Computing free convolutions via contour integrals

Alice Cortinovis – alicecor@stanford.edu

Joint work with Lexing Ying (Stanford University)

June 12, 2023
UC Berkeley Probabilistic Operator Algebra Seminar

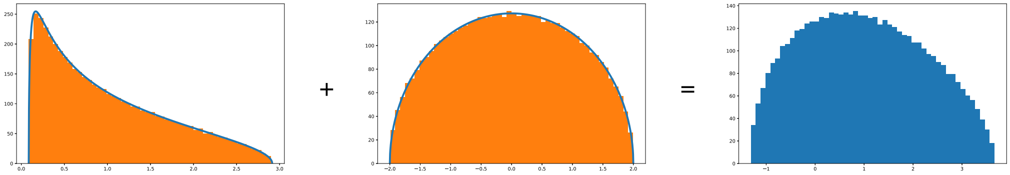
Goal: Develop efficient algorithm for computing free additive and multiplicative convolutions of "nice" measures.

Outline:

- Motivation
- Algorithm for the additive case
- Numerical examples
- Algorithm in the multiplicative case
- Conclusion

Motivation #1: Eigenvalues of a sum of random matrices

A_n, B_n sequences of asymptotically free random matrices (or one deterministic and one random), eig. distr. converging to μ_A, and μ_B, what about eig. distr. of $A_n + B_n$?



When $n \to \infty$, eig. distr. of the sum is (approximately) free additive convolution of μ_A and μ_B.

From a numerical linear algebra point of view: A_n = discretization of some operator, B_n models an unknown error, we can only access $A_n + B_n$...

Motivation #2: Distribution of covariance matrices

- Σ_p covariance matrices, for $p \to \infty$ the eigenvalues of Σ_p have a limit distribution μ.
- Take n samples $\sim N(0, \Sigma_p)$ and put them in matrix Z.
- Consider sample covariance matrix $\widehat{\Sigma}_p = \frac{1}{n} Z^T Z$
- (Also assume that $p/n \to \gamma$)

What is the distribution of $\widehat{\Sigma}_p$ (in the limit $p \to \infty$)?

Free multiplicative convolution of μ with a Marchenko-Pastur distribution

Free additive convolution

Setting: Measure μ with compact support $[a, b]$.

1. Cauchy transform: $G(z) = \int_a^b \frac{1}{z-t} \mathrm{d}\mu(t)$
   Analytic on $\mathbb{C} \backslash [a, b] \cup \{\infty\}$.
2. R-transform: defined by $G\left(R(z) + \frac{1}{z}\right) = z$
   Analytic on a disk around 0.
   If $G'(z) \neq 0$ for all $z \in \mathbb{C} \cup \{\infty\} \backslash [a, b]$ then R is analytic on the whole $G(\mathbb{C} \cup \{\infty\} \backslash [a, b])$.

Sum of freely independent random variables with measures μ_1 and μ_2

=

Free additive convolution of the corresponding measures $\mu := \mu_1 \boxplus \mu_2$

Theorem

Given two measures μ_1 and μ_2,

$$R_{\mu_1 \boxplus \mu_2}(z) = R_{\mu_1}(z) + R_{\mu_2}(z).$$

Numerical computation of free convolution: Existing algorithms

N. R. Rao and A. Edelman (2008). The polynomial method for random matrices. Found. Comput. Math.

Free additive convolution of measures whose Cauchy transform is an algebraic function.

Olver, S., & Nadakuditi, R. R. (2012). Numerical computation of convolutions in free probability theory. arXiv preprint.

Methods for smoothly decaying / Jacobi measures. We use several results in their paper for our (conceptually simpler) algorithm.

E. Dobriban. (2015). Efficient computation of limit spectra of sample covariance matrices. Random Matrices Theory Appl.

Distribution of covariance matrices resulting from (approximate) free multiplicative convolution of point measure with Marchenko-Pastur.

Assumptions on the input measures

Definition

μ is a Jacobi measure if its density has the form

$$\mathrm{d}\mu(x) = (x-a)^\alpha (b-x)^\beta \psi(x)\mathrm{d}x$$

for some $\alpha, \beta > -1$ and $\psi \in C^1([a,b])$. It has sqrt-behavior at the boundary if $\alpha = \beta = \frac{1}{2}$.

Examples:

- Semicircle $\to$ sqrt-behavior
- Marchenko-Pastur $\to$ sqrt-behavior
- Uniform on an interval $\to$ Jacobi

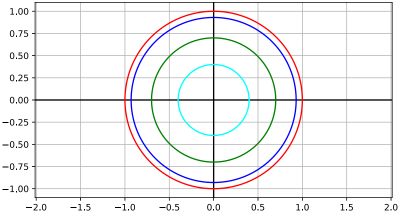
Assumption: We assume that μ_1 and μ_2 have compact support $[a_1, b_1]$ and $[a_2, b_2]$; one of the measures has sqrt-behavior at the boundary (and has an invertible Cauchy transform) and the other one is a Jacobi measure.

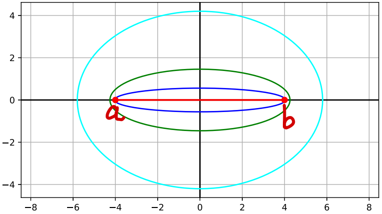
High-level idea of our algorithm:

- Express transforms $G(z)$ and $R(z)$ using Cauchy integral formula.
- Whenever we need to numerically compute an integral, we express it as integral of holomorphic function on the unit circle and use trapezoidal quadrature rule (which converges exponentially fast).

The Joukowski transform

$$J_{[a,b]}(v) = \frac{1}{2}\frac{b-a}{2}\left(v + \frac{1}{v}\right) + \frac{b+a}{2}$$





Conformal map from $\mathbb{D}$ to $\mathbb{C} \cup \{\infty\} \backslash [a, b]$

G analytic $\Rightarrow$ Can define $\tilde{G}$ in the unit disk

$$\tilde{G}(v) := G(J_{[a,b]}(v)) = \sum_{n \geq 1} g_n v^n.$$

1st ingredient:
Evaluation of the Cauchy transform

Evaluation of Cauchy transform

$$G(z) = \int_a^b \frac{1}{z-x} \mathrm{d}\mu(x) = \int_\gamma \frac{1}{2} \frac{b-a}{2} \left(1 - \frac{1}{w^2}\right) \frac{f(J_{[a,b]}(w))}{z - J_{[a,b]}(w)} \mathrm{d}w, \ \gamma \text{ upper unit semicircle}, \ \mathrm{d}\mu(x) = f(x)\mathrm{d}x$$

Numerically, we discretize the second integral with the trapezoidal quadrature rule in N equispaced points.

For measures with sqrt-behavior at the boundary,

$$G(z) = -\frac{1}{2} \int_{\partial\mathbb{D}} \frac{(b-a)^2}{4} \frac{1}{w} \left(w - \frac{1}{w}\right)^2 \frac{\psi(J_{[a,b]}(w))}{z - J_{[a,b]}(w)} \mathrm{d}w.$$

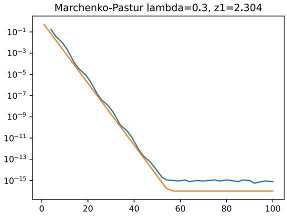
Exponential convergence of trapezoidal quadrature rule

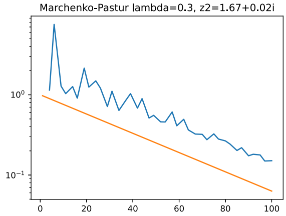
Theorem ([Trefethen/Weideman'2014 (but well known before)])

Suppose u is analytic and satisfies $|u(z)| \leq C$ in an annulus $\rho^{-1} < |z| < \rho$ of the complex plane, for some $\rho > 1$. Let $N > 1$ and consider the approximation I_N of the integral $I := \int_{\partial \mathbb{D}} u(z) \mathrm{d}z$ using N equispaced quadrature points and the trapezoidal quadrature rule. Then

$$|I_N - I| \leq \frac{4\pi C}{\rho^N - 1}.$$

Corollary: Trapezoidal quadrature rule converges exponentially for measures with sqrt-behavior at the boundary, and convergence rate ρ increases when we move further away from the support of μ.

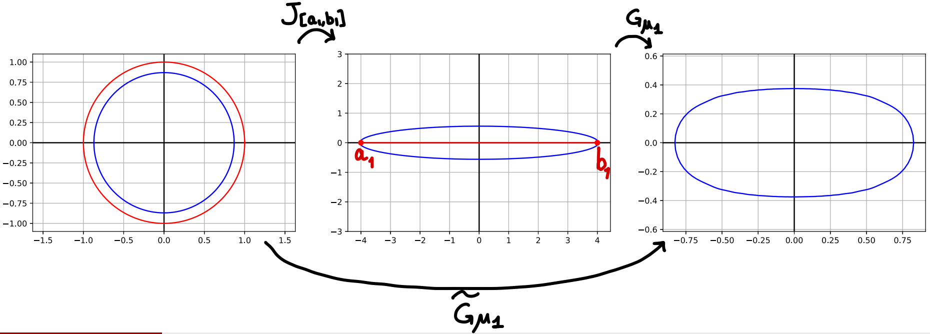




First step of the algorithm

First step of our algorithm: Choose N equispaces points $r\xi_N^0, r\xi_N^1, r\xi_N^2, \ldots, r\xi_N^{N-1}$ on a circle of radius $r < 1$ (but close to 1) and evaluate

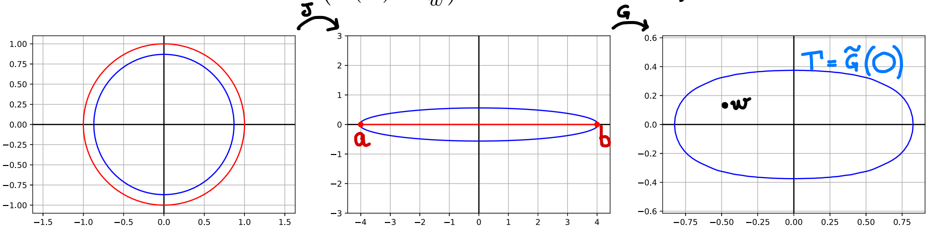
$$\tilde{G}_{\mu_1}(r\xi_N^0), \tilde{G}_{\mu_1}(r\xi_N^0), \ldots, \tilde{G}_{\mu_1}(r\xi_N^0),$$
$$\tilde{G}_{\mu_2}(r\xi_N^0), \tilde{G}_{\mu_2}(r\xi_N^0), \ldots, \tilde{G}_{\mu_2}(r\xi_N^0).$$



2nd ingredient:
Evaluation of the R transform

Numerical computation of R-transform

Recall: $G\left(R(w) + \frac{1}{w}\right) = w$ and R is analytic!



We can express $R(w)$ for w inside the blue curve on the right using Cauchy integral formula:

$$R(w) = \frac{1}{2\pi i} \int_\Gamma \frac{R(z)}{w - z} \mathrm{d}z$$

Convergence result

If Γ is parametrized using the blue circle on the left (radius r), we rewrite this as the integral of a holomorphic function on $\partial\mathbb{D} \Rightarrow$ trapezoidal quadrature rule converges **exponentially**!

$$R(w) = \frac{r}{2\pi i}\int_{\partial\mathbb{D}} \frac{J_{[a,b]}(rv) - 1/\widetilde{G}(rv)}{\widetilde{G}(rv) - w} G'(J_{[a,b]}(rv))J'_{[a,b]}(rv)\mathrm{d}v =: \int_{\partial\mathbb{D}} u(v)\mathrm{d}v.$$

Theorem ([C/Ying'2023])

Let w be a point inside the blue curve Γ. Let $\rho < \min\{r^{-1}, \operatorname{dist}(\tilde{G}^{-1}(z), r\partial\mathbb{D})\}$. Let $\xi_N = \exp(2\pi i/N)$. Assume we have computed approximations $c_j \approx G(J(r\xi_N^j)) = \widetilde{G}(r\xi_N^j)$ and $d_j \approx G'(J(r\xi_N^j))$ such that $|\tilde{G}(r\xi_N^j) - c_j| \le \varepsilon$ and $|G'(J(r\xi_N^j)) - d_j| \le \varepsilon$. Let $m_1 = \min\{\|\tilde{G}\|_\Gamma, c_0, \dots, c_{N-1}\}$. Let m_2 be the distance of z from Γ. Let us discretize the integral with the trapezoidal quadrature rule

$$R(w) \approx \frac{r}{N}\sum_{j=1}^{N} \xi_N^j \cdot d_j \cdot J'(\xi_N^j) \cdot \frac{J(\xi_N^j) - 1/c_j}{c_j - w} =: R_N(w).$$

Then

$$|R_N(w) - R(w)| \le \frac{4\pi}{\rho^N - 1} \max_{\rho^{-1} \le |v| \le \rho} |u(v)|$$

$$+ \varepsilon r \|J'\|_{r\partial\mathbb{D}} \left(\frac{\|J\|_{r\partial\mathbb{D}}}{m_2} + \frac{1}{m_1 m_2} + (\varepsilon + \|G'\|_{J(r\partial\mathbb{D})}) \frac{1}{m_2^2} \left(\|J\|_{r\partial\mathbb{D}} + \frac{|w|}{m_1^2} + \frac{2}{m_1} \right) \right).$$

3rd ingredient:
The support of $\mu := \mu_1 \boxplus \mu_2$

Computing the support of $\mu := \mu_1 \boxplus \mu_2$

Consider measures μ_1 and μ_2 with support in $[a_1, b_1]$ and $[a_2, b_2]$.

$$\text{Let } g(z) := G_{\mu_1}^{-1}(z) + G_{\mu_2}^{-1}(z) - \frac{1}{z} \qquad (= G_\mu(z) \text{ around } 0.)$$

Theorem ([Olver/Nadakuditi'2012])

Assume that μ_1 has sqrt-behavior at the boundary and its Cauchy transform is invertible, and μ_2 has the form $\mathrm{d}\mu_2(t) = (t-a_2)^\alpha (b_2 - t)^\beta \psi_2(t)\mathrm{d}t$ for some $\alpha, \beta > -1$ and $\psi_2 \in C^1[a_2, b_2]$. Then:

- *$\mu_1 \boxplus \mu_2$ has sqrt-behavior at the boundary;*
- *The support of μ is contained in the interval*

$$[a, b] := [g(\xi_a), g(\xi_b)],$$

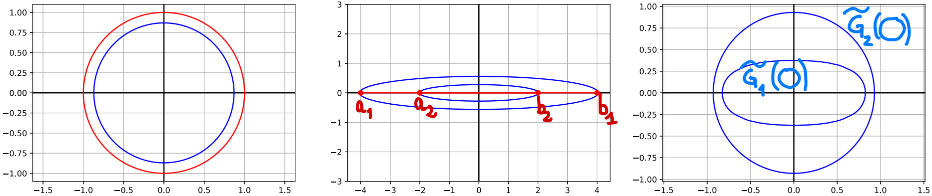
where ξ_a and ξ_b are the unique zeros of the derivative g' in the intervals $(\max(G_{\mu_1}(a_1), G_{\mu_2}(a_2)), 0)$ and $(0, \min(G_{\mu_1}(b_1), G_{\mu_2}(b_2)))$, respectively.

Step 2: Computing the support of $\mu_1 \boxplus \mu_2$

$$g'(z) = \frac{1}{G'_{\mu_1}(G^{-1}_{\mu_1}(z))} + \frac{1}{G'_{\mu_2}(G^{-1}_{\mu_2}(z))} + \frac{1}{z^2}.$$

To compute the support we need to be able to evaluate:

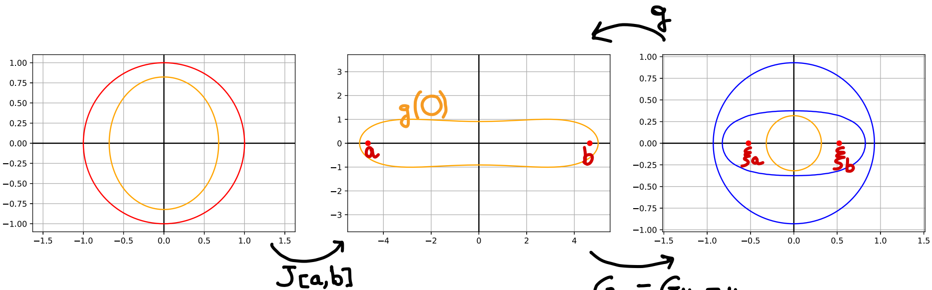
- R-transform $\rightsquigarrow$ by Cauchy integral formula
- $G'(z) = -\int_a^b \frac{1}{(z-t)^2} \mathrm{d}\mu(t)$ (again trapezoidal quadrature rule)



Do binary search! (with caution)

4th ingredient: Computing Cauchy transf. $G_{\mu_1 \boxplus \mu_2}$ (from its R-transform)

Step 3: Finding a circle in the image of $G_{\mu_1 \boxplus \mu_2}$



Want: Orange circle in $G_{\mu_1 \boxplus \mu_2}(\mathbb{C} \cup \{\infty\} \backslash [a,b])$

Theorem ([Olver/Nadakuditi'2011])

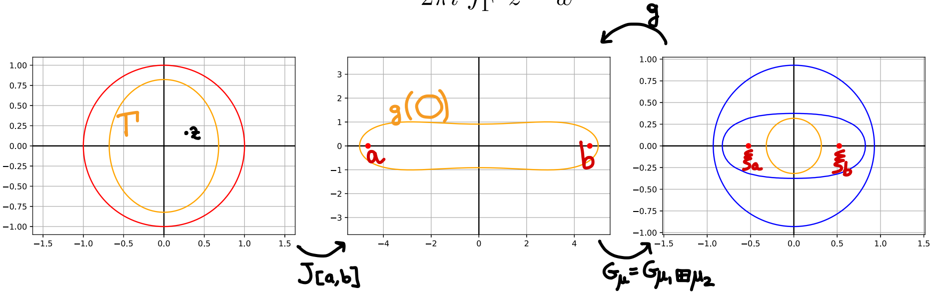
In the same assumptions of previous theorem,

$$z \in G_{\mu_1 \boxplus \mu_2}(\mathbb{C} \backslash [a,b]) \Leftrightarrow \operatorname{sgn}(\operatorname{Im}(g(z))) = -\operatorname{sgn}(\operatorname{Im}(z)).$$

Again, binary search! (on the radius of the orange circle)

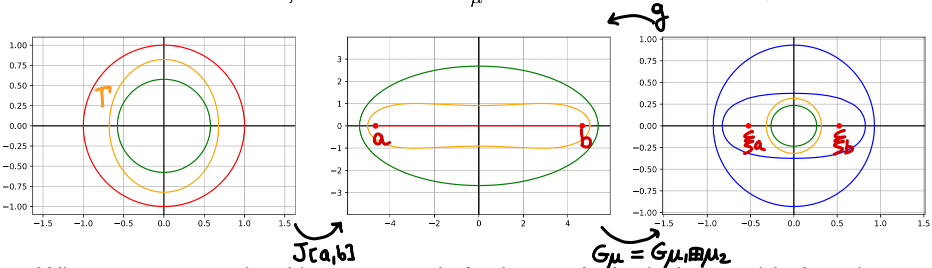
For z inside the orange curve on the left (Γ) we can use Cauchy integral formula again!

$$\tilde{G}_\mu(z) = \frac{1}{2\pi i}\int_\Gamma \frac{\tilde{G}_\mu(w)}{z-w}\mathrm{d}w$$



And we know $\tilde{G}_\mu$ on Γ because we constructed it as $\tilde{G}_\mu^{-1}$(orange circle).

To be able to recover the measure, we evaluate $\tilde{G}_\mu$ on a circle of radius $r < 1$.



When Γ parametrized by orange circle, integral of a holomorphic function
$\Rightarrow$ again, exponential convergence of trapezoidal quadrature rule!

Compute on green circle: $\tilde{G}_\mu(z) = \dfrac{r^2}{2\pi i} \displaystyle\int_{\partial \mathbb{D}} \frac{J_{[a,b]}^{-1}(g(rv))g'(rv)v}{J_{[a,b]}^{-1}(g(rv)) - z} \mathrm{d}v.$

5th ingredient:
Stieltjes inversion
Recovering $\mu_1 \boxplus \mu_2$ from $G_{\mu_1 \boxplus \mu_2}$

Stieltjes inversion theorem

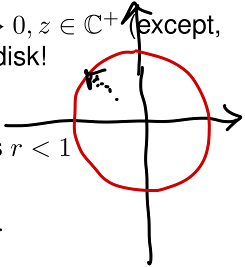
Objective: Given G_μ, recover the density of the measure μ

Theorem

Assume $\mathrm{d}\mu(t) = f(t)\mathrm{d}t$ *for a continuous function* f*. For any* $c < d \in (a, b)$ *we have that*

$$-\frac{1}{\pi}\lim_{\varepsilon\to 0^+}\int_c^d \mathrm{Im}\left(G_\mu(x+i\varepsilon)\right)\mathrm{d}x = \mu([c,d]).$$

Remark: Theorem also works when considering the limit of $G_\mu(x+z)$ with $z \to 0, z \in \mathbb{C}^+$ (except, possibly, in a and b) $\rightsquigarrow$ can consider limit for $\tilde{G}_\mu(z)$ when getting closer to unit disk!



Recall: $G_\mu(z) = \int_a^b \frac{1}{z-w}\mathrm{d}w$ and $\tilde{G}_\mu(z) = G_\mu(J_{[a,b]}(z)) = \sum_{n=1}^\infty g_n z^n$.
We know the value of $\tilde{G}_\mu$ on M equispaced points on the green circle of radius $r < 1$

$$p_j = r\xi_M^j, \qquad \xi_M = \exp\left(\frac{2\pi i}{M}\right), \qquad j = 0, 1, 2, \ldots, M-1.$$

Final step of the algorithm

If we truncate the power series corresponding to $\tilde{G}_\mu$ to the first M terms we get

$$\begin{bmatrix} \tilde{G}_\mu(p_0) \\ \tilde{G}_\mu(p_1) \\ \vdots \\ \tilde{G}_\mu(p_{M-1}) \end{bmatrix} = \begin{bmatrix} \xi_M^0 & & & \\ & \xi_M^1 & & \\ & & \ddots & \\ & & & \xi_M^{M-1} \end{bmatrix} \cdot F \cdot \begin{bmatrix} g_1 r \\ g_2 r^2 \\ \vdots \\ g_M r^M \end{bmatrix},$$

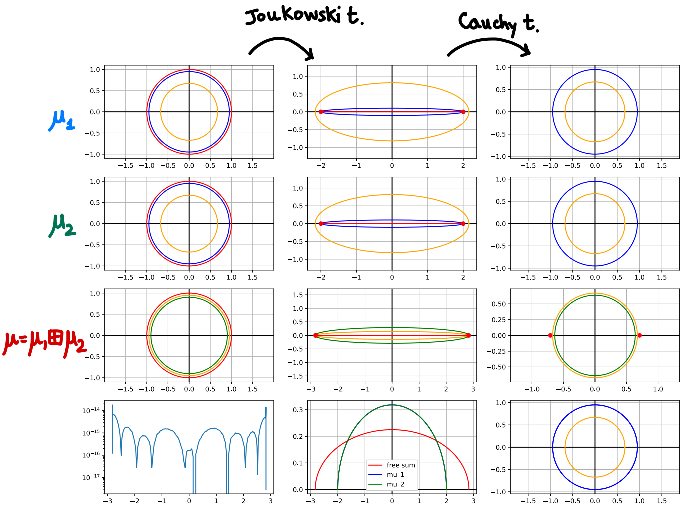
where F is the Fourier matrix of size $M \times M$.

$$\begin{bmatrix} f\left(\frac{b-a}{2}\cos\left(\frac{0\pi}{m}\right)+\frac{b+a}{2}\right) \\ f\left(\frac{b-a}{2}\cos\left(\frac{2\pi}{m}\right)+\frac{b+a}{2}\right) \\ \vdots \\ f\left(\frac{b-a}{2}\cos\left(\frac{2(m-1)\pi}{m}\right)+\frac{b+a}{2}\right) \end{bmatrix} \approx -\frac{1}{\pi}\mathrm{Im}\left(\begin{bmatrix} \xi_m^0 & & & \\ & \xi_m^1 & & \\ & & \ddots & \\ & & & \xi_M^{m-1} \end{bmatrix} \cdot F \cdot \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix}\right).$$

$\Rightarrow$ recover $g_1, \dots, g_M$ from IFFT and then density via FFT.

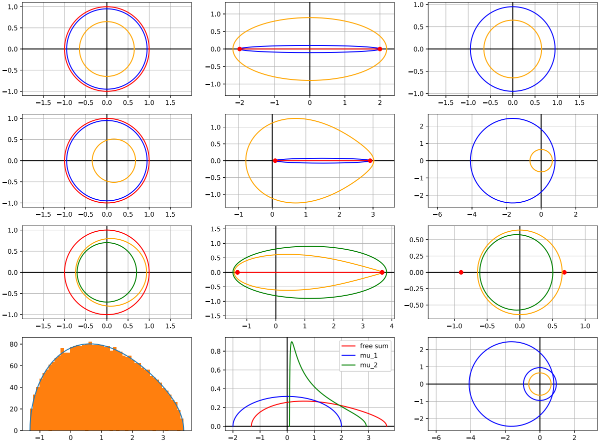
Numerical examples

Example #1: Semicircle + semicircle



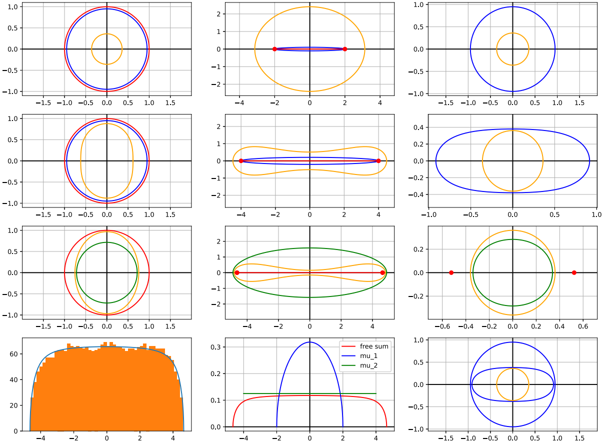
1. Compute Cauchy transforms $\widetilde{G}_{\mu_1}$ and $\widetilde{G}_{\mu_2}$ of blue circle so that we are able to compute R-transforms for μ_1 and μ_2
2. Find support of $\mu_1 \boxplus \mu_2$
3. Find orange circle inside images of G_{μ_1}, G_{μ_2}, and $G_{\mu_1 \boxplus \mu_2}$
4. Evaluate $\tilde{G}_{\mu_1 \boxplus \mu_2}$ on green circle inside $\tilde{G}^{-1}_{\mu_1 \boxplus \mu_2}$ (orange circle)
5. Recover measure from Stieltjes inversion

Example #2: Semicircle + Marchenko-Pastur ($\lambda = \frac{1}{2}$)



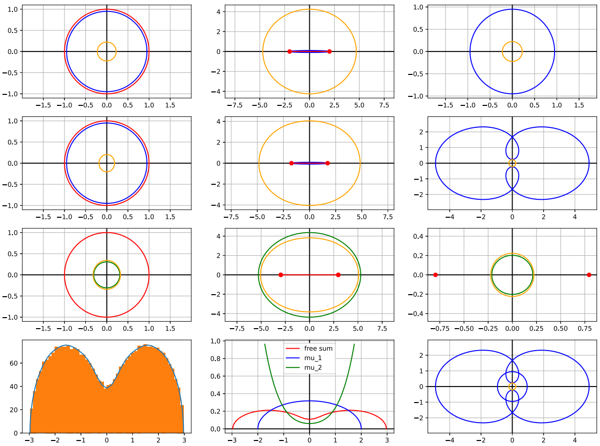
In this case we don't know the analytic expression of the resulting distribution, but we can compare it with the eigenvalue distribution of the sum of two random 3000×3000 matrices.

Example #3: Semicircle + Uniform



Uniform distribution does not have sqrt-behavior at the boundary, therefore quadrature rules converge more slowly. Algorithm works nonetheless.

Example #4: Semicircle + something weird



Consider distribution μ_2 with support on $[-\sqrt{3}, \sqrt{3}]$ and density $\mathrm{d}\mu_2(t) = \frac{5\sqrt{3}}{144}(t^2+1)^2\mathrm{d}t$. $G_2(z)$ not invertible, but not a problem for the algorithm (need to choose a small orange circle though).

Free multiplicative convolution

Theory for free multiplicative convolution

Definition

The T-transform and S-transform of a measure μ are

$$T(z) = \int_a^b \frac{t}{z-t} \mathrm{d}\mu(t) \quad \text{and} \quad S(w) = \frac{1+w}{wT^{-1}(w)},$$

respectively.

These, again, are analytic transforms!

Theorem

Given two measures μ_1 and μ_2,

$$S_{\mu_1 \boxtimes \mu_2}(w) = S_{\mu_1}(w) \cdot S_{\mu_2}(w).$$

Brief summary of the algorithm

1. Compute ~~Cauchy~~ T transforms T_{μ_1} and T_{μ_2} of circle of radius $r_A < 1$ so that we are able to compute ~~R-transforms~~ S-transforms for μ_1 and μ_2
2. Find support of ~~$\mu_1 \boxplus \mu_2$~~ $\mu_1 \boxtimes \mu_2$
3. Find circle inside images of T_{μ_1}, T_{μ_2}, and $T_{\mu_1 \boxplus \mu_2}$
4. Evaluate $\tilde{T}_{\mu_1 \boxplus \mu_2}$ on a suitable circle
5. Recover measure from a similar version of the Stieltjes inversion theorem.

Crucial step: Computing support of free multiplicative convolution

Assumptions: μ_1 and μ_2 satisfy the following properties.

- They have compact support $[a_1, b_1]$ and $[a_2, b_2]$, with $a_1, b_1, a_2, b_2 > 0$.
- Both μ_1 and μ_2 have sqrt-behavior at the boundary.
- The T-transforms T_{μ_1} and T_{μ_2} are invertible on their domain of definition.

Theorem ([C./Ying'2023])

Let us define

$$t(w) := \frac{w}{1+w} T_1^{-1}(w) T_2^{-1}(w),$$

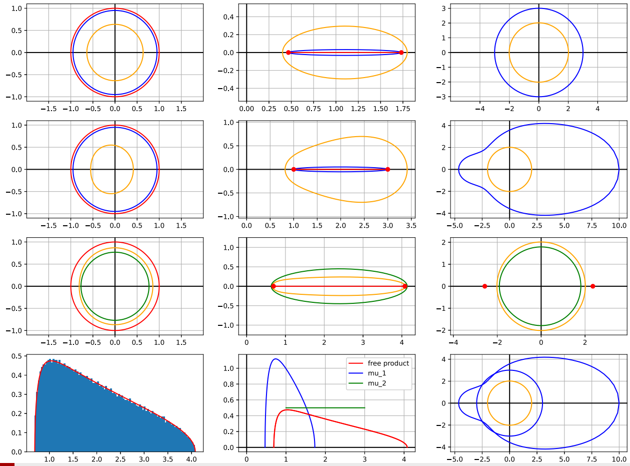
which coincides with $T_\mu^{-1}(w)$ in a neighborhood of zero. Then:

- $\mu := \mu_1 \boxtimes \mu_2$ *has sqrt-behavior at the boundary.*
- *The support of μ is contained in the interval $[a, b] := [t(\xi_a), t(\xi_b)]$, where ξ_a and ξ_b are the unique zeros of the derivative t' in the intervals $(\max(T_1(a_1), T_2(a_2)), 0)$ and $(0, \min(T_1(b_1), T_2(b_2)))$, respectively.*
- *To check whether a point w is in the image of $T_{\mu_1 \boxtimes \mu_2}$ we can use the following criterion:*

$$w \in T_{\mu_1 \boxplus \mu_2}(\mathbb{C} \backslash [a, b]) \Leftrightarrow \operatorname{sgn}(\operatorname{Im}(t(w))) = -\operatorname{sgn}(\operatorname{Im}(w)).$$

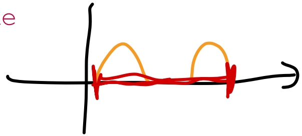
Numerical example

Free multiplicative convolution of Marchenko-Pastur distr. with $\lambda = 0.1$ and uniform distr. on $[1, 3]$



Conclusion

- Main ideas: Express the analytic transforms that define free convolutions with Cauchy integral formula, trapezoidal quadrature rule converges fast.
- Open questions:
  - How to reliably compute support for more general distributions?
  - What if it is not supported on a single interval?
  - Can the method be extended to polynomial/rational functions in free random variables?
- Python code is available on Github at https://github.com/Alice94/FreeConvolutionCode



Alice Cortinovis and Lexing Ying (2023).
Computing free convolutions via Cauchy integrals.
https://arxiv.org/abs/2305.01819

Thank you!

alicecor@stanford.edu