3 Gaussian random matrices

3a Introduction

The theory of random matrices makes the hypothesis that the characteristic energies of chaotic systems behave locally as if they were the eigenvalues of a matrix with randomly distributed elements.
. . . but when the complications increase beyond a certain point the situation becomes hopeful again, for we are no longer required to explain the characteristics of every individual state but only their average properties, which is much simpler.

M.L. Mehta[1]

Nothing is random in the famous number $\pi = 3.14159265358979323 84\dots$; nevertheless, for the best of our knowledge, statistical properties of its (decimal) digits do not differ from statistical properties of independent uniform random digits.[2]

Similarly, nothing is random in the energy levels of (say) the U^{239} nucleus; nevertheless, their statistical properties appear to be reasonably close to statistical properties of random matrices, considered below (in 3e).

See the introduction to a recent survey [2] for random matrices appearing in problems of statistics, physics, number theory, operator algebras and combinatorics.

Maintaining the tradition, I speak about random matrices, but in fact I think about random (linear) operators in an n-dimensional Euclidean space. If we choose an orthonormal basis in this space, then the operators become

[1]See Preface and Introduction to the book "Random matrices", second edition, Academic Press, 1991.

[2]Still, no one is able to *prove* even a small part of this observation.

$n \times n$-matrices $M \in \mathrm{M}_n(\mathbb{R})$. A different basis leads to a different matrix $O^{-1}MO$, where $O \in \mathrm{O}(n)$ is an orthogonal matrix (that is, $|Ox| = |x|$ for $x \in \mathbb{R}^n$, or equivalently, $O^{-1} = O^*$).

The unique (up to a coefficient) $\mathrm{O}(n)$-invariant (that is, invariant under $M \mapsto O^{-1}MO$ for $O \in \mathrm{O}(n)$) linear form (functional) on $\mathrm{M}_n(\mathbb{R})$ is the trace,

$$\operatorname{trace}(M) = m_{1,1} + \dots + m_{n,n} \, .$$

The so-called Hilbert-Schmidt norm $\| \cdot \|_{\mathrm{HS}}$ (denoted also $\| \cdot \|_2$),

$$\|M\|_{\mathrm{HS}} = \sqrt{\operatorname{trace}(M^*M)} = \Big(\sum_{k,l} m_{k,l}^2 \Big)^{1/2}$$

(mentioned also in 2a) is the square root of an $\mathrm{O}(n)$-invariant quadratic form on $\mathrm{M}_n(\mathbb{R})$. In contrast to the usual operator norm $\|M\| = \sup\{|Mx| : |x| \le 1\}$, the HS norm turns $\mathrm{M}_n(\mathbb{R})$ into a Euclidean (not just normed) space. The corresponding Gaussian measure (of dimension n^2) on $\mathrm{M}_n(\mathbb{R})$ is especially important: on one hand, it is $\mathrm{O}(n)$-invariant,[1] and on the other hand, it turns the matrix elements $m_{k,l}$ of the random matrix into *independent* (and identically distributed) random variables. A coefficient $1/\sqrt{n}$ is convenient, as we will see:

(3a1) $$m_{k,l} = \frac{1}{\sqrt{n}} \zeta_{k,l} \, ,$$

where $(\zeta_{k,l})_{k,l=1,\dots,n}$ are n^2 orthogaussian functions (on $(0,1)$, or $(\mathbb{R}^{n^2}, \gamma^{n^2})$, or another probability space).

The symmetric matrix (that is, self-adjoint operator)

$$A = \frac{1}{2}(M^* + M)$$

is distributed according to the standard Gaussian measure on the Euclidean space of all symmetric matrices (equipped with the norm $\sqrt{n}\| \cdot \|_{\mathrm{HS}}$);

(3a2) $$a_{k,k} = \frac{1}{\sqrt{n}} \zeta_{k,k} \, , \quad a_{k,l} = a_{l,k} = \frac{1}{\sqrt{2n}} \zeta_{k,l}$$

for $1 \le l < k \le n$; here $\zeta_{k,l}$ are $\frac{n(n+1)}{2}$ orthogaussian functions.

[1] Moreover, it is in fact invariant under $M \mapsto MO$ and $M \mapsto OM$ separately (for all $O \in \mathrm{O}(n)$), not only $M \mapsto O^{-1}MO$.

3b Estimating the norm

Let M be a random $n \times n$-matrix distributed according to (3a1).

The distribution of Mx does not depend on the choice of a unit vector $x \in \mathbb{R}^n$ (due to the $\mathrm{O}(n)$-invariance) and is equal to $\gamma^n_{1/\sqrt{n}}$. Thus,

$$|Mx| = 1 + O\Big(\frac{1}{\sqrt{n}}\Big) \tag{3b1}$$

in the sense that

$$\mathbb{P}\Big(1 - \frac{c}{\sqrt{n}} \le |Mx| \le 1 + \frac{c}{\sqrt{n}}\Big) \to 1 \quad \text{as } c \to \infty \text{, uniformly in } n \tag{3b2}$$

(recall 2c4, (2c5)). Using the fact that $\sqrt{n}(|Mx| - 1)$ is more concentrated than $\mathrm{N}(0,1)$ it is easy to see that $\max(|Me_1|, \dots, |Me_n|)$ typically is close to 1; however, it does not mean that $\|M\|$ is close to 1. In fact, it is not! Rather, $\|M\| = 2 + O(n^{-2/3})$, and moreover, the limiting distribution of $n^{2/3}(\|M\| - 2)$ exists.[1] The following weaker result is proven below.

3b3 Proposition. For every $\varepsilon > 0$,

$$\mathbb{P}\big(\|M\| \le 4 + \varepsilon\big) \to 1 \quad \text{as } n \to \infty \,.$$

Here is the idea of the proof. For a given x, $|Mx|$ typically is close to 1. Therefore, it holds for most pairs (x, M). Therefore, for a typical M, it holds for most points x. Therefore, $\|M\|$ is not too large.

We start the proof with the latter argument: if $\|M\|$ is large then $|Mx|$ is large for many x.

3b4 Exercise. For every $Z \in \mathrm{M}_n(\mathbb{R})$ and $c \in (0, \infty)$,

$$\gamma^n\{x : |Zx| \ge c\} \ge 2\Phi\Big(-\frac{c}{\|Z\|}\Big) \,.$$

Prove it.

Hint: take $y \in \mathbb{R}^n$ such that $|y| = 1$ and $|Z^* y| = \|Z\|$, then $|Zx| \ge |\langle Zx, y\rangle|$ and $\langle Zx, y\rangle = \langle x, Z^* y\rangle \sim \mathrm{N}(0, \|Z\|^2)$ for $x \sim \gamma^n$.

Now we estimate $|Mx|$ from above for random independent $x \sim \gamma^n$ and $M \sim \beta_n$, where β_n is the relevant Gaussian measure on $\mathrm{M}_n(\mathbb{R})$. We note that

$$|Mx| = \sqrt{n} \frac{|Mx|}{|x|} \frac{|x|}{\sqrt{n}} \,,$$

[1] The limiting distribution of $4n^{2/3}(\|M\| - 2)$ is the Tracy-Widom law of order 1, see I.M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis", Ann. Statist. **29**:2, 295–327 (2001).

$|x|/\sqrt{n}$ is concentrated near 1, and $|Mx|/|x|$ is (independent of x and) distributed like $|x|/\sqrt{n}$ (think, why).

By the Gaussian isoperimetry, $|x|$ is more concentrated than $\mathrm{N}(0,1)$. Also, the median of $|x|$ does not exceed $\sqrt{n}$ (I did not prove it, but I use it anyway; you may prove easily that the median does not exceed $\sqrt{n}+1$, and correct the proof accordingly). Thus,

$$\gamma^n\{x : |x| \ge 2\sqrt{n}\} \le \Phi(-\sqrt{n})\,. \tag{3b5}$$

3b6 Exercise. Prove that

$$(\gamma^n \times \beta_n)\{(x,Z) : |Zx| \ge 4\sqrt{n}\} \le \mathrm{e}^{-n/2}$$

for large n.

Hint: $|Zx| \ge 4\sqrt{n}$ implies $\frac{|Zx|}{|x|} \ge 2$ or $\frac{|x|}{\sqrt{n}} \ge 2$ (or both); use the Fubini theorem, and note that $\Phi(-\sqrt{n}) \sim \frac{\mathrm{const}}{\sqrt{n}}\mathrm{e}^{-n/2}$.

3b7 Exercise. Prove that

$$\int \Phi\Big(-\frac{4\sqrt{n}}{\|Z\|}\Big)\,\beta_n(\mathrm{d}Z) \le \mathrm{e}^{-n/2}$$

for large n.

Hint: integrate 3b4 in Z and use 3b6.

3b8 Exercise. Prove Proposition 3b3.

Hint: $\mathbb{P}\big(\|M\| \ge 4+\varepsilon\big) = \mathbb{P}\big(\Phi(-\frac{4\sqrt{n}}{\|Z\|}) \ge \Phi(-\frac{4\sqrt{n}}{4+\varepsilon})\big) \le \frac{\mathbb{E}\,\Phi(-\frac{4\sqrt{n}}{\|Z\|})}{\Phi(-\frac{4\sqrt{n}}{4+\varepsilon})} \to 0$.

The threshold, 4, in Proposition 3b3 can be improved to $2\sqrt{2} \approx 2.82$ by replacing the crude estimate $\frac{|Mx|}{|x|}\frac{|x|}{\sqrt{n}} \le \big(\max(\frac{|Mx|}{|x|}, \frac{|x|}{\sqrt{n}})\big)^2$ with a better estimate $\frac{|Mx|}{|x|}\frac{|x|}{\sqrt{n}} \le \frac14\big(\frac{|Mx|}{|x|} + \frac{|x|}{\sqrt{n}}\big)^2$. Moreover, it can be improved to 2.51 by replacing the Gaussian distribution of $x/\sqrt{n}$ with the uniform distribution on the unit sphere. However, the true threshold, 2, needs a different approach.

In fact, our proof shows that the convergence in Proposition 3b3 is exponentially fast. However, it could not be slow, because of the following concentration property.

3b9 Exercise. The distribution of $\sqrt{n}\|M\|$ is more concentrated than $\mathrm{N}(0,1)$.

Prove it.

Hint: $Z \mapsto \sqrt{n}\|Z\|$ is a 1-Lipschitz function w.r.t. the metric $Z \mapsto \sqrt{n}\|Z\|_{\mathrm{HS}}$, since $\|Z\| \le \|Z\|_{\mathrm{HS}}$ for all Z.

Some numerics, — sorted values of $\|M\|$ for samples of 5 matrices M:

n	$\|M\|$				
2	0.61	1.19	1.51	1.54	2.34
10	1.55	1.71	1.81	1.82	1.89
100	1.90	1.90	1.96	1.97	1.98
500	1.97	1.98	1.98	1.98	2.00

3c Orbit geometry: randomness disappears

By orbits I mean here such objects as the sequence $(x, Ax, A^2x, \dots)$, where A is a random symmetric matrix as in (3a2) and x is a unit vector. Another example is $(x, Mx, M^2x, \dots)$, where M is a random matrix as in (3a1). A more complicated case (for the same M) is the family of vectors Wx where W runs over arbitrary monomials (words) built from M and M^* (say, $W = MMM^*M^*M^*MM^*M$).

By geometry of the orbit $(x, Ax, A^2x, \dots)$ I mean scalar products $\langle A^k x, A^l x\rangle$ for all k, l (or just $\langle A^k x, x\rangle$, since $\langle A^k x, A^l x\rangle = \langle A^{k+l}x, x\rangle$). Similarly, geometry of the orbit $(x, Mx, M^2x, \dots)$ consists of $\langle M^k x, M^l x\rangle = \langle (M^*)^l M^k x, x\rangle$. Geometry of the (complicated) word-indexed orbit consists of the numbers $\langle Wx, x\rangle$. For any given dimension n the orbit geometry is random (its distribution does not depend on x due to $\mathrm{O}(n)$-invariance), but for large n the orbit geometry becomes nearly deterministic due to measure concentration, as we will see.

It is easy (but useless) to apply measure concentration to the linear function $Z \mapsto \langle Zx, x\rangle$. What about $Z \mapsto \langle Z^k x, x\rangle$? This is a polynomial of degree k, definitely not a Lipschitz function!

3c1 Exercise. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a measurable function and $B \subset \mathbb{R}^n$ a measurable set such that $|f(x) - f(y)| \le |x - y|$ for all $x, y \in B$. Then there exists $a \in \mathbb{R}$ such that

$$\gamma^n\{x \in \mathbb{R}^n : |f(x) - a| \ge c\} \le 2\Phi(-c) + \gamma^n(\mathbb{R}^n \setminus B)$$

for all $c \in (0, \infty)$.

Prove it.

Hint: there exists a 1-Lipschitz function $g : \mathbb{R}^n \to \mathbb{R}$ such that $g(x) = f(x)$ for all $x \in B$, for example, $g(x) = \sup_{y\in B}(f(y) - |x - y|)$.

We may choose $B = \{Z \in \mathrm{M}_n(\mathbb{R}) : \|Z\| \le 5\}$, then $\beta_n(\mathrm{M}_n(\mathbb{R}) \setminus B) \to 0$ as $n \to \infty$ by Proposition 3b3. The function (say) $Z \mapsto Z^3$ is continuously differentiable,

$$(Z + Y)^3 = Z^3 + Z^2Y + ZYZ + YZ^2 + o(\|Y\|)$$

as $\|Y\| \to 0$; we may estimate the gradient of the function $Z \mapsto \langle Z^3 x, x\rangle$ (for a given unit vector x):

$$\lim_{\varepsilon\to 0} \frac{\langle (Z+\varepsilon Y)^3 x, x\rangle - \langle Z^3 x, x\rangle}{\varepsilon} = \langle (Z^2 Y + ZYZ + YZ^2)x, x\rangle \le \\ \le \|Z^2 Y + ZYZ + YZ^2\| \le 3\|Z\|^2 \|Y\| \,.$$

For $Z \in B$ we get $75\|Y\| \le 75\|Y\|_{\mathrm{HS}}$, thus, the function $Z \mapsto \langle Z^3 x, x\rangle$ restricted to B is 75-Lipschitz w.r.t. $\|\cdot\|_{\mathrm{HS}}$, therefore $(75/\sqrt{n})$-Lipschitz w.r.t. $\sqrt{n}\|\cdot\|_{\mathrm{HS}}$; by 3c1 (and 3b3),

$$\beta_n\{Z \in \mathrm{M}_n(\mathbb{R}) : |\langle Z^3 x, x\rangle - r_{3,n}| \ge \varepsilon\} \to 0 \quad \text{as } n \to \infty$$

(for some numbers $r_{3,n}$), and the same for any $\langle Z^k x, x\rangle$. We see that

$$\text{(3c2)} \qquad \langle M^k x, x\rangle - r_{k,n} \to 0 \quad \text{in probability as } n \to \infty$$

for some $r_{k,n} \in \mathbb{R}$. The choice of unit vector $x \in \mathbb{R}^n$ does not matter (due to $\mathrm{O}(n)$-invariance). It is tempting to take

$$\text{(3c3)} \qquad r_{k,n} = \mathbb{E}\,\langle M^k x, x\rangle = \int_{\mathrm{M}_n(\mathbb{R})} \langle Z^k x, x\rangle \, \beta_n(\mathrm{d}Z)\,.$$

To this end, we need L_1-convergence rather than convergence in probability in (3c2). No doubt, the integral (3c3) exists for any k, l (indeed, in finite dimension, every polynomial is integrable w.r.t. every Gaussian measure); the problem is that $\int_{\mathrm{M}_n(\mathbb{R})\setminus Z} \langle Z^k x, x\rangle \, \beta_n(\mathrm{d}Z)$ could be large even though $\beta_n(\mathrm{M}_n(\mathbb{R}) \setminus B)$ is small. Fortunately, it is forbidden by 3b9 (together with 3b3); these imply

$$\sup_n \mathbb{E}\, \|M\|^p < \infty \quad \text{for all } p \in (0,\infty)\,.$$

3c4 Exercise. Fill in the gaps in the arguments above, thus proving that (3c2) holds for $r_{k,n}$ of (3c3) and moreover, the convergence in (3c2) holds in all L_p ($p < \infty$).

3c5 Exercise. Prove that

$$\frac{1}{n} \operatorname{trace}(M^k) - r_{k,n} \to 0 \quad \text{in probability as } n \to \infty\,.$$

Hint: all said above about the function $Z \mapsto \langle Z^k x, x\rangle$ holds also for the function $Z \mapsto \frac{1}{n}\big(\langle Z^k e_1, e_1\rangle + \dots + \langle Z^k e_n, e_n\rangle\big) = \frac{1}{n}\operatorname{trace}(Z^k)$.

The formula

$$r_{k,n} = \frac{1}{n}\mathbb{E}\,\operatorname{trace}(M^k) = \frac{1}{n}\int_{\mathrm{M}_n(\mathbb{R})} \operatorname{trace}(Z^k)\,\beta_n(\mathrm{d}Z) \tag{3c6}$$

is equivalent to (3c3), since $\operatorname{trace}(Z) = \langle Ze_1, e_1\rangle + \dots + \langle Ze_n, e_n\rangle$ for any Z. Moreover,

$$\mathbb{E}\,M^k = \int_{\mathrm{M}_n(\mathbb{R})} Z^k\,\beta_n(\mathrm{d}Z) = r_{k,n}\mathbf{1}_n\,, \tag{3c7}$$

since the $\mathrm{O}(n)$-invariant (that is, commuting with $\mathrm{O}(n)$) operator $\mathbb{E}\,M^k$ must be a scalar multiple of the identity operator $\mathbf{1}_n \in \mathrm{M}_n(\mathbb{R})$.

3d Orbit geometry via random rotations

Let M be a random matrix as in (3a1). The geometry of the orbit $(M^k x)_{k=0,1,\dots}$ remains intact if we replace M with $O^{-1}MO$ for any $O \in \mathrm{O}(n)$ such that $Ox = x$. For large n the geometry is nearly nonrandom, which means that one can choose O (dependent on M) such that the orbit $\big((O^{-1}MO)^k x\big)_k = (O^{-1}M^kOx)_k = (O^{-1}M^kx)_k$ is nearly nonrandom. We will do it via an explicit construction that reveals the orbit geometry.

3d1 Exercise. There exist measurable maps $O_n : \mathbb{R}^n \to \mathrm{O}(n)$ for $n = 1, 2, \dots$ such that

$$|x - O_n(x)e_1| \to 0 \quad \text{in probability as } n \to \infty, \text{ for } x \sim \gamma^n_{1/\sqrt{n}}\,.$$

Prove it.

Hint: try the rotation by the right angle in the plane spanned by e_1 and x; namely,

$$O_n(x)e_1 = x_1, \quad O_n(x)x_1 = -e_1,$$
$$O_n(x)y = y \quad \text{for all } y \text{ orthogonal to } e_1, x_1,$$

where $x_1 = (x - \langle x, e_1\rangle e_1)/|x - \langle x, e_1\rangle e_1|$. That is,

$$O_n(x)y = \langle y, e_1\rangle x_1 - \langle y, x_1\rangle e_1 + y - \langle y, e_1\rangle e_1 - \langle y, x_1\rangle x_1\,.$$

(Alternatively, one may use the rotation by the angle between e_1 and x, that is, $O_n(x)e_1 = x/|x|$.)

We denote $x_{2:n} = x - \langle x, e_1\rangle e_1 = \sum_{k=2}^n \langle x, e_k\rangle e_k$, $x_{3:n} = \sum_{k=3}^n \langle x, e_k\rangle e_k$ and so on. Also, we denote by $O_{2:n}(x_{2:n})$ the construction of 3d1 applied on the $(n-1)$-dimensional subspace $\mathbb{R}^{2:n} \subset \mathbb{R}^{1:n} = \mathbb{R}^n$ spanned by $e_2, \dots, e_n$. We have

$$O_{2:n}(x_{2:n})e_1 = e_1 \, , \qquad |x_{2:n} - O_{2:n}(x_{2:n})e_2| \to 0$$

despite the (small) distinction between $\gamma^{n-1}_{1/\sqrt{n}}$ and $\gamma^{n-1}_{1/\sqrt{n-1}}$. We use this rotation for transforming the random matrix M as follows:

$$\begin{gathered} O_1 = O_{2:n}\big((Me_1)_{2:n}\big) \, , \\ M_1 = O_1^{-1} M O_1 \, . \end{gathered}$$

We have $|O_1 e_2 - (Me_1)_{2:n}| \to 0$, which may be written as $O_1 e_2 = (Me_1)_{2:n} + o(1)$. However, $(Me_1)_{2:n} = Me_1 + o(1)$, thus $O_1 e_2 = Me_1 + o(1)$, therefore $O_1^{-1} Me_1 = e_2 + o(1)$, and so,

$$M_1 e_1 = O_1^{-1} M e_1 = e_2 + o(1) \, .$$

The first column of the random matrix M_1 typically is close to $(0, 1, 0, 0, \dots, 0)$, which shows that the distribution of M_1 differs from β_n. However, the distinction affects the first column only, as we will see soon.

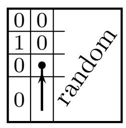
Denote by $M_{2:n}$ columns $2 \dots n$ of M; that is, $M_{2:n} e_1 = 0$ and $M_{2:n} e_k = Me_k$ for $k = 2, \dots, n$. The distribution $\beta_{2:n}$ of $M_{2:n}$ is invariant under $Z \mapsto OZ$ for all $O \in \mathrm{O}(n)$ and under $Z \mapsto ZO$ for all $O \in \mathrm{O}(2:n) = \{O \in \mathrm{O}(n) : Oe_1 = e_1\}$, therefore, under $Z \mapsto O^{-1} Z O$ for $O \in \mathrm{O}(2:n)$. Note that $O_1 \in \mathrm{O}(2:n)$ depends only on the first column $M_{1:1}$ of M (independent of $M_{2:n}$). For every bounded continuous (or just measurable) function $f_n : \mathrm{M}_n(\mathbb{R}) \to \mathbb{R}$,

$$\begin{aligned} \int f\big((M_1)_{2:n}\big) \, \beta_n(\mathrm{d}M) &= \int \beta_{1:1}(\mathrm{d}M_{1:1}) \int \beta_{2:n}(\mathrm{d}M_{2:n}) f\big(\underbrace{(O_1^{-1} M O_1)_{2:n}}_{=O_1^{-1} M_{2:n} O_1}\big) = \\ &= \int \beta_{1:1}(\mathrm{d}M_{1:1}) \int \beta_{2:n}(\mathrm{d}M_{2:n}) f(M_{2:n}) = \int f \, \mathrm{d}\beta_{2:n} \, , \end{aligned}$$

which means that $(M_1)_{2:n}$ is distributed $\beta_{2:n}$.

We continue the process recursively,

$$\begin{gathered} M_0 = M \, , \\ O_k = O_{k+1:n}\big((M_{k-1} e_k)_{k+1:n}\big) \, , \\ M_k = O_k^{-1} M_{k-1} O_k \end{gathered}$$



for $k = 1, 2, \dots$ and prove by induction that

$$\begin{gathered} M_k e_k = e_{k+1} + o(1) \, , \\ (M_k)_{k+1:n} \text{ is distributed } \beta_{k+1:n}. \end{gathered}$$

Also, $M_{k+1}e_k = e_{k+1} + o(1)$, $M_{k+2}e_k = e_{k+1} + o(1)$ and so on; indeed, e_k and e_{k+1} are invariant under $O_{k+1}, O_{k+1}^{-1}, O_{k+2}, O_{k+2}^{-1}, \dots$

The relations $M_k e_1 = e_2 + o(1)$, $M_k e_2 = e_3 + o(1)$, …, $M_k e_k = e_{k+1} + o(1)$ imply $M_k^k e_1 = e_{k+1} + o(1)$. Therefore $\langle M_k^k e_1, e_1 \rangle \to 0$, that is, $\langle O^{-1} M^k O e_1, e_1 \rangle \to 0$ where $O = O_1 \dots O_k$ satisfies $Oe_1 = e_1$; we get

$$\langle M^k e_1, e_1 \rangle \to 0 \quad \text{in probability as } n \to \infty$$

for $k = 1, 2, \dots$ In terms of the numbers $r_{k,n}$ (recall (3c2)–(3c7)) it means that

$$r_{k,n} \to 0 \quad \text{as } n \to \infty, \text{ for } k = 1, 2, \dots$$

Similarly, $\langle M^k e_1, M^l e_1 \rangle \to 0$ for $k \ne l$ (and 1 for $k = l$); the limiting orbit geometry describes just an orthonormal sequence. (Still, for now we have no information about $\langle Wx, x \rangle$ in general, where W is a word build from M and M^*.)

Some numerics:

n	2	10	100	500
$\langle Me_1, e_1 \rangle$	0.20	−0.05	0.08	0.03
$\lvert Me_1 \rvert^2 = \langle M^* M e_1, e_1 \rangle$	0.38	1.19	1.00	0.95
$\langle M^2 e_1, e_1 \rangle$	−0.42	−0.01	−0.03	0.02
$\lvert M^2 e_1 \rvert^2 = \langle M^{*2} M^2 e_1, e_1 \rangle$	0.26	0.92	1.11	0.95
$\langle M^2 e_1, M e_1 \rangle = \langle M^* M^2 e_1, e_1 \rangle$	0.09	−0.37	0.02	0.01

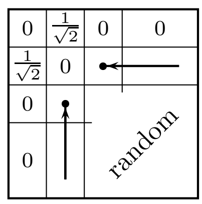
3e Wigner's semi-circle law

Let A be a symmetric random matrix as in (3a2). We may use the same rotation $O_1 = O_{2:n}\big((Ae_1)_{2:n}\big)$ as in 3d and get $A_1 = O_1^{-1} A O_1$ such that $A_1 e_1 = e_2 + o(1)$. However, the distribution of $(A_1)_{2:n}$ differs from the distribution of $A_{2:n}$. Indeed, $A_1 = O_1^* A O_1$ is symmetric; its first row, being equal to its first column, is close to $(0, 1, 0, 0, \dots, 0)$.

Consider the matrix $A_{2:n,2:n}$ consisting of $a_{k,l}$ for $k > 1, l > 1$. Its distribution $\beta_{2:n,2:n}$ is invariant under $Z \mapsto ZO$ and $Z \mapsto OZ$ for all $O \in \mathrm{O}(2\!:\!n)$. Similarly to 3d we conclude that $(A_1)_{2:n,2:n}$ is distributed $\beta_{2:n,2:n}$.

Similarly to 3d we continue the process,

$$\begin{gathered} A_0 = A\,, \\ O_k = O_{k+1:n}\big((A_{k-1}e_k)_{k+1:n}\big)\,, \\ A_k = O_k^{-1} A_{k-1} O_k \end{gathered}$$



and prove that

$$A_k^* = A_k \,,$$
$$\sqrt{2} A_k e_k = e_{k-1} + e_{k+1} + o(1) \,, \qquad (e_0 = 0)$$
$$(A_k)_{k+1:n,k+1:n} \text{ is distributed } \beta_{k+1:n,k+1:n} \,;$$

as before, $\sqrt{2}A_{k+1}e_k = e_{k-1} + e_{k+1} + o(1)$, $\sqrt{2}A_{k+2}e_k = e_{k-1} + e_{k+1} + o(1)$ and so on. The limiting (as $n \to \infty$) geometry of the random orbit $\big((\sqrt{2}A)^k e_1\big)_k$ is the geometry of the nonrandom orbit $\big((S^* + S)^k e_1\big)_k$, where S is the one-sided shift operator in $l_2 = \{(x_1, x_2, \dots) : \sum x_k^2 < \infty\}$, that is, $Se_k = e_{k+1}$ for $k = 1, 2, \dots$; of course, $S^* e_k = e_{k-1}$ for $k = 2, 3, \dots$ and $S^* e_1 = 0$. Thus, for example, $(S^* + S)^2 e_1 = e_1 + e_3$ and $(S^* + S)^3 e_1 = 2e_2 + e_4$. In terms of the numbers $r_{k,n}$ (as in (3c2)–(3c7) but for A instead of M) we have

(3e1) $$r_{k,n} \to r_k = 2^{-k/2} \langle (S^* + S)^k e_1, e_1 \rangle \quad \text{as } n \to \infty \,.$$

In fact, $r_{2k-1} = 0$ and $2^k r_{2k} = \binom{2k}{k} - \binom{2k}{k-1} = \frac{(2k)!}{k!(k+1)!}$, but we do not need it now.

Instead of S we may use the two-sided shift operator T on the space $l_2(\mathbb{Z})$ of all two-sided sequences $(x_k)_{k \in \mathbb{Z}}$ such that $\sum_{-\infty}^{+\infty} x_k^2 < \infty$. That is, $Te_k = e_{k+1}$ and $T^* e_k = e_{k-1}$ for all $k \in \mathbb{Z}$.

3e2 Exercise. The orbit $\big((T^* + T)^k (e_1 - e_{-1})/\sqrt{2}\big)_k$ in $l_2(\mathbb{Z})$ has the same geometry as the orbit $\big((S^* + S)^k e_1\big)_k$ in l_2.

Prove it.

Hint: $\langle (T^* + T)^k (e_1 - e_{-1})/\sqrt{2}, e_{\pm l} \rangle = \pm \frac{1}{\sqrt{2}} \langle (S^* + S)^k e_1, e_l \rangle$ for $l = 1, 2, \dots$ (and for $l = 0$ the left-hand side vanishes).

Combining (3e1) and 3c5 (for A instead of M) we get

(3e3) $$\frac{1}{n} \operatorname{trace}(A^k) \to r_k \quad \text{in probability as } n \to \infty$$

for $k = 0, 1, 2, \dots$ However, $\frac{1}{n} \operatorname{trace}(A^k) = \frac{1}{n}(\lambda_1^k + \dots + \lambda_n^k)$ where $(\lambda_1, \dots, \lambda_n)$ is the spectrum of A (recall 2a). Thus,

(3e4) $$\frac{1}{n}(\lambda_1^k + \dots + \lambda_n^k) \to r_k \quad \text{in probability as } n \to \infty$$

for $k = 0, 1, 2, \dots$ We may hope that the spectrum converges (in probability as $n \to \infty$) to a (nonrandom) distribution μ whose moments are equal to r_k. First of all we want to find such a distribution.

We need $\int \lambda^k \, \mu(\mathrm{d}\lambda) = r_k$, that is, $\langle \Lambda^k \mathbf{1}, \mathbf{1} \rangle = r_k$, where Λ is the multiplication operator, $(\Lambda f)(\lambda) = \lambda f(\lambda)$, in the space $L_2(\mu)$, and $\mathbf{1} \in L_2(\mu)$,

$\mathbf{1}(\lambda) = 1$ for all λ. In other words, the geometry of the orbit $(\Lambda^k \mathbf{1})_k$ in $L_2(\mu)$ should be (described by r_k, therefore) the same as the geometry of the orbit $\Big(\big(\frac{T^*+T}{\sqrt{2}}\big)^k \frac{e_1 - e_{-1}}{\sqrt{2}} \Big)_k$ in $l_2(\mathbb{Z})$.

The shift operator T is diagonalized by Fourier transform $\mathcal{F} : l_2(\mathbb{Z}) \to L_2\big((0,2\pi), \frac{\mathrm{mes}}{2\pi}\big)$,

$$\mathcal{F}(x)(u) = \sum_{k\in\mathbb{Z}} \langle x, e_k \rangle \mathrm{e}^{\mathrm{i}ku} \,, \qquad \| \mathcal{F}(x) \| = \|x\| \,,$$
$$\mathcal{F}(Tx)(u) = \mathrm{e}^{\mathrm{i}u} \mathcal{F}(x)(u) \,,$$
$$\mathcal{F}\big(\tfrac{T^*+T}{\sqrt{2}} x \big)(u) = \sqrt{2} \cos u \cdot \mathcal{F}(x)(u) \,.$$

Taking into account that $\mathcal{F}\big(\frac{e_1 - e_{-1}}{\sqrt{2}}\big)(u) = \sqrt{2}\, \mathrm{i} \sin u$ we see that the orbit $\Big(\big(\frac{T^*+T}{\sqrt{2}}\big)^k \frac{e_1 - e_{-1}}{\sqrt{2}} \Big)_k$ in $l_2(\mathbb{Z})$ is isometric to the orbit $\big((\sqrt{2}\cos u)^k \sqrt{2}\, \mathrm{i} \sin u \big)_k$ in $L_2\big((0,2\pi), \frac{\mathrm{mes}}{2\pi}\big)$, that is,

$$r_k = \frac{1}{2\pi} \int_0^{2\pi} (\sqrt{2}\cos u)^k |\sqrt{2}\, \mathrm{i} \sin u|^2 \, \mathrm{d}u \,.$$

3e5 Exercise. Prove that

$$r_k = \frac{1}{\pi} \int_{-\sqrt{2}}^{+\sqrt{2}} \lambda^k \sqrt{2 - \lambda^2} \, \mathrm{d}\lambda$$

for $k = 0, 1, 2, \dots$

Hint: $\lambda = \sqrt{2} \cos u$.

The measure μ is found,

$$\mu(\mathrm{d}\lambda) = \frac{1}{\pi} \sqrt{(2-\lambda^2)^+} \, \mathrm{d}\lambda = \begin{cases} \frac{1}{\pi} \sqrt{2-\lambda^2} \, \mathrm{d}\lambda & \text{for } -\sqrt{2} < \lambda < \sqrt{2}, \\ 0 & \text{otherwise;} \end{cases}$$

it is the well-known Wigner's semi-circle law on $(-\sqrt{2}, \sqrt{2})$. It has a compact support (as it should in the light of 3b3). Therefore it is uniquely determined by its moments r_k,[1] and for any sequence of distributions $\mu_1, \mu_2, \dots$, convergence of their moments to r_k ensures $\mu_n \to \mu$.[2] The latter may be written as

$$\sup_{\lambda\in\mathbb{R}} | \mu_n((-\infty,\lambda]) - \mu((-\infty,\lambda]) | \to 0 \quad \text{as } n \to \infty \,,$$

since the measure μ is nonatomic.[3]

[1] See for instance [1], XV.4 (Appendix).
[2] See for instance [1], VIII.6 (Example b).
[3] For the general case see [1], VIII.10 (Problem 11).

3e6 Theorem.

$$\sup_{a\in\mathbb{R}} \left| \frac{1}{n}\#\{k : \lambda_k \le a\} - \frac{1}{\pi}\int_{-\infty}^{a} \sqrt{(2-\lambda^2)^+}\,\mathrm{d}\lambda \right| \to 0 \quad \text{in probability as } n \to \infty\,,$$

where $(\lambda_1, \dots, \lambda_n)$ is the spectrum of a Gaussian random matrix distributed according to (3a2).

3e7 Exercise. Prove the theorem.

Hint: for any ε there exist k and δ such that δ-closeness of the first k moments to $r_1, \dots, r_k$ ensures ε-closeness between the cumulative distribution functions; use (3e4).

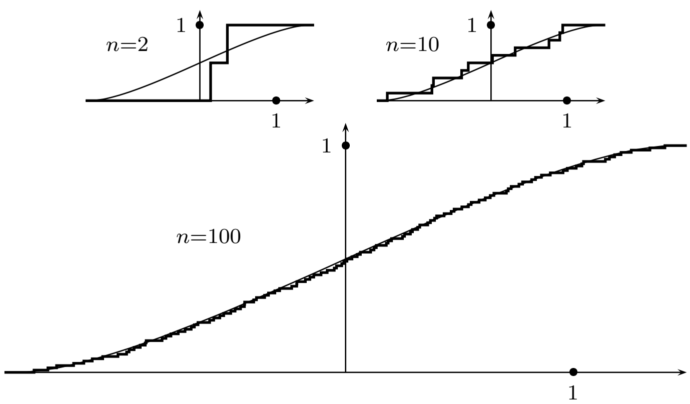
3e8 Exercise. Prove that

$$\mathbb{P}\big(\|A\| \ge \sqrt{2} - \varepsilon\big) \to 1 \quad \text{as } n \to \infty\,.$$

Hint: the norm of a *symmetric* matrix is equal to the highest eigenvalue; use Theorem 3e6.

Do not think that $\mathbb{P}\big(\|A\| \le \sqrt{2} + \varepsilon\big) \to 1$ by Theorem 3e6; it does not follow! The distribution of $\|A\|$ is concentrated near some point (similarly to 3b9), and the point can be chosen from $[\sqrt{2}, 4]$ according to 3e8 and 3b3. In fact, it can be chosen at $\sqrt{2}$, and moreover, the limiting distribution of $n^{2/3}(\|A\| - \sqrt{2})$ exists.[1]

Some numerics. The two cumulative distribution functions, empirical $a \mapsto \frac{1}{n}\#\{k : \lambda_k \le a\}$ and theoretical $a \mapsto \frac{1}{\pi}\int_{-\infty}^{a} \sqrt{(2-\lambda^2)^+}\,\mathrm{d}\lambda = \frac{1}{2} + \frac{1}{\pi}\arcsin\frac{\lambda}{\sqrt{2}} + \frac{1}{2\pi}\lambda\sqrt{2-\lambda^2}$ for $|\lambda| < \sqrt{2}$.



[1]The limiting distribution of $\sqrt{2}n^{2/3}(\|A\| - \sqrt{2})$ is the Tracy-Widom law of order 1, see C.A. Tracy and H. Widom, "Distribution functions for largest eigenvalues and their applications", Proceedings of the International Congress of Mathematicians (2002), Vol. 1, 587–596. arXiv:math-ph/0210034.

References

My approach to Wigner's semi-circle law shaped under the influence of Daniel Slutsky who in turn was influenced by the course "Random Matrices" given in Hebrew University by Andrzej Szankowski who used the paper [3].

[1] W. Feller, *An introduction to probability theory and its applications,* second edition, vol. 2, Wiley, 1971.

[2] A. Guionnet, *Large deviations and stochastic calculus for large random matrices,* Probability Surveys **1**, 72–172 (2004).

[3] H.F. Trotter, *Eigenvalue distributions of large Hermitian matrices; Wigner's semi-circle law and a theorem of Kac, Murdock and Szegö,* Advances in Math. **54**, 67–82 (1984).

Index