A variational approach to free convolutions

Octavio Arizmendi
CIMAT

joint work with Samuel G. G. Johnston

Probabilistic Operator Algebra Seminar
Berkeley, November 27, 2023

1. Finite free convolutions

Finite Free Convolution

- For a matrix M, let $\chi_M(x) = \det(xI - M)$ denote the characteristic polynomial of the matrix M.
- For d-dimensional hermitian matrices A_d and B_d, let p_d and q_d be their characteristic polynomials.
- We define the **finite free additive convolution** of p_d and q_d to be

$$p_d(x) \boxplus_d q_d(x) = \mathbf{E}_{\mathcal{U}_n}[\chi_{A_d+UB_dU^{-1}}(x)],$$

where the expectation is taken over unitary matrices U sampled according to the Haar measure on $\mathcal{U}_n$.
- The convolution does not depend on the specific choice of A_d and B_d, but only on p_d and q_d.

Finite Free Additive Convolution

Theorem (Marcus, Spielman, Srivastava)

For $p(x) = \sum_{i=0}^{d} x^{d-i}(-1)^i a_i^p$ *and* $q(x) = \sum_{i=0}^{d} x^{d-i}(-1)^i a_i^q$*, the finite free convolution of* p *and* q *is given by*

$$p(x) \boxplus_d q(x) = \sum_{k=0}^{d} x^{d-k}(-1)^k \sum_{i+j=k} \frac{(d-i)!(d-j)!}{d!(d-i-j)!} a_i^p a_j^q.$$

Corollary (Theorem of Walsh)

If p *and* q *are real rooted then* $p \boxplus_d q$ *is real rooted.*

Proof: This convolution was known ~ 100 years ago.

Finite Free Multiplicative Convolution

Similarly, when A and B are positive semidefinite, the *finite free multiplicative convolution* of p and q is defined to be

$$p(x) \boxtimes_d q(x) = \mathbf{E}_{U(n)}[\chi_{AUBU^{-1}}(x)], \tag{1}$$

In this case they proved

$$p(x) \boxtimes_d q(x) = \sum_{i=0}^{d} x^{d-i}(-1)^i \frac{a_i b_i}{\binom{d}{i}}.$$

Real rootness was proved by Szego.

Finite Free Compression

The *finite free compression* of $p = p_A$ is defined to be

$$[p(z)]_k = \mathbb{E}[\det(z - [U^* AU]_k)] = \frac{k!}{N!}\left(\frac{\partial}{\partial z}\right)^{N-k} p(z)$$

Marcus and Gorin 2020 observe:

$$[p(z)]_k = \left(\frac{\partial}{\partial z}\right)^{N-k} p(z)$$

$\boxplus_d$ approximates $\boxplus$

Theorem (A., Perales 18, Marcus 16, Guionnet, Shlyakthenko 16?)

Let $(p_d)_{d=1}^{\infty}$ *and* $(q_d)_{d=1}^{\infty}$ *with*

$$\mu_{p_d} \to \mu, \quad \mu_{q_d} \to \nu$$

as $d \to \infty$ *then*

$$\mu_{(p_d \boxplus_d q_d)} \to \mu \boxplus \nu.$$

Theorem (Garza-Vargas, Perales 23)

Let $(p_d)_{d=1}^{\infty}$ and $(q_d)_{d=1}^{\infty}$ with positive real roots and such that

$$\mu_{p_d} \to \mu, \quad \mu_{q_d} \to \nu$$

as $d \to \infty$ then

$$\mu_{(p_d \boxtimes_d q_d)} \to \mu \boxtimes \nu.$$

Derivatives approximate free compresison

Theorem (Hoskins- Kabluchko 2021, A.,Garza-Vargas, Perales 23)

Let $(p_d)_{d=1}^{\infty}$ *and* $(q_d)_{d=1}^{\infty}$ *with*

$$\mu_{p_d} \to \mu$$

as $d \to \infty$ *then*

$$\mu_{[p_d]_{td}} \to [\mu]_{\boxplus t}$$

Finite Free Convolution

Usual approach in RMT:

- Take A and B, $d \times d$ matrices.
- We consider $X = A + UBU^{-1}$
- Calculate $m_n := \mathbb{E}(tr(X^n))$
- Find a measure such that

$$\int x^n \mu(dx) = m_n$$

New approach:

- Take A and B, $d \times d$ matrices.
- We consider $X = A + UBU^{-1}$
- Calculate the **expected characteristic polynomial** of the matrix X.
- Consider the measure supported on the roots of this polynomial.

2. Quadrature formulas and Two Large N Limits

Quadrature formulas

Theorem (Marcus, Spielman, Srivastava)

For A_N and B_N selfadjoint matrices the

$$\mathbf{E}_{\mathcal{U}_N}[\det(z - A - UBU^{-1})] = \mathbf{E}_{\hat{\mathcal{S}}_N}[\det(z - A - \Sigma B \Sigma^{-1})].$$

where U is a Haar unitary matrix under $\mathcal{U}(N)$, and Σ is a uniform signed permutation matrix over $\hat{\mathcal{S}}(N)$.

Quadrature formulas

Theorem (Marcus,Spielman, Srivastava)

For A_N and B_N diagonal matrices,

$$\mathbf{E}_{\mathcal{U}_N}[\det(z - A - UBU^{-1})] = \mathbf{E}_{\mathcal{S}_N}[\det(z - A - \Sigma B \Sigma^{-1})].$$

where U is a Haar unitary matrix under $\mathcal{U}(N)$, and Σ is a uniform permutation matrix over $\mathcal{S}(N)$.

Asymptotic freeness for sums of matrices

Let $(A_n)_{N\geq 1}$ and $(B_n)_{N\geq 1}$ be selfadjoint random matrices such that, in distribution

$$\mu_{A_N} \to \mu \quad \text{and} \quad \mu_{B_N} \to \nu.$$

Consider

$$X_N = A_N + U_N B U_N^{-1}.$$

where U_N is sampled according to the Haar measure on $\mathcal{U}(N)$. Then

$$\mu_{X_n} \to \mu \boxplus \nu \quad a.s.$$

- **Note that we can assume that A_N and B_N are diagonal.**

In particular,

$$\tilde{\mu}_{A_N + U_N B U_N^{-1}} \to \mu \boxplus \nu$$

where $\tilde{\mu}_{X_n}$ denotes the *average empirical eigenvalue distribution*. Equivalently,

$$\mathbb{E}(tr(X_n^l)) \to \int_{\mathbb{R}} x^n \mu \boxplus \nu(dx)$$

or

$$\mathbb{E}[tr((zI - X_N)^{-1})] \to \int_{\mathbb{R}} \frac{1}{z-x} \mu \boxplus \nu(dx), \quad z \in \mathbb{C}^+$$

Quadrature formulas

Now, for A_N and B_N seladjoint diagonal matrices such that, in distribution

$$\mu_{A_N} \to \mu \quad \text{and} \quad \mu_{B_N} \to \nu.$$

Then if we consider

$$Y_N = A_N + \Sigma_N B \Sigma_N^{-1}$$

where Σ_N is a uniform permutation matrix over $\mathcal{S}(N)$
Then

$$\mu_{X_n} \to \mu * \nu \quad a.s.$$

- **Note that A_N and B_N must be diagonal.**

Classical Convolution

If

$$\mu_{A_N+\Sigma_N B_M \Sigma_N^{-1}} = \mu * \nu$$

where $\tilde{\mu}_{X_n}$ denotes the *average empirical eigenvalue distribution*. Equivalently,

$$\mathbb{E}[tr((zI - X_N)^{-1})] \to \int_{\mathbb{R}} \frac{1}{z-x} \mu * \nu(dx), \quad z \in \mathbb{C}^+$$

Summary

Now, for A_N and B_N seladjoint diagonal matrices with $\mu_{A_N} \to \mu$ and $\mu_{B_N} \to \nu$. Then

$$\mu_{A_N + \Sigma_N B_M \Sigma_N^{-1}} \to \mu * \nu$$

and

$$\mu_{A_N + U_N B U_N^{-1}} \to \mu \boxplus \nu$$

but

$$\mathbf{E}_{\mathcal{U}_N}[\det(z - A - UBU^{-1})] = \mathbf{E}_{\mathcal{S}_N}[\det(z - A - \Sigma B \Sigma^{-1})].$$

We want to study both sides and relate them.

Asymptotics for U_N

Now for a matrix A_N and $z >> ||A_N||$ we have that

$$\det(z - A_N) = \prod_i^N (z - \lambda_i) = e^{N \int log(z-x) d\mu_{A_N}} \approx e^{Nf(z)}$$

with $log(z + ||A||) \leq f(z) \leq log(z - ||A||)$.
So we will be interested in

$$\frac{1}{N} \log \mathbf{E}_{\mathcal{U}_N}[\det(z - A_N - U_N B_N U_N^{-1})]$$

Asymptotics for U_N

Now for a matrix A_N and $z >> ||A_N||$ we have that

$$\det(z - A_N) = \prod_i^N (z - \lambda_i) = e^{N \int log(z-x) d\mu_{A_N}} \approx e^{Nf(z)}$$

with $log(z + ||A||) \leq f(z) \leq log(z - ||A||)$.
So we are interested in

$$\frac{1}{N} \log \mathbf{E}_{\mathcal{U}_N}[\det(z - A_N - U_N B_N U_N^{-1})]$$

Theorem (A. Johnston 2023)

For $(A_N)_N$ and $(B_N)_N$ selfadjoint matrices

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}_{\mathcal{U}_N}[\det(z - A_N - U_N B_N U_N^{-1})]$$
$$= \int_{-\infty}^{\infty} \log(z-\lambda)(\mu \boxplus \nu)(\mathrm{d}\lambda).$$

- This says that in the convergence

$$\mu_{A_N + U_N B U_N^{-1}} \to \mu \boxplus \nu$$

we can swap log and **E**.

Asymptotics for S_N

Theorem (A. Johnston 2023)

For $(A_N)_N$ and $(B_N)_N$ **diagonal** *matrices*

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}_{\mathcal{S}_N}[\det(z - A_N - \Sigma_N B_N \Sigma_N^{-1})]$$
$$= \int_{-\infty}^{\infty} \log(z-\lambda)(\mu \boxplus \nu)(\mathrm{d}\lambda).$$

- This says that in the convergence

$$\mu_{A_N + \Sigma_N B_N \Sigma_N^{-1}} \to \mu * \nu$$

we can not swap log and **E** and there should be some terms with high contribution.

Remark: Lower Bounds

Note that since log is concave, by Jensen's

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbf{E}_{\mathcal{U}_N}[\det(z - A_N - U_N B_N U_N^{-1})] \geq \int_{-\infty}^{\infty} \log(z-x)(\mu \boxplus \nu)(\mathrm{d}x).$$

and

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbf{E}_{\mathcal{S}_N}[\det(z - A_N - \Sigma_N B_N \Sigma_N^{-1})] \geq \int_{-\infty}^{\infty} \log(z-x)(\mu * \nu)(\mathrm{d}x).$$

3. Asymptotic for $\mathcal{S}_N$

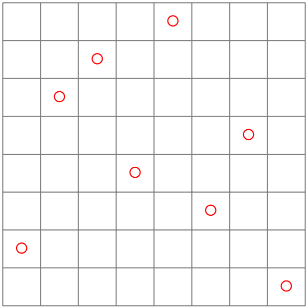
Asymptotics for S_N

If the intuition of $\mu_{A_N+\Sigma_N B_N \Sigma_N^{-1}} \to \mu * \nu$ is incorrect because the contributions of $\mathbf{E}_{\mathcal{S}_N}$ are not "uniform" in

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}_{\mathcal{S}_N}[\det(z - A_N - \Sigma_N B_N \Sigma_N^{-1})]$$

For a permutation σ_N on S_N, we associate a measure π_{σ_N} on $[0,1]^2$ defined by

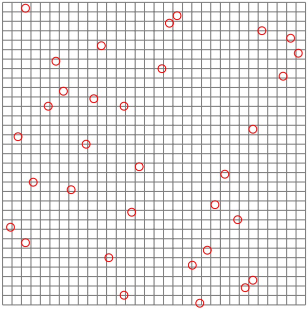
$$\pi_{\sigma_N} := \frac{1}{N}\sum_{i=1}^{N} \delta_{(i/N, \sigma_N(i)/N)}.$$



Large deviations for S_N

If the permutation σ_N is uniformly sampled on S_N, then

$$\pi_{\sigma_N} \approx Leb([0,1]^2)$$



If the permutation σ_N is uniformly sampled on S_N, then

$$\pi_{\sigma_N} \approx Leb([0,1]^2)$$

- However, we are interested in the probability $P_{S_N}(\pi_{\sigma_N} \approx f(s,t))$ for a density function $f(s,t)$, which satisfies
$$\int_0^1 f(s_0,t)\mathrm{d}t = \int_0^1 f(s,t_0)\mathrm{d}s = 1$$
for all $s_0, t_0 \in [0,1]$, in some suitable sense.
- f is called a **coupling density** or a **permuton**.

Theorem (Wu 91, Trashorras 08, KKRW 03)

Under the topology of weak convergence of probability measures on $[0,1]^2$, the random measures π_{σ_N} satisfy a large deviation principle with good rate function $I(\pi)$ given by

$$I(\pi) := \begin{cases} \int_{[0,1]^2} f(s,t) \log f(s,t) \mathrm{d}s\mathrm{d}t & \text{if } \pi \text{ is a permuton} \\ +\infty & \text{otherwise.} \end{cases}$$

That is

$$P_{S_N}(\pi_{\sigma_N} \approx f(s,t)) \approx e^{-NI(f)}$$

Back to finite free convolution

Going back to our problem we rewrite the expected characteristic polynomial as follows,

$$\begin{aligned}
&\mathbf{E}_{\mathcal{S}_N}[\det(z - A_N - \Sigma_N B_N \Sigma_N^{-1})] \\
&= \mathbf{E}_{\mathcal{S}_N}[\prod_i (z - \lambda_i^A - \lambda_{\sigma_N(i)}^B)] \\
&\approx \mathbf{E}_{\mathcal{S}_N} exp[N \int_{[0,1]^2} log(z - \rho_\mu(s) - \rho_\nu(s))\pi_{\sigma_N}]
\end{aligned}$$

where $\rho_\mu(s)$ and $\rho_\nu(s)$ are the quantile functions of μ and ν.

Back to finite free convolution

Going back to our problem we rewrite the expected characteristic polynomial as follows,

$$\begin{aligned} & \mathbf{E}_{\mathcal{S}_N}[\det(z - A_N - \Sigma_N B_N \Sigma_N^{-1})] \\ & = \mathbf{E}_{\mathcal{S}_N}[\prod_i (z - \lambda_i^A - \lambda_{\sigma_N(i)}^B)] \\ & \approx \mathbf{E}_{\mathcal{S}_N} exp[N \int_{[0,1]^2} log(z - \rho_\mu(s) - \rho_\nu(s))\pi_{\sigma_N}] \end{aligned}$$

where $\rho_\mu(s)$ and $\rho_\nu(s)$ are the quantile functions of μ and ν.

Back to finite free convolution

Theorem (A. Johnston)

For w nice enough,we have

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}[exp[N \int_{[0,1]^2} w d\pi_{\sigma_N}]] = \sup_{f \ permuton} \mathcal{G}[w, f].$$

where

$$\mathcal{G}[w, f] := \int_{[0,1]^2} (w(s,t) - \log f(s,t)) f(s,t) ds dt.$$

Back to finite free convolution

Theorem (A. Johnston)

For w nice enough, we have

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}[exp[N \int_{[0,1]^2} w \mathrm{d}\pi_{\sigma_N}]] = \sup_{f \text{ permuton}} \mathcal{G}[w, f].$$

where

$$\mathcal{G}[w, f] := \int_{[0,1]^2} (w(s,t) - \log f(s,t)) f(s,t) \mathrm{d}s\mathrm{d}t.$$

Back to finite free convolution

Theorem

For $(A_N)_N$ *and* $(B_N)_N$ **diagonal** *matrices*

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}_{\mathcal{S}_N}[\det(z - A_N - \Sigma_N B_N \Sigma_N^{-1})]$$
$$= \sup_f \mathcal{G}\left[\log(z - \rho_\mu(s) - \rho_\nu(t)), f\right].$$

Theorem

For $(A_N)_N$ *and* $(B_N)_N$ **diagonal** *matrices*

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbf{E}_{\mathcal{S}_N}[\det(z - A_N - \Sigma_N B_N \Sigma_N^{-1})]$$
$$= \int_{-\infty}^{\infty} \log(z - \lambda)(\mu \boxplus \nu)(\mathrm{d}\lambda).$$

Main theorem

Theorem (Hydrodynamic quadrature theorem)

Let μ and ν be probability measures on the real line with compact support, and let ρ_μ, ρ_ν be the right-continuous inverses of their distribution functions. Then

$$\int_{-\infty}^{\infty} \log(z-\lambda)(\mu \boxplus \nu)(\mathrm{d}\lambda) = \sup_f \mathcal{G}\left[\log(z-\rho_\mu(s)-\rho_\nu(t)), f\right].$$

Likewise,

$$\int_{-\infty}^{\infty} \log(z-\lambda)(\mu \boxtimes \nu)(\mathrm{d}\lambda) = \sup_f \mathcal{G}\left[\log(z-\rho_\mu(s)\cdot\rho_\nu(t)), f\right],$$

and

$$\tau \int_{-\infty}^{\infty} \log(z-\lambda)[\mu]_\tau(\mathrm{d}\lambda) = \sup_f \mathcal{G}\left[\log(z-\rho_\mu(s))1_{[0,\tau]}(t), f\right].$$

Main theorem

Theorem (Hydrodynamic quadrature theorem, coupling version)

Let μ and ν be probability measures on the real line with compactly supported density functions. Then

$$\int_{-\infty}^{\infty} \log(z-\lambda)(\mu \boxplus \nu)(\mathrm{d}\lambda)$$
$$= \sup_{\Pi} \left\{ \mathbf{E}_{\Pi}[\log(z-(X+Y)] - \mathcal{E}[\Pi] + \mathcal{E}[\mu] + \mathcal{E}[\nu] \right\},$$

where $\mathcal{E}[\pi] := \int_{\mathbb{R}^k} g(\mathbf{x}) \log g(\mathbf{x}) \mathrm{d}\mathbf{x}$. *and the supremum is taken over all couplings* Π *on* $\mathbb{R}^2$ *of the measures* μ *and* ν.

Inequality with classical convolution

Theorem

Let μ and μ have compact support and z larger than the sum of supremums of this supports then

$$\int_{-\infty}^{\infty} \log(z-x)\mu \boxplus \nu(\mathrm{d}x) \geq \int_{-\infty}^{\infty} \log(z-x)\mu * \nu(\mathrm{d}x),$$

Main theorem

Theorem (Optimal Entropic Transport Version)

Let μ and ν be probability measures on the real line with bounded support. Then for $z > E_\mu + E_\nu$ we have

$$\int_{-\infty}^{\infty} \log(z - x)(\mu \boxplus \nu)(\mathrm{d}x) = \sup_{\Pi} \{\mathbf{E}_\Pi[\log(z - (X + Y)] - H(\Pi|\mu \otimes \nu)\}.$$

$H(\Pi_1|\Pi_2)$ denotes the relative entropy $\int_E \log \frac{\mathrm{d}\Pi_1}{\mathrm{d}\Pi_2} \, \mathrm{d}\Pi_1$ and the supremum is taken over all couplings Π on $\mathbb{R}^2$ of the measures μ and ν.

For $c : \mathbb{R}^2 \to \mathbb{R}$ measurable function and $\kappa \geq 0$, the entropic optimal transport problem is

$$\sup_{\Pi} \left\{ \mathbf{E}_{\Pi}[c(X,Y)] - \kappa H(\Pi | \mu \otimes \nu) \right\},$$

- Sending $\kappa \downarrow 0$, recovers standard Monge-Kantorovich problem and
- When $\kappa \to \infty$, the infimum is attained by couplings approach the product measure $\mu \otimes \nu$.

Explicit Solution

Lemma (Bernton, Ghosal and Nutz (2022))

Let $c : \mathbb{R}^2 \to \mathbb{R}$ be measurable and bounded. Then there is a unique coupling Π_* of probability measures μ and ν maximising the functional

$$\mathcal{F}_c[\Pi] := \mathbf{E}_\Pi[c(X,Y)] - H(\Pi|\mu \otimes \nu).$$

The optimal coupling Π_* has Radon-Nikodym derivative against product measure taking the form

$$\frac{\mathrm{d}\Pi_*}{\mathrm{d}(\mu \otimes \nu)} = F(X)G(Y)e^{c(X,Y)} \tag{2}$$

where F and G are measurable functions. Moreover, Π_* is the unique coupling of μ and ν whose Radon-Nikodym derivative takes the form (3).

Explicit Solution

Plugging

$$\frac{\mathrm{d}\Pi_*}{\mathrm{d}(\mu \otimes \nu)} = F(X)G(Y)e^{c(X,Y)} \tag{3}$$

we see that

$$\sup_{\Pi} \mathcal{F}_c[\Pi] = \mathcal{F}_c[\Pi_*] = -\mu(\log F(X)) - \nu(\log G(Y)).$$

In our case, one can prove that

$$F(X)G(Y) := \frac{\omega(z)}{(\omega_\mu(z) - X)(\omega_\nu(z) - Y)}$$

Subordination functions

Theorem

$$\int_{-\infty}^{\infty} \log(z-\lambda)(\mu \boxplus \nu)(\mathrm{d}\lambda)$$
$$= -\log(\omega - z) + \int_{-\infty}^{\infty} \log(\omega_\mu - \lambda)\mu(\mathrm{d}\lambda) + \int_{-\infty}^{\infty} \log(\omega_\nu - \lambda)\nu(\mathrm{d}\lambda).$$

where $\omega, \omega_\mu, \omega_\nu$ are the unique solutions to the relations

$$\omega(z) = \omega_\mu(z) + \omega_\nu(z)$$
$$\frac{1}{\omega(z) - z} = G_\mu(\omega_\mu(z)) = G_\nu(\omega_\nu(z)).$$

Taking derivatives we obtain

$$\int_{-\infty}^{\infty} \frac{1}{z-\lambda}(\mu \boxplus \nu) = \frac{1}{\omega}$$

Final Remarks and Questions

1. Other polynomials? Other quadrature formulas?
2. Can we consider sup over non commutative distributions?
3. Can we consider make a sense to the solution of κ in the optimal entropic transport problem.
4. Can we relate with variational properties of atoms in free random variables from ACSY 21.

Thanks!