Large-N behavior of random matrix models through non-commutative function theory

David A. Jekel

University of California, Los Angeles

June 9, 2020

Overview

- Like the previous talk, this one will be big-picture.
- The large-N behavior of some models for several self-adjoint random matrices can be described through free probability.
- We'll fill in a "PDE viewpoint" related to the "SDE viewpoint."
- We aim to describe the large-N limit of some functions on $M_N(\mathbb{C})_{sa}^d$ related to the random matrix models.

Notation

- $M_N(\mathbb{C})$ denotes $N \times N$ matrices, $M_N(\mathbb{C})_{sa}$ denotes the self-adjoint (Hermitian) matrices, $M_N(\mathbb{C})_{sa}^d$ denotes d-tuple of self-adjoint matrices.
- $\|A\|_\infty$ denotes the *operator norm* (equivalently, the largest singular value). If $\mathbf{A} = (A_1, \dots, A_d)$, then $\|\mathbf{A}\|_\infty := \max_j \|A_j\|_\infty$.
- tr_N denotes the normalized trace $\operatorname{tr}_N(A) = (1/N)\operatorname{Tr}(A)$.
- Define (real) inner product on $M_N(\mathbb{C})_{sa}^d$ by $\langle \mathbf{A}, \mathbf{B} \rangle_2 = \sum_{j=1}^d \operatorname{tr}_N(A_j B_j)$ where $\mathbf{A} = (A_1, \dots, A_d)$ and $\mathbf{B} = (B_1, \dots, B_d)$.
- The associated norm is written $\|\mathbf{A}\|_2$.
- Thus, $M_N(\mathbb{C})_{sa}^d \cong \mathbb{R}^{N^2 d}$ isometrically.
- Using such an isometric identification, we can apply all the concepts of classical analysis to functions on $M_N(\mathbb{C})_{sa}^d$.

The need for good observables

- We want to study functions that make sense on $M_N(\mathbb{C})_{sa}^d$ for every N simultaneously.
- This is like the way that in mean field games, you focus on the functions that only depend on the measure produced by your N particles, rather than arbitrary functions of all the positions of the particles.
- As another analogy, in statistical mechanics, you study macroscopic observables and not all the observables you could define in terms of the microstate.

Trace polynomials

Fact (Procesi/Razmyslov)

Every classical polynomial function $M_N(\mathbb{C})_{sa}^d \to \mathbb{C}$ that is invariant under unitary conjugation is a *trace polynomial*. That is, it is in the span of the functions

$$\operatorname{tr}_N(p_1(\mathbf{A})) \dots \operatorname{tr}_N(p_k(\mathbf{A})),$$

where $p_1, \dots, p_k$ are non-commutative polynomials.

Fact (Procesi/Razmyslov)

Every classical polynomial function $M_N(\mathbb{C})_{sa}^d \to M_N(\mathbb{C})$ that is *equivariant* under unitary conjugation is a *non-commutative trace polynomial*. That is, it is in the span of the functions

$$p_0(\mathbf{A}) \operatorname{tr}_N(p_1(\mathbf{A})) \dots \operatorname{tr}_N(p_k(\mathbf{A})),$$

where $p_0, p_1, \dots, p_k$ are non-commutative polynomials.

Completion of trace polynomials

In order to do *analysis*, we don't want to work only with polynomials ...

For N fixed, every unitarily equivariant continuous function $M_N(\mathbb{C})_{sa}^d \to M_N(\mathbb{C})$ can be approximated uniformly on compact sets by a unitarily invariant polynomial (a non-commutative trace polynomial).

What happens if we take limits of non-commutative trace polynomials that converge *for all N simultaneously*?

Definition

Let TrP_d be the space of NC trace polynomials. For $R > 0$, define a norm on TrP_d by

$$\|f\|_{2,R} = \sup_N \sup\{\|f(\mathbf{A})\|_2 : \mathbf{A} \in M_N(\mathbb{C})_{sa}^d, \|\mathbf{A}\|_\infty \le R\}.$$

Let $\overline{\mathrm{TrP}}_d$ be the completion of TrP_d with respect to the seminorms $\|\cdot\|_{2,R}$ for $R > 0$.

Asymptotic approximation

The space $\overline{\mathrm{TrP}}_d$ represents "functions of d non-commuting self-adjoint variables." It will be used to describe the asymptotic behavior of certain functions on $M_N(\mathbb{C})_{sa}^d$ as $N \to \infty$.

Definition

Let $f^{(N)} : M_N(\mathbb{C})_{sa}^d \to M_N(\mathbb{C})$ for each N and $f \in \overline{\mathrm{TrP}}_d$. Then we say that $f^{(N)}$ is *asymptotic to f*, written $f^{(N)} \rightsquigarrow f$, if for every $R > 0$,

$$\lim_{N \to \infty} \sup\{\|f^{(N)}(\mathbf{A}) - f(\mathbf{A})\|_2 : \mathbf{A} \in M_N(\mathbb{C})_{sa}^d, \|\mathbf{A}\|_\infty \leq R\} = 0.$$

In other words, $f^{(N)} \rightsquigarrow f$ means that the function $f^{(N)}$ is approximated uniformly on operator-norm balls by the abstract function f as $N \to \infty$.

Remark

The previous talk explained Guionnet and Shlyakhtenko's result that certain free transport functions were approximated by the finite-N optimal transport functions in L^2 norm.

The notion of asymptotic approximation in this talk is stronger, because it gives a uniform approximation on an operator-norm ball rather than an L^2 approximation.

Due to concentration of measure, L^2 approximation is very different than uniform approximation since the random matrix measures are asymptotically supported on smaller sets where our macroscopic observables are close to being constant.

Convergence of expectations

Theorem

Consider a probability measure $\mu^{(N)}$ *on* $M_N(\mathbb{C})_{sa}^d$ *of the form*

$$d\mu^{(N)}(\mathbf{A}) = \frac{1}{Z^{(N)}} e^{-N^2 V^{(N)}(\mathbf{A})}\, d\mathbf{A},$$

where $V^{(N)} : M_N(\mathbb{C})_{sa}^d \to \mathbb{R}$ *such that*

1. $V^{(N)}(\mathbf{A}) - (c/2)\|\mathbf{A}\|_2^2$ *is convex and* $V^{(N)}(\mathbf{A}) - (C/2)\|\mathbf{A}\|_2^2$ *is concave for some* $0 < c < C$.
2. $\nabla V^{(N)}(\mathbf{A})$ *is asymptotic to some* $g \in (\overline{\mathrm{TrP}_d})^d$ *as* $N \to \infty$.

If $f^{(N)} : M_N(\mathbb{C})_{sa}^d \to \mathbb{R}$ *is asymptotic to some* f *and satisfies some mild growth conditions at* ∞*, then* $\lim_{N\to\infty} \int f^{(N)}\, d\mu^{(N)}$ *exists.*

Heat semigroup for $\mu^{(N)}$

Due to $e^{-N^2 V^{(N)}}$ having an N^2 in the exponent, it is not easy to evaluate the large N limit of $\int f^{(N)}\, d\mu^{(N)}$ directly. Rather, we use indirect methods which are *dimension-independent*.

As mentioned in Dima's talk, $\mu^{(N)}$ is the stationary distribution for the SDE

$$dX_t^{(N)} = dS_t^{(N)} - \frac{1}{2}\nabla V^{(N)}(X_t^{(N)})\, dt.$$

Here $S_t^{(N)}$ is a Brownian motion on $M_N(\mathbb{C})_{sa}^d$ such that $E\|S_t^{(N)}\|_2^2 = d \cdot t$.

Studying the large-N behavior of the matrix Brownian motion is one approach to proving the theorem (or at least a similar theorem), used in several papers by Dabrowski, Guionnet, and Shlyakhtenko.

Heat semigroup for $\mu^{(N)}$

The SDE is closely related to the heat semigroup associated to $\mu^{(N)}$. Indeed, for $t \geq 0$, define an operator $T_t^{(N)}$ acting on Lipschitz functions on $M_N(\mathbb{C})_{sa}^d$ by

$$T_t^{(N)} f^{(N)}(\mathbf{A}) = E\left[f^{(N)}(X_t^{(N)}) | X_0^{(N)} = \mathbf{A}\right].$$

Then $(T_t^{(N)})_{t \geq 0}$ is a semigroup, and $F^{(N)}(x,t) = T_t^{(N)} f(x)$ solves the heat equation

$$\partial_t F^{(N)} = \frac{1}{2N^2} \Delta F^{(N)} - \frac{1}{2} \langle \nabla V^{(N)}, \nabla F^{(N)} \rangle_2.$$

This implies that $\int T_t^{(N)} f^{(N)}\, d\mu^{(N)} = \int f^{(N)}\, d\mu^{(N)}$ for all t.

Heat semigroup for $\mu^{(N)}$

It turns out that if $f^{(N)}$ is uniformly $\|\cdot\|_2$-Lipschitz and $f^{(N)} \rightsquigarrow f$, then $T_t^{(N)} f^{(N)}$ is also asymptotic to some $T_t f$ (and thus there is a "free version" T_t of the heat semigroup acting on $\overline{\mathrm{TrP}}_d$ in the large-N limit).

Due to the uniform convexity of $V^{(N)}$ (that is, $V^{(N)}(\mathbf{A}) - (c/2)\|\mathbf{A}\|_2^2$ is convex), we find that $\|T_t^{(N)} f^{(N)}\|_{\mathrm{Lip}} \leq e^{-ct/2}\|f^{(N)}\|_{\mathrm{Lip}}$, which implies that $T_t f^{(N)}$ converges to the constant $\int f^{(N)}\,d\mu^{(N)}$ as $t \to \infty$.

The rate of convergence is independent of N. Thus, we can exchange limits: Since $T_t^{(N)} f^{(N)}$ has a large-N limit in $\overline{\mathrm{TrP}}_d$ for each t, we conclude that $\int f^{(N)}\,d\mu^{(N)}$ has a limit as $N \to \infty$. And the limit can be evaluated using the semigroup T_t on the abstract function space $\overline{\mathrm{TrP}}_d$.

The large-N limit of entropy

Theorem

Let $V^{(N)}$ and $\mu^{(N)}$ be as above. Let h denote the classical entropy of a probability measure, given by $-\int \rho \log \rho$, where ρ is the density. Then

$$\lim_{N \to \infty} \left(\frac{1}{N^2} h(\mu^{(N)}) + d \log N \right) \text{ exists,}$$

and it is equal to something known in free probability as the non-microstates free entropy χ^* of the limiting non-commutative law.

The large-N limit of entropy

Again, we approach the problem rather indirectly. Let $\sigma_t^{(N)}$ be the Gaussian probability measure on $M_N(\mathbb{C})_{sa}^d$ with density proportional to $e^{-N^2\|\mathbf{A}\|_2^2/2}\,d\mathbf{A}$.

Let $\mu_t^{(N)} = \mu^{(N)} * \sigma_t^{(N)}$ (the law of a random variable obtained by adding an independent Gaussian). We can write

$$d\mu_t^{(N)}(\mathbf{A}) = \frac{1}{Z^{(N)}} e^{-N^2 V_t^{(N)}(\mathbf{A})}\,d\mathbf{A}.$$

It is not hard to show that

$$\frac{1}{N^2}\partial_t h(\mu_t^{(N)}) = \frac{1}{2}\int \|\nabla V_t^{(N)}\|_2^2\,d\mu_t^{(N)}.$$

(The derivative of entropy is Fisher's information.)

The large-N limit of entropy

Using well-known facts about entropy and Fisher information, the theorem can be reduced to showing that

$$\lim_{N \to \infty} \int \left\| \nabla V_t^{(N)} \right\|_2^2 d\mu_t^{(N)} \text{ exists for every} t \geq 0.$$

In order to do this, it suffices to show that $\nabla V_t^{(N)}$ is asymptotic to some $g_t \in (\overline{\mathrm{TrP}}_d)^d$ (since we can the first theorem to the measure $\mu_t^{(N)}$ and the function $\|\nabla V_t^{(N)}\|_2^2$).

There are at least two different proofs of this, and I won't explain either one, only give some general remarks about the ideas and some open questions.

Hamilton-Jacobi-Bellman equation

This late in the talk, we finally get to some HJ equations ...

The potential $V_t^{(N)}$ evolves according to the equation

$$\partial_t V_t^{(N)} = \frac{1}{2N^2} \Delta V_t^{(N)} - \frac{1}{2} \|\nabla V_t^{(N)}\|_2^2.$$

There is an SDE viewpoint on this as well, but it is more complicated than the case of the heat equation; it involves forward-backward SDE and has a stochastic variational characterization.

The methods in my paper are purely PDE / semigroup based, but they heavily rely on the dimension-independent regularity that comes from the uniform convexity of $V^{(N)}$.

Hamilton-Jacobi-Bellman equation

The difficulty in general is that the regularizing effect of the free Laplacian in the large-N limit is not as strong as the regularizing effect of the classical Laplacian in $\mathbb{R}^d$, and it is something we don't understand well yet.

So with the present technological limitations, the regularity (and asymptotic behavior as $N \to \infty$) that we can obtain for the HJB equation is not much better than the HJ equation

$$\partial_t V^{(N)} = -\frac{1}{2}\|\nabla V_t^{(N)}\|_2^2.$$

This is a good toy problem to consider because the solution has a very explicit variational form given by the *Hopf-Lax formula*:

$$V_t^{(N)}(\mathbf{A}) = \inf_{\mathbf{B}} \left[V^{(N)}(\mathbf{B}) + \frac{1}{2t}\|\mathbf{A} - \mathbf{B}\|_2^2 \right].$$

Inf-convolution

Let $V_t^{(N)}$ be given by the Hopf-Lax formula above. When will $\nabla V_t^{(N)}$ will be asymptotically approximable by functions in $\overline{\mathrm{TrP}}_d$?

Recall some basic facts. If $\mathbf{A}$ is given and $\mathbf{B}$ achieves the infimum defining $V_t^{(N)}(\mathbf{A})$, then we have

$$\mathbf{B} = \mathbf{A} - t\nabla V^{(N)}(\mathbf{B}).$$

Also, if $V_t^{(N)}$ is differentiable at $\mathbf{A}$, then

$$\nabla V_t^{(N)}(\mathbf{A}) = \nabla V^{(N)}(\mathbf{B}).$$

Inf-convolution

In particular, if $\nabla V^{(N)}$ is C-Lipschitz and $t < 1/C$, then the infimizer $\mathbf{B}$ can be obtained from a fixed-point iteration, and hence we can obtain $\nabla V_t^{(N)}(\mathbf{A})$.

This strategy shows that $\nabla V_t^{(N)}$ is asymptotic to some tuple of functions from $\overline{\mathrm{TrP}}_d$, since our notion of asymptotic approximation is respected by composition and limits.

In fact, if $V^{(N)}$ is convex and semi-concave, then this strategy will work for all $t \geq 0$. The reason is that $V_t^{(N)}$ remains convex for all t.

Inf-convolution

More generally, if $cI \leq HV^{(N)} \leq CI$, then

$$\frac{c}{1+ct} I \leq HV_t^{(N)} \leq \frac{C}{1+Ct} I$$

as long as $1 + ct < 0$. If $c < 0$, then at some time we will hit $1 + ct = 0$ and then $V_t^{(N)}$ could fail to be differentiable.

There are well-known classical examples where this differentiability breaks down, but what about trace polynomials?

Inf-convolution

Let $V^{(N)}(A_1, A_2) = -\|e^{iA_1}e^{iA_2} - e^{iA_2}e^{iA_1}\|_2^2$, which is a smooth bounded function. If $N > 1$, then there are self-adjoint matrices A_1 and A_2 which don't commute, which implies that $\inf V^{(N)} < 0$. Hence, for sufficiently large t,

$$V_t^{(N)}(0) = \inf_{\mathbf{B}} \left[V^{(N)}(\mathbf{B}) + \frac{1}{2t}\|\mathbf{B}\|_2^2 \right] < 0.$$

If it happens that $\nabla V_t^{(N)}$ is well-defined approximated by trace polynomials, then the infimizer for $V_t^{(N)}(0)$ would be a trace polynomial of the pair $(0,0)$, hence it would have be $(\beta_1 I, \beta_2 I)$ for some scalars β_1 and β_2. But scalar multiples of the identity would commute, and hence $V^{(N)}(\mathbf{B})$ would be zero, which contradicts $V_t^{(N)}(0,0) < 0$.

Inf-convolution

This illustrates a conceptual obstacle to studying the large-N limit of this HJ equation: The infimum could be achieved at some point $\mathbf{B}$ that is not in the algebra generated by $\mathbf{A}$. In this case, the function no longer depends only on the behavior of $\mathbf{A}$ but could also depend on how $\mathbf{A}$ sits within the larger matrix algebra. In that case, $V_t^{(N)}$ itself cannot be approximated using only trace polynomials.

In principle, the same problem could occur when studying the stochastic variational formula for the HJB equation. This issue underlies many of the difficulties in the theory of free entropy and the large deviation principles for random matrices — for instance, the only known theorems relating the free entropy χ^* coming from the classical entropy assume that $V^{(N)}$ is convex.