and J. Lindenstrauss (Eds.), Handbook of the Geometry of Banach Spaces, Vol. II, North-Holland, Amsterdam, 2003.

[17] A. Soshnikov, Universality at the edge of the spectrum in Wigner random matrices, Comm. Math. Phys. 207 (3) (1999) 697–733.

[18] V.N. Sudakov, B.S. Tsirel'son, Extremal properties of half-spaces for spherically invariant measures, J. Soviet. Math. 9 (1978) 9–18, 1978 (translated from Zap. Nauch. Sem. L.O.M.I. 41 (1974) 14–24).

[19] M. Talagrand, An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities, Proc. Amer. Math. Soc. 104 (3) (1988) 905–909.

[20] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, Inst. Hautes Études Sci. Publ. Math. 81 (1995) 73–205.

[21] M. Talagrand, A new look at independence, Ann. Probab. 24 (1) (1996) 1–34.

[22] C.A. Tracy, H. Widom, Level-spacing distributions and the Airy kernel, Comm. Math. Phys. 159 (1) (1994) 151–174.

[23] C.A. Tracy, H. Widom, On orthogonal and symplectic matrix ensembles, Comm. Math. Phys. 177 (3) (1996) 727–754.