Random Matrices (2017)

Introductory course on random matrices from *October 10 to November 23, 2017* at *IST Austria*. The course instructor is László Erdős and the teaching assistant is Dominik Schröder.

Description

Random matrices were first introduced in statistics in the 1920's, but they were made famous by Eugene Wigner's revolutionary vision. He predicted that spectral lines of heavy nuclei can be modelled by the eigenvalues of random symmetric matrices with independent entries (Wigner matrices). In particular, he conjectured that the statistics of energy gaps is given by a universal distribution that is independent of the detailed physical parameters. While the proof of this conjecture for realistic physical models is still beyond reach, it has recently been shown that the gap statistics of Wigner matrices is independent of the distribution of the matrix elements. Students will be introduced to the fascinating world of random matrices and presented with some of the basic tools for their mathematical analysis in this course.

Target audience

Students with orientation in mathematics, theoretical physics, statistics and computer science. No physics background is necessary. Calculus, linear algebra and some basic familiarity with probability theory is expected.

Evaluation

The final grade will be obtained as a combination of the student's performance on the example sheets and an oral exam.

Credits

3 ECTS

Lecture notes

Related notes from the recent PCMI summer school on random matrices.

Schedule

The course lasts from October 10 – November 23, 2017.

Day		Time		Room
Oct 12	*Thu*	`11.20am-12.35pm`	Lecture	Mondi 3
Oct 17	*Tue*	`10.15am-11.30am`	Lecture	Mondi 3
Oct 17	*Tue*	`11.45am-12.35pm`	Recitation	Mondi 3
Oct 18	**Wed**	`11.30am-12.45pm`	Lecture	**Mondi 1**
Oct 24	*Tue*	`10.15am-11.30am`	Lecture	Mondi 3
Oct 24	*Tue*	`11.45am-12.35pm`	Recitation	Mondi 3
Oct 25	**Wed**	`11.30am-12.45pm`	Lecture	**Mondi 1**
Nov 2	*Thu*	`11.20am-12.35pm`	Lecture	Mondi 3
Nov 7	*Tue*	`10.15am-11.30am`	Lecture	Mondi 3
Nov 7	*Tue*	`11.45am-12.35pm`	Recitation	Mondi 3
Nov 9	*Thu*	`11.20am-12.35pm`	Lecture	Mondi 3
Nov 14	*Tue*	`10.15am-11.30am`	Lecture	Mondi 3
Nov 14	*Tue*	`11.45am-12.35pm`	Recitation	Mondi 3
Nov 16	*Thu*	`11.20am-12.35pm`	Lecture	Mondi 3
Nov 21	*Tue*	`10.15am-11.30am`	Lecture	Mondi 3
Nov 21	*Tue*	`11.45am-12.35pm`	Recitation	Mondi 3
Nov 23	*Thu*	`11.20am-12.35pm`	Lecture	Mondi 3

Content

October 12

1. Basic facts from probability theory. Law of large numbers (LLN) and the central limit theorem (CLT), viewed as universality statements. In the LLN the limit is deterministic, while in CLT the limit is a random variable, namely the Gaussian (normal) one. No matter which distribution the initial random variables had, their appropriately normalized sums always converge to the same distribution — in other words the limiting Gaussian distribution is universal.

2. Wigner random matrices. Real symmetric and complex hermitian. GUE and GOE. Wishart matrices and their relation to Wigner-type matrices. Scaling so that eigenvalues remain bounded. Statement on the concentration of the largest eigenvalue. Introducing the semicircle law as a law of large numbers for the empirical density of the eigenvalues.

3. Linear statistics of eigenvalues (with a smooth function as observable) leads to CLT but with an unusual scaling — indicating very strong correlation among eigenvalues.

4. Statement of the gap universality, Wigner surmise. The limit behavior of the gap is a new universal distribution; in this sense this is the analogue of the CLT.

 Reading. PCMI lecture notes up to middle of Section 1.2.3.

October 17

1. Main questions in random matrix theory:

   - Density on global scale (like LLN)
   - Extreme eigenvalues (especially relevant for sample covariance matrices)
   - Fluctuation on the scale of eigenvalue spacing (like CLT)
   - Mesoscopic density — follows the global behaviour, but it is a non-trivial fact.
   - Eigenfunction (de)localization

2. Definition of k-point correlation functions. Relation of the gap distribution to the local correlation functions on scale of the eigenvalue spacing (inclusion-exclusion formula)

3. Rescaled (local) correlation functions. Determinant structure. Sine kernel for complex Hermitian Wigner matrices. Statement of the main universality result in the bulk spectrum (for energies away from the edges of the semicircle law).

   Reading. PCMI lecture notes up to the end of Section 1.2.3.

October 17 (Recitation)

1. Definition of Stieltjes transform
   $$ m_\mu(z):=\int_{\mathbb R} \frac{d\mu(\lambda)}{\lambda-z},\quad z\in\mathbb{H}:=\{z\in\mathbb C,\, \Im z>0\} $$ (eq-Stieltjes) of probability measure μ and statement of elementary properties (analyticity, trivial bounds on derivatives). Interpretation of the Stieltjes transform of the empirical spectral density as the normalized trace of the resolvent.

2. Interpretation of the imaginary part of **eq-Stieltjes** as the convolution with the Poisson kernel,
   $$ \Im m_\mu(x+i\eta)= \pi (P_\eta\ast \mu)(x),\quad P_\eta(x):=\frac{1}{\pi}\frac{\eta}{x^2+\eta^2}. $$ The Stieltjes transform $m_\mu(x+i\eta)$ thus contains information about μ at a scale of η around x.

3. *Stieltjes continuity theorem for sequences of random measures:* A sequence of random probability measures $\mu_1,\mu_2,\dots$ converges vaguely, a) in expectation b) in probability c) almost surely to a deterministic probability measure μ if and only if for all $z\in\mathbb H$, the sequence of numbers $m_{\mu_N}(z)$ converges a) in expectation b) in probability c) almost surely to $m_{\mu}(z)$.

4. Derivation of the *Helffer-Sjöstrand formula*
   $$ f(\lambda)=\frac{1}{2\pi i}\int_{\mathbb C} \frac{\partial_{\overline z} f_{\mathbb C} (z)}{\lambda-z}d \overline z \wedge d z,\quad f_{\mathbb C}(x+i\eta):= \chi(\eta)\big[f(x)+i\eta f'(x)\big] $$ for compactly supported C^2-functions $f\colon\mathbb R\to\mathbb R$ and some smooth cut-off function χ.

October 18

Main motivations for random matrices:

1. Wigner's original motivation: to model energy levels of heavy nuclei. The distribution of the gaps very well matched that of the Wigner random matrices. The density of states depends on the actual nucleus (and it is not the semicircle), but the local statistics (e.g. gap statistics) are universal.

2. Random Schrodinger operators, Anderson transition

3. Gap statistics of the zeros of the Riemann zeta function.

   Quantum Mechanics in nutshell:

- Configuration space: S (with a measure)

- State space: $\ell^2(S)$ (square integrable functions on S)

- Observables: self-adjoint (symmetric) operators on $\ell^2(S)$

- A distinguished observable: the Hamilton (or energy) operator

- Time evolution — Schrödinger equation.

Random Schrödinger operator describes a single electron in an ionic (metallic) lattice. $S = \mathbb{Z}^d$ or a subset of that. H is the sum of the discrete (lattice) Laplace operator and a random potential.

Anderson phase transition: depending on the strength of the disorder, the system is either in delocalized (conductor) or localized (insulator) phase. Localized phase is characterized by

- Localized eigenfunctions

- Localized time evolution (no transport)

- Pure point spectrum (for the infinite volume operator)

- Poisson local spectral statistics, no level repulsion (for the finite volume model)

In the delocalized phase, we have delocalized eigenfunctions ("almost" ℓ^2-normalizable solutions to the eigenvalue equation), quantum transport, absolutely continuous spectrum and random matrix eigenvalue statistics, in particular level repulsion.

Reading. PCMI lecture notes Sections 5.1 — 5.3

October 24

Phase diagram for the Anderson model (= random Schrödinger operator on the $\mathbb{Z}^d$ lattice) in $d \ge 3$ dimensions. Localized regime can be proven, delocalized regime is conjectured to exist but no mathematical result.

In $d=1$ dimension the Anderson model is always localized (transfer matrix method helps). In $d=2$ nothing is known, even there is no clear agreement in the physics whether it behaves more like $d=1$ (localization) or more like $d=3$ (delocalization); majority believes in localization.

4

Delocalized regime, at least for small disorder, sounds easier to prove because it looks like a perturbative problem (zero disorder corresponds to the pure Laplacian which is perfectly understood). Resolvent perturbation formulas were discussed; major problem: lack of convergence.

We gave some explanation why the localization regime is easier to handle mathematically: off-diagonal resolvent matrix elements decay exponentially. This fact provides an effective decoupling and makes localized resolvents almost independent.

Random band matrices: naturally interpolate between $d=1$ dimensional random Schrödinger operators (bandwidth $W=O(1)$) and mean field Wigner matrices (bandwidth $W = N$, where N is the linear size of the system). Phase transition is expected at $W = \sqrt{N}$; this is a major open question. There are similar conjectures in higher dimensional band matrices, but we did not discuss them.

Finally, we discussed a mysterious connection between the Dyson sine kernel statistics and the location of the zeros of of the zeta function on the critical line. There is only one mathematical result in this direction, Montgomery proved that the two point function of the (appropriately rescaled) zeros follows the sine kernel behavior, but only for test functions with Fourier support in $[-1,1]$. No progress has been made in the last 40 years to relax this condition.

Reading. PCMI lecture notes Section 5.3 and the entertaining article "Tea Time in Princeton" by Paul Bourgade about Montgomery's theorem.

October 24 (Recitation)

1. Analytic definition of (multivariate) cumulants κ_α of a random vector $X=(X_1,\dots,X_n)$ as the coefficients of the log-characteristic function

   $$ \log \mathbf E e^{i t\cdot X} = \sum_\alpha \kappa_\alpha \frac{(it)^\alpha}{\alpha!}. $$

2. Proof of the cumulant expansion formula

   $$ \mathbf E X_i f(X)=\sum_{\alpha} \frac{\kappa_{\alpha, i }}{\alpha!}\mathbf E f^{(\alpha)}(X) $$ via Fourier transform.

3. Expression of moments in terms of cumulants as the sum of all partitions

   $$ \mathbf{E} X_1\dots X_n=\sum_{\mathcal{P}\vdash [n]} \kappa^{\mathcal{P}}=\sum_{\mathcal{P}\vdash [n]}\prod_{P_i\in\mathcal{P}} \kappa(X_j \mid j\in P_i) $$

4. Derivation of the inverse relationship

   $$ \kappa(X_1,\dots,X_n)=\sum_{\mathcal P\vdash [n]}(-1)^{\lvert\mathcal P\rvert-1}(\lvert\mathcal P\rvert-1)! \prod_{P_i\in\mathcal P} \mathbf E \prod_{j\in P_i} X_j $$ (eq-combcum) through Möbius inversion on abstract incidence algebras. Note that **eq-combcum** can also serve as a purely combinatorial definition of cumulants.

5. Proof that cumulants of random variables which split into two independent subgroups vanish.

October 25

There are two natural ways to put a measure on the space of (hermitian) matrices, hence defining two major classes of random matrix ensembles:

1. Choose matrix elements independently (modulo the hermitian symmetry) from some distribution on the complex or real numbers. This results in Wigner matrices (and possible generalizations, when identicality of the distribution is dropped).

2. Equip the space of hermitian matrices with the usual Lebesgue measure and multiply it by a Radon-Nikodym factor that makes the measure finite. We choose the factor invariant under unitary conjugation in the form $\exp(-\text{Tr}\, V(H))$ for some real valued function V. These are called invariant ensembles.

Only Gaussian matrices belong to both families.

For invariant ensembles, the joint probability density function of the eigenvalues can be computed explicitly and it consists of the Vandermonde determinant (to the first or second power, $\beta=1,2$, depending on the symmetry class). We sketched of the proof by change of variables.

Invariant ensembles can also be represented as Gibbs measure of N points on the real line with a one-body potential V and a logarithmic two-body interaction. This interpretation allows for choosing any $\beta>0$, yielding the beta-ensembles, even though there is no matrix or eigenvalues behind them. There are analogous universality statements for beta-ensembles, which assert that the local statistics depend only on the parameter beta and are independent of the potential V.

Reading. PCMI lecture notes Section 1.1.2

November 2

Precise statement of the Wigner semicircle law (for i.i.d. case) in the form of weak convergence in probability. In general, there are two methods to prove the semicircle law:

1. Moment method: computes $\text{Tr}\, H^k$, obtains the distribution of the moments of the eigenvalues. The moments are given by the Catalan numbers and they uniquely identify the semicircle law (calculus exercise) using Carleman theorem on the uniqueness of the measure if the moments do not grow too fast.

2. Resolvent method: derives an equation for the limiting Stieltjes transform of the empirical density.

The resolvent method in general is more powerful, it works well inside as well as neat the edge of the spectrum. The moment method is powerful only at the extreme edges.

Proof of the Wigner semicircle law by moment method

Compute
$$ \frac{1}{N} \mathbb E \text{Tr}\, H^k=\frac{1}{N}\mathbb E\sum_{i_1,\dots,i_k} h_{i_1i_2}h_{i_2i_3}\dots h_{i_{k-1}i_k}h_{i_ki_1} $$
in terms of the number of backtracking paths (only those path give a relevant contribution where every edge is travelled exactly twice and the skeleton of the graph is a tree). We reduced the problem to counting such path — it is an N independent problem.

November 7

We completed the proof of the Wigner semicircle law by moment method. Last time we showed that to evaluate $\mathbb E \text{Tr}\, H^{2k}$ is sufficient to count the number of backtracking path of total length $2k$. This number has many other combinatorial interpretations. It is the same as the number of rooted, oriented trees on $k+1$ vertices by a simple one to one correspondance. It is also the same as the number of Dyck paths of length $2k$, where a Dyck path is a random walk on the nonnegative integers starting and ending at 0. Finally, we counted the Dyck paths by deriving the recursion
$$ C_k = C_{k-1} C_0 + C_{k-2} C_1 + \ldots + C_0 C_{k-1} $$
with $C_0=1$ for their number C_k. This recursion can be solved by considering the generating function
$$ f(x) = \sum_{k=0}^\infty C_k x^k $$
and observe that
$$ xf^2(x) = f(x) - 1. $$
Thus $f(x)$ can be explicitly computed by the solution formula for the quadratic equation and Taylor expanding around $x=0$. After some calculation with the fractional binomial coefficients, we obtain that $C_k = 1/(k+1) {2k \choose k}$, i.e. the Catalan numbers.

Since the Catalan numbers are the moments of the semicircle law (calculus exercise), and these moments do not grow too fast, they identify the measure.

This proved that the expectation of the empirical eigenvalue density converges to the semicircle in the sense of moments. Using compact support of the measures (for the empirical density we know it from the homework problem since the norm of H is bounded), by Weierstrass theorem we can extend the convergence for any bounded continuous functions.

Finally, the expectation can be removed, by computing the variance of $N^{-1} \text{Tr}\, H^k$, again by the graphical representation (now we have two cycles and studied which edge-coincidences give rise to nonzero contribution). We showed that the variance vanishes in the large N limit and then a Chebyshev inequality converts it into a high probability bound.

Reading. PCMI lecture notes Section 2.3

November 7 (Recitation)

We found yet another combinatorial description of Catalan numbers. C_k is the number of non-crossing pair partitions of the set $\{1,\dots,2k\}$. Indeed, denote the number in question by N_k. Then there exists some j such that 1 is paired with $2j$ since due to the absence of crossings there has to be an even number of other integers between 1 and its partner. The number of non-crossing pairings of the integers $\{2,\dots,2j-1\}$ and $\{2j+1,\dots,2k\}$ are given by N_{j-1} and N_{k-j} respectively and it follows that $$ N_k=\sum_{j=1}^k N_{j-1}N_{k-j}, \qquad N_1=1 $$ and thus $N_k=C_k$ since they satisfy the same recursion.

We defined a commonly used notion of stochastic domination $X\prec Y$ and stated the following *large deviation estimates* for families of random variables X_i,Y_i of zero mean $\mathbf E X_i=\mathbf E Y_i=0$ and unit variance $\mathbf E \lvert X_i\rvert^2=\mathbf E \lvert Y_i\rvert^2=1$ and deterministic coefficients b_i, a_{ij},
$$ \left\lvert\sum_{i} b_i X_i\right\rvert\prec \left(\sum_i\lvert b_i\rvert^2\right)^{1/2} $$
$$ \left\lvert\sum_{i,j} a_{ij} X_i Y_j\right\rvert\prec \left(\sum_{i,j}\lvert a_{ij}\rvert^2\right)^{1/2} $$ (eq-LDE)

$$ \left\lvert\sum_{i\not=j} a_{ij} X_i X_j\right\rvert\prec \left(\sum_{i\not=j}\lvert a_{ij}\rvert^2\right)^{1/2} $$ We proved **eq-LDE** only for uniformly subgaussian families of random variables but not that uniformly finite moments of all orders are also sufficient for them to hold.

November 9

- Precise statement of the local semicircle laws (entrywise, isotropic, averaged) for Wigner type matrices with moment condition of arbitrary order.

- Definition of stochastic dominations, some properties.

- We started the proof of the weak law for Wigner matrices.

- Schur complement formula. Almost selfconsistent equation for $m_N = N^{-1} \text{Tr}\, G$ assuming that the fluctuation of the quadratic term is small (will be proven later).

- The other two errors were shown to be small. The smallness of the single diagonal element h_{ii} directly follows from the moment condition. The difference of the Stieltjes transform of the resolvent and its minor was estimated via interlacing and integration by parts.

  Reading. PCMI lecture notes Section 3.1.1.

November 14

- Proof of the weak local law in the bulk.

- Stability of the equation for m_{sc}, the Stieltjes transform of the semicircle law.

- Proof for the large eta regime.

- Breaking the circularity of the argument in two steps: In the first step one proves a weaker bound that allows one to approximate m via m_{sc}, then run the argument again but with improved inputs. The bootstrap argument will have the same philosophy next time.

- Discussion of the uniformity in the spectral parameter. Grid argument to improve the bound for supremum over all z. This argument works because (i) the probabilistic bound for any fixed z is very strong (arbitrary $1/N$-power) and (ii) the function we consider $(m-m_{sc})(z)$ has some weak but deterministic Lipschitz continuity.

November 14 (Recitation)

We presented a cumulant approach to proving local laws for correlated random matrices. Specifically, we gave a heuristic argument that the resolvent G should be well apprixmated by the unique solution $M=M(z)$ to the matrix Dyson equation (MDE) $$ 0=1+zM+\mathcal S[M]M, \quad \Im M>0,\qquad \mathcal S[R]:= \sum_{\alpha,\beta}\text{Cov}(h_\alpha,h_\beta)\Delta^\alpha R\Delta^\beta. $$ We furthermore proved that the error matrix $$ D=1+zG+\mathcal S[G]G=HG+\mathcal S[G]G $$ satisfies $$ \mathbf E\lvert\langle x,Dy \rangle\rvert^2 $$

$$\lesssim \left(\frac{\lVert x\rVert \lVert y\rVert}{\sqrt{N\eta}}\right)^2,\qquad \mathbf E\lvert\langle BD \rangle\rvert^2 \lesssim \left(\frac{\lVert B\rVert}{N\eta}\right)^2$$ in the case of Gaussian entries h_α.

November 16

- We completed the rigorous proof of the weak local semicircle law by the bootstrap argument.

- Then we mentioned two improvements: (i) Strong local law (error bound improved from $(N \eta)^{-1/2}$ to $(N\eta)^{-1}$ and (ii) entrywise local law.

- Proof of the entrywise local law via the self-consistent vector equation. Stability operator mentioned in the more general setup of Wigner type matrices (when the variance matrix S is stochastic). Diagonal and offdiagonal elements are estimated separately via a joint control parameter Λ. Main ideas are sketched, the rigorous bootstrap argument was omitted.

  Reading. PCMI lecture notes Sections 4.1–4.3

November 21

- Fluctuation averaging phenomenon. Proof of the strong local law in the bulk. Some remarks on the modifications at the edge. Corollaries of the strong local law: optimal estimates on the eigenvalue counting function and rigidity (location of the individual eigenvalues).

- Bulk universality for Hermitian Wigner matrices. Basic idea: interpolation. Ornstein Uhlenbeck process for matrices (preserves expectation and variance). Crash course on Brownian motion, stochastic integration and Ito's formula. Dyson Brownian motion (DBM) for the eigenvalues. Local equilibration phenomenon due to the strong level repulsion in the DBM.

November 23

We summarized the three step strategy to prove local spectral universality of Wigner matrices. We discussed the second step: fast convergence to equilibrium of the Dyson Brownian Motion. Relation between SDE and PDE: introduction of the generator. Laplacian is the generator of the standard Brownian motion.

Basics of large dimensional analysis: Gibbs measure, entropy, Dirichlet form, generator. The total mass of a probability measure is preserved under the dynamics. Relation between various concepts of closeness to equilibrium. Entropy inequality (total variation norm is bounded by the entropy). Logarithmic Sobolev inequality. Spectral gap inequality. Bakry-Emery theory: (i) the Gibbs measure with a convex Hamiltonian satisfies LSI, (ii) entropy and Dirichlet form decays exponentially fast.

Homework

The problem sheets can either be handed in during the lecture or put in the letter box of Dominik Schröder in LBW, 3rd floor.

		published	due
Problem sheet I	Solutions	Oct 18	Oct 25
Problem sheet II	Solutions	Oct 25	Nov 7
Problem sheet III	Solutions	Nov 9	Nov 21