Random matrix theory - an introduction

February 21, 2024

1 Matrices and their spectra

1.1 Linear operators and their matrices

Let V be a finite-dimensional linear space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. If $\{e(1), \ldots, e(n)\}$ is a basis for V and $\phi \in V$, then there exist unique $\phi_1, \ldots, \phi_n \in \mathbb{K}$ such that

$$\phi = \sum_{i=1}^{n} \phi_i e(i).$$

Let V, W be finite-dimensional linear spaces over $\mathbb{K}$ equipped with bases $\{e(1), \ldots, e(n)\}$ and $\{f(1), \ldots, f(m)\}$. Let $\mathcal{L}(V, W)$ be the space of linear operators $A : V \to W$. Then for each $A \in \mathcal{L}(V, W)$ there exist unique $A_{ij} \in \mathbb{K}$ with $1 \leq i \leq m$ and $1 \leq j \leq n$ such that

$$(A\phi)_i = \sum_{j=1}^{n} A_{ij}\phi_j \qquad (1 \leq i \leq m).$$

We call $(A_{ij})_{1 \leq i \leq m,\ 1 \leq j \leq n}$ the *matrix* of A and we call A_{ij} the (i, j)-th *entry* of the matrix. One has

$$(AB)_{ik} = \sum_{j} A_{ij} B_{jk}.$$

The *trace* of an operator $A \in \mathcal{L}(V) := \mathcal{L}(V, V)$

$$\mathrm{tr}(A) := \sum_{i} A_{ii}$$

does not depend on the choice of the basis and satisfies $\mathrm{tr}(AB) = \mathrm{tr}(BA)$. Let $A \in \mathcal{L}(V)$. By definition, $0 \neq \phi \in V$ is an *eigenvector* with *eigenvalue* $\lambda \in \mathbb{K}$ if $A\phi = \lambda\phi$. We call

$$\sigma(A) := \big\{\lambda : \lambda \text{ is an eigenvalue of } A\big\}$$

the *spectrum* of A. One has

$$\sigma(A) := \big\{\lambda : (\lambda - A) \text{ is not invertible}\big\}.$$

Lemma 1 (Nonempty spectrum) *Assume that $A \in \mathcal{L}(V)$ and $\mathbb{K} = \mathbb{C}$. Then $\sigma(A) \neq \emptyset$.*

Proof (sketch) One has $\sigma(A) := \big\{\lambda : \det(\lambda - A) = 0\big\}$. The equation $\det(\lambda - A) = 0$ is a polynomial in λ of degree $\dim(V)$ which is guaranteed to have at least one complex root. ∎

1.2 Inner product spaces

Let V be a finite-dimensional linear space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. An *inner product* on V is a map that assigns to two vectors $\phi, \psi \in V$ a number $\langle \phi, \psi \rangle \in \mathbb{K}$ such that

(i) $\psi \mapsto \langle \phi, \psi \rangle$ is linear,

(ii) $\langle \phi, \psi \rangle = \overline{\langle \psi, \phi \rangle}$,

(iii) $\langle \phi, \phi \rangle \geq 0$,

(iv) $\langle \phi, \phi \rangle = 0 \quad \Rightarrow \quad \phi = 0$.

Here $\bar{c}$ denotes the complex conjugate of a number $c \in \mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. In the complex case, $\phi \mapsto \langle \phi, \psi \rangle$ is not linear but *colinear* in the sense that

$$\langle c_1 \phi(1) + c_2 \phi(2), \psi \rangle = \bar{c}_1 \langle \phi(1), \psi \rangle + \bar{c}_2 \langle \phi(2), \psi \rangle.$$

The norm associated with the inner product is $|\phi| := \sqrt{\langle \phi, \phi \rangle}$. A basis $\{e(1), \ldots, e(n)\}$ is *orthogonal* if $\langle e(i), e(j) \rangle = 0$ for $i \neq j$ and *orthonormal* if in addition $\langle e(i), e(i) \rangle = 1$ for each i. For each $\phi \in V$ we define $\langle \phi | \in \mathcal{L}(V, \mathbb{K})$ and $|\phi\rangle \in \mathcal{L}(\mathbb{K}, V)$ by

$$\langle \phi | \psi := \langle \phi, \psi \rangle \quad \text{and} \quad |\phi\rangle c := c\phi.$$

The space $V' := \mathcal{L}(V, \mathbb{K})$ is called the *dual linear space* of V. On the other hand, $\mathcal{L}(\mathbb{K}, V)$ can naturally be identified with V itself. Also, $\langle \phi | \, | \psi \rangle \in \mathcal{L}(\mathbb{K}, \mathbb{K}) \cong \mathbb{K}$ can be identified with the number $\langle \phi, \psi \rangle$. For $A \in \mathcal{L}(V)$ one has

$$A|\psi\rangle = |A\psi\rangle.$$

The coordinates of a vector and operator with respect to an *orthonormal* basis are given by

$$\phi_i = \langle e(i), \phi \rangle \quad \text{and} \quad A_{ij} = \langle e(i), Ae(j) \rangle = \langle e(i) | A | e(j) \rangle.$$

Note that $|\phi\rangle\langle\psi| \in \mathcal{L}(V, V) = \mathcal{L}(V)$. One has

$$A = \sum_{i,j} A_{ij} |e(i)\rangle\langle e(j)|,$$

and the operators $|e(i)\rangle\langle e(j)|$ form a basis for $\mathcal{L}(V)$. In particular

$$1 = \sum_i |e(i)\rangle\langle e(i)|$$

is the identity operator.

Each $A \in \mathcal{L}(V)$ has a unique *adjoint* $A^* \in \mathcal{L}(V)$ such that

$$\langle \phi, A\psi \rangle = \langle A^* \phi, \psi \rangle \qquad (\phi, \psi \in V).$$

the map $A \mapsto A^*$ is colinear with

$$(A^*)^* = A \quad \text{and} \quad (AB)^* = B^* A^*.$$

In coordinates

$$A^*_{ij} = \overline{A_{ji}}.$$

An operator A is *normal* if it commutes with its adjoint:

$$AA^* = A^* A.$$

We say that A is *hermitian* if $A^* = A$.

Theorem 2 (Diagonalisation of normal operators) *Assume that* $\mathbb{K} = \mathbb{C}$. *Then* $A \in \mathcal{L}(V)$ *is normal if and only if there exists an orthonormal basis* $\{e(1), \ldots, e(n)\}$ *and complex numbers* $\lambda_1, \ldots, \lambda_n$ *such that*

$$A = \sum_i \lambda_i |e(i)\rangle\langle e(i)|. \tag{1}$$

Moreover, A *is hermitian if and only if* $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$.

Proof It is straightforward to check an operator of the form (1) is normal, and hermitian if and only if $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. It remains to show each normal operator can be written in the form (1). Assume A is normal. Then for each $\phi \in V$,

$$\langle A^*\phi, A^*\phi\rangle = \langle \phi, AA^*\phi\rangle = \langle \phi, A^*A\phi\rangle = \langle A\phi, A\phi\rangle,$$

which shows that

$$|A^*\phi| = |A\phi|.$$

Since $\mathbb{K} = \mathbb{C}$, the operator A has at least one eigenvector ϕ with some eigenvalue λ. Then

$$A\phi = \lambda\phi \quad \Rightarrow \quad |(A-\lambda)\phi| = 0 \quad \Rightarrow \quad |(A-\lambda)^*\phi| = 0 \quad \Rightarrow \quad A^*\phi = \lambda^*\phi.$$

Let $\{\phi\}^\perp := \{\psi \in V : \langle \phi, \psi\rangle = 0\}$. Then

$$\psi \in \{\phi\}^\perp \quad \Rightarrow \quad \langle \phi, A\psi\rangle = \langle A^*\phi, \psi\rangle = \lambda\langle \phi, \psi\rangle = 0 \quad \Rightarrow \quad A\psi \in \{\phi\}^\perp.$$

Now A restricted to $\{\phi\}^\perp$, is again a normal operator so repeating the argument we can find an orthonormal basis of eigenvectors. ∎

Although our proof of Theorem 2 used the complex numbers in an essential way, for *symmetric* real matrices one can prove something similar.

Theorem 3 (Diagonalisation of symmetric matrices) *Let* V *be a real vector space and let* $A \in \mathcal{L}(V)$ *satisfy* $A^* = A$. *Then there exists an orthonormal basis* $\{e(1), \ldots, e(n)\}$ *consisting of eigenvectors of* A.

Proof This will follow from the same arguments as in the proof of Theorem 2 provided we show that each symmetric real matrix has at least one eigenvector. By compactness, the function $\phi \mapsto \langle \phi, A\phi\rangle$ assumes its maximum over the ball surface $\{\phi \in V : |\phi| = 1\}$ in some point ψ. This means that in the point ψ the derivatives of the function $\phi \mapsto \langle \phi, A\phi\rangle$ in directions tangential to the ball are all zero. By the method of Lagrange multipliers, there exists a $\lambda \in \mathbb{R}$ such that derivatives of the function $\phi \mapsto \langle \phi, A\phi\rangle - \lambda\langle \phi, \phi\rangle$ in the point ψ are zero in *all* directions. Thus, for our ψ, we can find $\lambda \in \mathbb{R}$ so that

$$\tfrac{\partial}{\partial\varepsilon}\big[\langle \psi + \varepsilon\phi, A(\psi + \varepsilon\phi)\rangle - \lambda\langle \psi + \varepsilon\phi, \psi + \varepsilon\phi\rangle\big]\Big|_{\varepsilon=0} = 0 \qquad (\phi \in V).$$

This gives

$$\langle \psi, A\phi\rangle + \langle \phi, A\psi\rangle - \lambda\langle \psi, \phi\rangle - \lambda\langle \phi, \psi\rangle = 0 \qquad (\phi \in V).$$

Using the fact that $A^* = A$ and dividing out a factor 2, we get

$$\langle \phi, A\psi\rangle - \lambda\langle \phi, \psi\rangle = 0 \qquad (\phi \in V).$$

But this says that $\langle \phi, A\psi - \lambda\psi\rangle = 0$ for all $\phi \in V$, which means that $A\psi = \lambda\psi$, i.e., we have found an eigenvector. ∎

1.3 The operator norm

Let V be a finite-dimensional linear space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. The *operator norm* of an operator $A \in \mathcal{L}(V)$ is defined as

$$\|A\|_{\mathrm{op}} := \sup_{|\phi| \leq 1} |A\phi|.$$

By linearity, $\|A\|_{\mathrm{op}}$ is the least constant for which that the inequality

$$|A\phi| \leq \|A\|_{\mathrm{op}} \cdot |\phi|$$

holds for all $\phi \in V$. From this, it is easy to see that

$$\|AB\|_{\mathrm{op}} \leq \|A\|_{\mathrm{op}} \cdot \|B\|_{\mathrm{op}}.$$

Lemma 4 (Operator norm and spectrum) *If A is normal and $\mathbb{K} = \mathbb{C}$, then*

$$\|A\|_{\mathrm{op}} = \sup\big\{|\lambda| : \lambda \in \sigma(A)\big\}. \tag{2}$$

Proof Let $C := \sup\big\{|\lambda| : \lambda \in \sigma(A)\big\}$. If ϕ is an eigenvector with eigenvalue λ, then

$$|A\phi| = |\lambda| \cdot |\phi|,$$

so $|\lambda| \leq \|A\|_{\mathrm{op}}$ for each $\lambda \in \sigma(A)$, which implies $\|A\|_{\mathrm{op}} \geq C$. To prove the other inequality, we use that by Theorem 2 there exists an orthonormal basis $\{e(1), \ldots, e(n)\}$ of eigenvectors. Denoting the corresponding eigenvalues by $\lambda_1, \ldots, \lambda_n$, we can write for arbitrary $\phi \in V$,

$$|A\phi|^2 = \Big|\sum_{i=1}^{n} \lambda_i \phi_i e(i)\Big|^2 = \sum_{ij} \overline{\lambda}_i \overline{\phi}_i \lambda_j \phi_j \langle e(i), e(j)\rangle = \sum_i |\lambda_i|^2 \cdot |\phi_i|^2 \leq C^2 \sum_i |\phi_i|^2 = C^2 |\phi|^2.$$

It follows that $|A\phi| \leq C|\phi|$ for all ϕ and hence $\|A\|_{\mathrm{op}} \leq C$. ∎

Question Is the assumption in Lemma 4 that A is normal needed? Note that the definition of $\sigma(A)$ does not depend on the choice of the inner product on V, so if (2) would hold for all $A \in \mathcal{L}(V)$, then this would mean that the definition of the operator norm does not depend on the choice of the inner product on V. This seems strange. Can someone find a counterexample to show that (2) may fail if A is not normal?

2 Random matrices

2.1 Matrix ensembles

The central question we will be interested in is the following:

> **Question** Let M be a random matrix of size $n \times n$. What can we say about its spectrum as $n \to \infty$?

To make this question more precise, we must say what we mean with a "random matrix", i.e., we must describe its law, and we must also be more specific about what we want to know about its spectrum. In random matrix theory, a sequence of probability laws on the spaces of $n \times n$ real or complex matrices is called a *matrix ensemble*. There are several natural choices. In what follows, for each integer $n \geq 1$, we consider a random matrix $M = (\xi_{ij})_{1 \leq i,j \leq n}$.

- **I.i.d. matrix ensembles** These are ensembles where the entries are i.i.d. according to some common law. Examples are:
  - *Bernoulli ensemble* ξ_{ij} uniformly distributed on $\{-1, 1\}$,

- *Real Gaussian ensemble* ξ_{ij} standard normally distributed on $\mathbb{R}$,
- *Complex Gaussian ensemble* ξ_{ij} standard normally distributed on $\mathbb{C} \cong \mathbb{R}^2$.

- **Symmetric Wigner matrix ensembles** These are ensembles where $(\xi_{ij})_{i \geq j}$ are independent real random variables and $\xi_{ij} := \xi_{ji}$ for $i > j$. Examples are:
  - *Symmetric Bernoulli ensemble* $(\xi_{ij})_{i \leq j}$ uniformly distributed on $\{-1, 1\}$,
  - *Gaussian orthogonal ensemble (GOE)* $(\xi_{ij})_{i \leq j}$ standard normally distributed on $\mathbb{R}$, and $(\xi_{ii})_{1 \leq i \leq n}$ normally distributed with mean zero and variance 2.

- **Hermitian Wigner matrix ensembles** These are ensembles where $(\xi_{ij})_{i<j}$ are independent complex random variables, $(\xi_{ii})_{1 \leq i \leq n}$ are independent real random variables, and $\xi_{ij} := \overline{\xi_{ji}}$ for $i > j$. The symmetric Wigner matrix ensembles are special cases of this. An important additional example is:
  - *Gaussian unitary ensemble (GUE)* $(\xi_{ij})_{i \leq j}$ standard normally distributed on $\mathbb{C} \cong \mathbb{R}^2$, and $(\xi_{ii})_{1 \leq i \leq n}$ standard normally distributed on $\mathbb{R}$.

Note that in all these examples, we can start with an infinite matrix

$$M = (\xi_{ij})_{i,j \in \mathbb{N}_+},$$

and then define $M_n := (\xi_{ij})_{1 \leq i,j \leq n}$. This provides a natural coupling between matrixes of different size.

Question If V is an inner product space and $\{e(1), \ldots, e(n)\}$ is an orthonormal basis, then each of the "ensembles" above naturally defines a probability law on $\mathcal{L}(V)$, when we identify a linear operator with its matrix. For which of the probability distributions above (if any) is this law independent of the choice of the orthonormal basis? I suspect this may be true for GOE and GUE (and is probably known) but I haven't found this yet. Note that in general (for example for the Bernoulli ensembles), if we change the basis, then for the new matrix it will not even be true that the (upper diagional) matrix entries are independent.

2.2 Limit behaviour of the spectrum

One of the highlights of the course is the proof of Theorem 2.4.2 in the book, which reads:

> **Wigner semicircle law** Let $M = (\xi_{ij})_{i,j \in \mathbb{N}_+}$ be an infinite hermitian Wigner matrix. For each $n \geq 1$, let $M_n := (\xi_{ij})_{1 \leq i,j \leq n}$ and let $\lambda_1(M_n) \leq \cdots \leq \lambda_n(M_n)$ denote its eigenvalues. Let
>
> $$\mu_n := \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j(M_n)/\sqrt{n}} \qquad (n \geq 1).$$
>
> Then almost surely $\mu_n \Rightarrow \mu$, where $\Rightarrow$ denotes weak convergence of probability measures on $\mathbb{R}$ and μ is the *semicircle law*
>
> $$\mu(\mathrm{d}x) := \frac{1}{2\pi} \sqrt{0 \vee (4 - |x|^2)}\, \mathrm{d}x.$$

Note that we are able to formulate this as almost sure convergence only due to the coupling described above. Of course, almost sure convergence to a deterministic limit also implies convergence in probability.

The theorem holds for the symmetric Bernoulli ensemble and the GOE and GUE ensembles, as well as many other hermitian Wigner matrix ensembles. Thus, similar to the central

limit theorem, the Wigner semicircle law is highly *universal*. For i.i.d. matrix ensembles the limit law is different, however: in this case the limit law is the uniform distribution on the ball $\{z \in \mathbb{C} : |z| \leq 2\}$. This limit law is technically more difficult to prove, so we will not see its proof in this course.

The Wigner semicircle law implies that for large n, with high probability, most of the eigenvalues of M_n lie between $-2\sqrt{n}$ and $2\sqrt{n}$. Since there are n eigenvalues, their average spacing is $4/\sqrt{n}$. There are other questions about the spectrum one could be interested in, such as:

- Let $x_n \in \mathbb{R}$ satisfy $x_n/\sqrt{n} \to x \in (-2, 2)$. Consider the point process

$$\big\{\lambda \in \mathbb{R} : x_n + \lambda/\sqrt{n} \in \sigma(M_n)\big\}.$$

  Does this point process have a limit (in law)?

- Same question as above, but for a suitable sequence x_n such that $x_n/\sqrt{n} \to 2$, perhaps with a different scaling of space to compensate for the lower density of points near the end of the spectrum.

- What can we say about $\|M_n\|_{\mathrm{op}}$, i.e., the largest eigenvalue in absolute value?. Is it true that $\|M_n\|_{\mathrm{op}}/\sqrt{n} \to 2$?

These questions have been studied in detail and a lot is known. We will only look at the last question. We will prove that indeed $\|M_n\|_{\mathrm{op}}$ is with high probability close to $2\sqrt{n}$. The methods used to prove this turn out to be very useful also later when we prove the Wigner semicircle law.

2.3 Why would we care?

The book we use is written by Terence Tao, who is a pure mathematician who has worked in a variety of fields such as number theory. For him, there is one clear reason for studying random matrices:

> It is a rich field with beautiful mathematics and many interesting problems.

There are also more practical reasons why one could be interested in random matrices. Some of the first people to study them were in fact physicists who were interested in the absorption spectrum of large atoms. One of the first great achievements of quantum mechanics was the description of the spectrum of the hydrogen atom, which has just one electron. The spectrum of helium, which has two electrons, is already much more complicated. As one moves up in the periodic system of elements things quickly get really messy. Mathematically, the possible energy values of an electron orbiting an atom are given by the spectrum of a hermitian matrix. This matrix becomes so complicated that at some point a physicist wondered what would happen if one assumes it is completely random -and actually got reasonable results.

Apart from the reasons mentioned above, one may also ask if within mathematics, random matrix theory is more or less isolated, or on the other hand connected to lots of other problems (for which one may then again have practical reasons to study them). The answer seems to lie a bit in the middle. On the one hand, interesting connections have been found to other subjects such as the totally asymetric exclusion process (TASEP) or free probability. On the other hand, it is fair to say that within probability theory, the theory of random matrices is a bit of an outlier. On conferences, one sometimes meets the true specialists in random matrix theory, that seem to live a bit in a world of their own, with limited interaction with the rest of probability. This is mostly concerned with very subtle questions of the field, however, which require difficult and long proofs using specialised methods. As for the basic topics covered by Tao's book, it is probably a good thing for every probabilist to know a bit about them.

2.4 How do we start

We will focus on hermitian Wigner ensembles. Our first aim will be to prove that $\|M_n\|_{\mathrm{op}}$ is with high probability close to $2\sqrt{n}$. As mentioned before, the techniques we use for that will also be useful when proving the Wigner semicircle law. Thus, perhaps surprisingly, the strategy is to first prove some sort of (weak) law of large numbers, as a first step towards something that is a bit like the CLT. Actually, there is much more similarity than might be expected at first sight between the classical limit theorems of probability theory (LLN, CLT) and their random matrix "counterparts" (LLN for $\|M_n\|_{\mathrm{op}}$, Wigner semicircle law).

Let $X_1, \ldots, X_n$ be i.i.d. real random variables with finite mean μ. Then the weak law of large numbers says that $\frac{1}{n}\sum_{i=1}^n X_i$ is with high probability close to μ. One can generalise the problem and ask what we can say about $F_n(X_1, \ldots, X_n)$, where $F_n : \mathbb{R}^n \to \mathbb{R}$ is a sequence of "nice" functions. For such functions, one often observes that for large n, the distribution of $F_n(X_1, \ldots, X_n)$ is closely concentrated around one deterministic value. If the F_n are not linear, then it may be difficult to say what that value is precisely. But it turns out that there are nice methods that give sufficient conditions for the law of $F_n(X_1, \ldots, X_n)$ to be closely concentrated around one deterministic value, *even if we can's say precisely what value that is.* This is the problem of *concentration of measure.* There exists a large literature on this and it obviously has many applications in probability theory outside of random matrix theory. Our first aim will be *Talagrand's concentration inequality* (Theorem 2.1.13 in the book).