

Computing Free Convolutions via Contour Integrals

Alice Cortinovis*

Lexing Ying†

Abstract

This work proposes algorithms for computing additive and multiplicative free convolutions of two given measures. We consider measures with compact support whose free convolution results in a measure with a density function that exhibits a square-root decay at the boundary (for example, the semicircle distribution or the Marchenko-Pastur distribution). A key ingredient of our method is rewriting the intermediate quantities of the free convolution using the Cauchy integral formula and then discretizing these integrals using the trapezoidal quadrature rule, which converges exponentially fast under suitable analyticity properties of the functions to be integrated.

1 Introduction

Given two independent commutative random variables X and Y , the density of the sum can be obtained by computing the convolution of the densities of X and Y . This no longer holds when X and Y do not commute. As an example, consider two symmetric matrices X and Y of size $n \times n$ whose eigenvalues are ± 1 , each chosen with probability $\frac{1}{2}$, and whose eigenvector matrices are two independently chosen random orthogonal matrix (that is, from the Haar distribution). When the size n grows to infinity, the eigenvalues of $X + Y$ tend to follow an arcsine distribution – a continuous distribution that does not correspond to the convolution of the random variables that define the eigenvalues of X and Y . The theory of free probability, introduced by Voiculescu in the 1980s, is a powerful tool to explain the behavior of the eigenvalue distribution of the sum of random matrices via the concept of *free additive convolution*; we refer the reader to the book [7] for an explanation of the topic and its connections with random matrix theory.

Another application of free probability is the following. When Z is a random matrix of size $n \times p$ whose rows are zero-mean random vectors with variance $\Sigma_p \in \mathbb{R}^{p \times p}$, one can consider the sample covariance $\hat{\Sigma}_p := \frac{1}{n} Z^T Z$. If we consider increasing values of n and p with $\frac{p}{n} \rightarrow \gamma$, and the eigenvalues of Σ_p have a limit distribution, then the eigenvalue distribution of $\hat{\Sigma}_p$ also has a deterministic limit, under mild conditions [10, 11]. Again, such a limit can be expressed using the tools of *free multiplicative convolution*.

Related literature. The free additive and multiplicative convolutions of measures are defined, theoretically, via Cauchy transforms, R-transforms, T-transforms, and S-transforms, which will be recalled in Section 2. It is possible to compute the free convolution analytically only in some very special cases; for all other measures, one must resort to numerical techniques. An approach for the numerical computation of free convolutions is to use fixed-point iterations, for example,

*Department of Mathematics, Stanford University, CA, USA. Email: alicecor@stanford.edu

†Department of Mathematics and ICME, Stanford University, CA, USA. Email: lexing@stanford.edu

based on the results in [5, Theorem 6.5]. The work [3] proposes an algorithm for computing the free multiplicative convolution of a sum of point measures with a Marchenko-Pastur distribution (which fits into the application of free multiplicative convolution mentioned above). The work [9] addresses the free additive convolution of measures whose Cauchy transform is an algebraic function. The work [8] proposes an algorithm for the free convolution of measures that result in a smoothly-decaying measure supported on the entire real line or a measure that is supported on a compact interval and exhibits a square-root behavior at the boundary; see Remark 3.6 for a comparison with our approach.

Contributions. In this work, we propose a new algorithm for computing the free (additive and multiplicative) convolution of two measures μ_1 and μ_2 . We assume that μ_1 and μ_2 have compact support and that their free convolution has a square-root behavior at the boundary, which is made precise in Definition 2.1. Our approach is based on approximating the analytic transforms needed in the computation of the free convolution using the Cauchy integral theorem along suitably defined curves. The numerical evaluation of integrals is done using the trapezoidal quadrature rule, which results in a fast convergence with respect to the number of quadrature points because the functions to be integrated are analytic.

Outline. The rest of the paper is organized as follows. In Section 2, we recall the relevant definitions and results on the free convolution of measures. In Section 3, we outline the proposed algorithm for free additive convolution. A similar algorithm for the free multiplicative convolution is summarized in Section 4. In Section 5, we discuss the approximation errors in the algorithms for free convolutions. We report various numerical examples in Section 6, and the conclusions are given in Section 7.

2 Preliminaries on free convolution

In this work, we consider measures with compact support and with a continuous density without atoms.

Notation. We denote by μ a measure with support in $[a, b]$ and density $f(x)$. The input measures are μ_1 (with support $[a_1, b_1]$) and μ_2 (with support $[a_2, b_2]$). We denote by $\mu_1 \boxplus \mu_2$ their free additive convolution and by $\mu_1 \boxtimes \mu_2$ their free multiplicative convolution. We denote \mathbb{D} the open unit disk in \mathbb{C} and $\partial\mathbb{D}$ the unit circle (the border of the unit disk) in \mathbb{C} . When computing integrals on $\partial\mathbb{D}$ or any other curve in the complex plane, these are intended in an anti-clockwise direction. \mathbb{C}^+ and \mathbb{C}^- denote the half of the complex plane containing numbers with strictly positive or negative imaginary parts, respectively.

2.1 Measures with sqrt-behavior at the boundary and Jacobi measures

The most regular distributions we consider in this paper are those with an sqrt-behavior at the boundary in the following sense.

Definition 2.1. *The measure μ has square-root behavior at the boundary (sqrt-behavior) if its density has the form*

$$d\mu(x) = f(x)dx = \psi(x)\sqrt{x-a}\sqrt{b-x}dx$$

for some $\psi \in C^1([a, b])$ with $\psi'(x)$ of bounded variation.

We recall two examples of measures of this type that arise naturally from random matrix theory.

Example 2.2 (Semicircle). *The semicircle distribution has support in $[a, b] = [-2, 2]$ and its density is*

$$d\mu(x) = \frac{1}{2\pi} \sqrt{4 - x^2} dx.$$

It has sqrt-behavior at the boundary with $\psi(x) = \frac{1}{2\pi}$.

The semicircle distribution is the limit (for $n \rightarrow \infty$) eigenvalue distribution of a random symmetric matrix of size $n \times n$ that has random i.i.d. entries with mean 0 and variance $\frac{1}{n}$ above the diagonal, and i.i.d. entries with mean 0 and variance $\frac{C}{n}$ on the diagonal, for some constant $C \geq 0$.

Example 2.3 (Marchenko-Pastur). *The Marchenko-Pastur distribution with parameter $\lambda \in (0, 1)$ has support in $[\lambda_-, \lambda_+]$ with $\lambda_{\pm} := (1 \pm \sqrt{\lambda})^2$, and its density is*

$$d\mu(x) = \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{2\pi\lambda x} dx.$$

It has sqrt-behavior at the boundary, with $\psi(x) = \frac{1}{2\pi\lambda x}$.

The Marchenko-Pastur distribution is the limit (for $n \rightarrow \infty$) eigenvalue distribution of random matrices of the form $B_n B_n^T$ for matrices $B_n \in \mathbb{R}^{[n] \times n}$ made of i.i.d. entries with zero mean and variance $\frac{1}{n}$.

A more general class of measures is the following.

Definition 2.4. μ is a Jacobi measure if its density has the form

$$d\mu(x) = (x - a)^\alpha (b - x)^\beta \psi(x) dx$$

for some $\alpha, \beta > -1$ and $\psi \in C^1([a, b])$.

Example 2.5 (Uniform distribution). *As an example of a Jacobi measure that does not have an sqrt-behavior at the boundary (that we will use in our numerical experiments), let us consider the uniform distribution with support in the interval $[a, b]$, for some $a, b \in \mathbb{R}$. The density is $d\mu(x) = \frac{1}{b-a} dx$.*

2.2 The Cauchy transform and Stieltjes inversion

Definition 2.6. *The Cauchy transform of the point $z \in \mathbb{C} \setminus [a, b]$ is defined as*

$$G(z) = \int_a^b \frac{1}{z - x} d\mu(x).$$

When ambiguous, we will put a subscript indicating the measure, for example, $G_\mu(z)$.

Theorem 2.7. *The Cauchy transform G is analytic on $\mathbb{C} \setminus [a, b]$ and at infinity.*

This is a version of Proposition A.7 in [8] with different assumptions, and it is slightly better than Lemma 2 in Chapter 3 in [7] because we have stronger assumptions in our case (i.e. compact support). The proof follows [7, Chapter 3, Lemma 2]; we report the proof in Appendix A for completeness.

To see the Cauchy transform as a map from the unit disk in \mathbb{C} to \mathbb{C} , we use the Joukowski map.

Definition 2.8. *The Joukowski map relative to the segment $[a, b]$ is*

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

This is a conformal map from the unit disk to $\mathbb{C} \cup \{\infty\} \setminus [a, b]$. The unit circle is mapped into the segment $[a, b]$ (“twice”), and circles are mapped into ellipses. For $z \in \mathbb{C} \cup \{\infty\} \setminus [a, b]$ we denote by $J_{[a,b]}^-$ and $J_{[a,b]}^+$ the inverse Joukowski functions which are inside and outside the unit disk, respectively. The Joukowski map allows us to define

$$\mathcal{G}(v) := G(J_{[a,b]}(v)),$$

which is a holomorphic function that maps the unit disk into \mathbb{C} , and such that $\mathcal{G}(0) = 0$. Therefore, we can write it as a power series

$$\mathcal{G}(v) = \sum_{n \geq 1} g_n v^n \tag{1}$$

that converges everywhere inside \mathbb{D} .

Remark 2.9. *If the measure μ has sqrt-behavior at the boundary, the series (1) converges also on the unit circle. Indeed, as discussed in [8, Section 3.0.1], the coefficients are $g_n = \frac{(b-a)\pi}{2} \psi_{n-1}$, where ψ_n is the n -th coefficient in the series expansion of $\psi(x)$ with Chebyshev polynomials of the second kind. When $\psi(x)$ is analytic in a (complex) neighborhood of the interval $[a, b]$, the coefficients ψ_n decay exponentially fast; see, e.g., [12, Theorem 8.1]. To obtain convergence of (1) on the unit circle, it is sufficient to assume that $\psi(x)$ is Lipschitz-continuous, see, e.g., [12, Theorem 3.1].*

Example 2.10. *For the semicircle and Marchenko-Pastur distribution, we can write the Cauchy transform explicitly (and their inverse and $\mathcal{G}(z)$ as well). For the semicircle distribution, we have*

$$G(z) = \frac{z - \sqrt{z^2 - 4}}{2}, \quad G^{-1}(w) = w + \frac{1}{w}, \quad \mathcal{G}(v) = v.$$

For the Marchenko-Pastur distribution with parameter λ , we have

$$G(z) = \frac{z + \lambda - 1 - \sqrt{(z - \lambda - 1)^2 - 4\lambda}}{2\lambda z}, \quad G^{-1}(w) = \frac{1}{1 - \lambda w} + \frac{1}{w}, \quad \mathcal{G}(v) = \frac{v}{\lambda(1 - v\sqrt{\lambda})}.$$

In particular, note that \mathcal{G} is analytically continuable in the disk of radius $\frac{1}{\sqrt{\lambda}} > 1$. For the uniform distribution on the interval $[-m, m]$, instead, we have that

$$G(z) = \frac{1}{2m} \log \frac{z+m}{z-m}, \quad G^{-1}(w) = -m + \frac{2m}{1 - \exp(-2mw)}.$$

We have

$$\mathcal{G}(v) = \frac{1}{m} \log \left(1 + \frac{2}{1-v} \right) = \frac{2}{m} \sum_{n \geq 0} \frac{v^{2n+1}}{2n+1},$$

which means that \mathcal{G} is not analytically continuable on the border of the unit disk.

The Stiltjes inversion formula allows us to recover the density of a distribution from its Cauchy transform.

Theorem 2.11 ([7, Theorem 6]). *Assume $d\mu(x) = f(x)dx$ for a continuous function f . For any $c < d \in (a, b)$ we have that*

$$-\frac{1}{\pi} \lim_{y \rightarrow 0^+} \int_c^d \operatorname{Im}(G(x + iy)) dx = \mu([c, d]).$$

In our algorithm, we will use the following corollary, whose proof we report in Appendix A, for completeness.

Corollary 2.12. *Let $d\mu(x) = f(x)dx$ for a continuous function f . Then for all $\theta \in (0, \pi)$ we have*

$$f\left(\frac{b-a}{2} \cos \theta + \frac{b+a}{2}\right) = \frac{1}{\pi} \lim_{r \rightarrow 1^-} \operatorname{Im}\left(\mathcal{G}(re^{i\theta})\right).$$

2.3 The free additive convolution

The key function needed to define the free additive convolution of measures is the R-transform, which is defined as follows.

Definition 2.13. *The R-transform of a measure μ is a function R defined in a neighborhood of 0 such that $G(R(w) + \frac{1}{w}) = w$.*

When the Cauchy transform G is invertible, we have $R(w) = G^{-1}(w) - \frac{1}{w}$. When we need to avoid confusion, we will denote the R-transform of the measure μ by $R_\mu(w)$.

Theorem 2.14. *Let μ be a measure with compact support in $[a, b]$. Then its R-transform is analytic on $G(\mathbb{R} \setminus [a, b])$ and on a disk centered in the origin with radius $\frac{1}{6 \min\{|a|, |b|\}}$.*

If $G'(z) \neq 0$ for all $z \in \mathbb{C} \cup \{\infty\} \setminus [a, b]$ then R is invertible on $G(\mathbb{C} \cup \{\infty\} \setminus [a, b])$ and it is analytic on the whole $G(\mathbb{C} \cup \{\infty\} \setminus [a, b])$. We have the following result for the free additive convolution of μ_1 and μ_2 .

Theorem 2.15 ([7, Chapter 2, Theorem 18]). *Given two measures μ_1 and μ_2 ,*

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

2.4 The free multiplicative convolution

For two measures μ_1 and μ_2 with compact support, analogously to the additive case, a suitable transform (S-transform) allows us to compute the free multiplicative convolutions.

Definition 2.16. *The T-transform and S-transform of a measure μ are*

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

respectively.

The T-transform maps \mathbb{C}^+ into \mathbb{C}^- and maps \mathbb{C}^- into \mathbb{C}^+ , and it is a holomorphic function from $\mathbb{C} \cup \{\infty\} \setminus [a, b]$ to \mathbb{C} . Therefore, the map $\mathcal{T}(v) := T(J(v))$ has a power expansion $\mathcal{T}(v) = \sum_{n \geq 1} t_n v^n$ that converges inside the unit disk. The S-transform is analytic in a neighborhood of 0. We skip the details because this is similar to the additive case.

Theorem 2.17 ([7, Chapter 4, Theorem 23]). *Given two measures μ_1 and μ_2 ,*

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

3 An algorithm for free additive convolution

In this section, we propose an efficient algorithm for computing the free additive convolution of two measures μ_1 and μ_2 that satisfy the following set of assumptions.

Assumptions 3.1. *We assume that μ_1 and μ_2 satisfy the following properties.*

- *They have compact support $[a_1, b_1]$ and $[a_2, b_2]$.*
- *One of them, without loss of generality μ_1 , has sqrt-behavior at the boundary.*
- *The other one, without loss of generality μ_2 , is a Jacobi measure.*
- *The Cauchy transforms G_{μ_1} and G_{μ_2} are invertible on their domain of definition.*

We recall that the Cauchy integral formula states that, for an analytic function f and a contour Γ , for any point z inside Γ

$$f(z) = \frac{1}{2\pi i} \int_{\Gamma} \frac{f(w)}{w - z} dw.$$

This is a tool that we will use frequently in our algorithm, to evaluate Cauchy transforms and R-transforms (which are analytic functions).

Another central theoretical result for our algorithm is the following theorem that allows us to characterize the support of $\mu_1 \boxplus \mu_2$ and the image of $G_{\mu_1 \boxplus \mu_2}$.

Theorem 3.2 (Theorem 2.2 and Theorem 5.2 in [8]). *Under the Assumptions 3.1, let*

$$g(w) := G_{\mu_1}^{-1}(w) + G_{\mu_2}^{-1}(w) - \frac{1}{w}, \tag{2}$$

which coincides with $G_{\mu}^{-1}(w)$ in a neighborhood of 0. We have the following properties.

- *$\mu := \mu_1 \boxplus \mu_2$ 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

$$(\max(G_{\mu_1}(a_1), G_{\mu_2}(a_2)), 0) \quad \text{and} \quad (0, \min(G_{\mu_1}(b_1), G_{\mu_2}(b_2))),$$

respectively.

- To check whether a point w is in the image of $G_{\mu_1 \boxplus \mu_2}$ we can use the following criterion:

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

Our proposed algorithm consists of four main steps:

1. Set up the computation of the R-transform for μ_1 and μ_2 .
2. Compute the support of μ
3. Compute the Cauchy transform of μ on a suitable set of points.
4. Recover the density of μ from its Cauchy transform.

3.1 Step 1: Setting up the computation of the R-transform

In this section, we explain how to set up the computation of the R-transform for a measure μ with support in $[a, b]$. We apply this procedure to μ_1 and μ_2 . Let us consider the curve $\Gamma := \mathcal{G}(r_A \partial \mathbb{D})$, for some $r_A \in (0, 1)$. Since the R-transform is analytic, given a point w inside Γ we can write $R(w)$ with the Cauchy integral formula:

$$R(w) = \frac{1}{2\pi i} \int_{\Gamma} \frac{R(s)}{w - s} ds.$$

As Γ is parameterized by $s = \mathcal{G}(r_A v)$ for v on the unit circle, we have $ds = r_A \mathcal{G}'(r_A v) dv$ and $G^{-1}(s) = J_{[a, b]}(r_A v)$ and the integral becomes

$$R(w) = \frac{r_A}{2\pi i} \int_{\partial \mathbb{D}} \frac{J_{[a, b]}(r_A v) - 1/\mathcal{G}(r_A v)}{\mathcal{G}(r_A v) - w} G'(J_{[a, b]}(r_A v)) J'_{[a, b]}(r_A v) dv =: \int_{\partial \mathbb{D}} u(v) dv. \quad (3)$$

We perform an additional change of variables $v = \exp(2\pi i \theta)$, for $\theta \in [0, 1]$, and aim at discretizing (3) with the trapezoidal quadrature rule, which reads

$$R(w) \approx \frac{r_A}{N} \sum_{j=0}^{N-1} \xi_N^j \cdot G'(J_{[a, b]}(\xi_N^j)) \cdot J'_{[a, b]}(\xi_N^j) \cdot \frac{J_{[a, b]}(\xi_N^j) - \frac{1}{G(J_{[a, b]}(\xi_N^j))}}{G(J_{[a, b]}(\xi_N^j)) - w}, \quad \xi_N = \exp\left(\frac{2\pi i}{N}\right).$$

Note that the quantities $G(J_{[a, b]}(\xi_N^j))$ and $G'(J_{[a, b]}(\xi_N^j))$ do not depend on the point w and can be computed ahead of time. For their computation, we again use the trapezoidal quadrature rule. More precisely, for evaluating the Cauchy transform $G(z)$ for a point $z \in \mathbb{C} \setminus [a, b]$, we rewrite the integral that defines it as

$$G(z) = \int_{\gamma} \frac{1}{2} \frac{b-a}{2} \left(1 - \frac{1}{v^2}\right) \frac{f(J_{[a, b]}(v))}{z - J_{[a, b]}(v)} dv,$$

where γ is the upper unit semicircle in the complex plane, oriented clockwise, where we made the change of variables $x = J_{[a, b]}(v)$. With an additional change-of-variable $v = \exp(2\pi i \theta)$ for $\theta \in [0, 1]$ we get

$$G(z) = \pi \frac{b-a}{2} \int_1^0 \sin(2\pi \theta) \frac{f\left(\frac{b-a}{2} \cos(2\pi \theta) + \frac{b+a}{2}\right)}{z - \left(\frac{b-a}{2} \cos(2\pi \theta) + \frac{b+a}{2}\right)} d\theta. \quad (4)$$

The discretization in N quadrature points using the trapezoidal quadrature rule reads:

$$G(z) \approx -\frac{\pi}{N} \sum_{k=0}^N {}'' \frac{b-a}{2} \sin\left(\frac{2\pi k}{N}\right) \frac{f\left(\frac{b-a}{2} \cos\left(\frac{2\pi k}{N}\right) + \frac{b+a}{2}\right)}{z - \left(\frac{b-a}{2} \cos\left(\frac{2\pi k}{N}\right) + \frac{b+a}{2}\right)}, \quad (5)$$

where the $''$ denotes that the first and last terms are halved. This makes sense for any measure μ , but it has a provably exponential convergence when μ has sqrt-behavior at the boundary (see Corollary 5.2 below). Now note that the derivative of the Cauchy transform can be written as

$$G'(z) = - \int_a^b \frac{1}{(z-x)^2} d\mu(x)$$

and this integral can be discretized in the same way as the Cauchy transform itself.

Practically speaking, the first step of our proposed algorithm consists in choosing a radius $r_A < 1$ (for example, $r_A = 0.95$), a number of quadrature points N (for example, $N = 1000$), and approximating

$$c_j^{(k)} \approx G_{\mu_k}(J_{[a_k, b_k]}(\xi_N^j)), \quad d_j^{(k)} \approx G'_{\mu_k}(J_{[a_k, b_k]}(\xi_N^j))$$

for $j = 0, 1, \dots, N$ and for $k = 1, 2$ via (5). We call Γ_1 and Γ_2 the images of the circle of radius r_A by \mathcal{G}_{μ_1} and \mathcal{G}_{μ_2} , respectively. This allows us to compute the R-transform for μ_1 and μ_2 , when needed, from the expression

$$R_{\mu_k}(w) \approx \frac{r_A}{N} \sum_{j=0}^{N-1} \xi_N^j \cdot d_j^{(k)} \cdot J'_{[a_k, b_k]}(\xi_N^j) \cdot \frac{J_{[a_k, b_k]}(\xi_N^j) - 1/c_j^{(k)}}{c_j^{(k)} - w}, \quad k \in \{1, 2\}. \quad (6)$$

Example 3.3. Throughout this section, we will illustrate the behavior of our proposed algorithm on an explicit example: we consider μ_1 to be a Marchenko-Pastur distribution with parameter $\lambda = 0.5$ and μ_2 to be the semicircle distribution. In Figure 1, we show, in the left block, a *blue* circle of radius $r_A = 0.95$, in the middle block the image of this circle via the Joukowski transform associated with μ_1 , and in the right block the image of this curve (an ellipse) via the Cauchy transform G_{μ_1} . The *red* circle in the left plot is the unit circle. In Figure 2, we do the same for μ_2 . The number of discretization points used in the expression (5) is $N = 400$ in these examples.

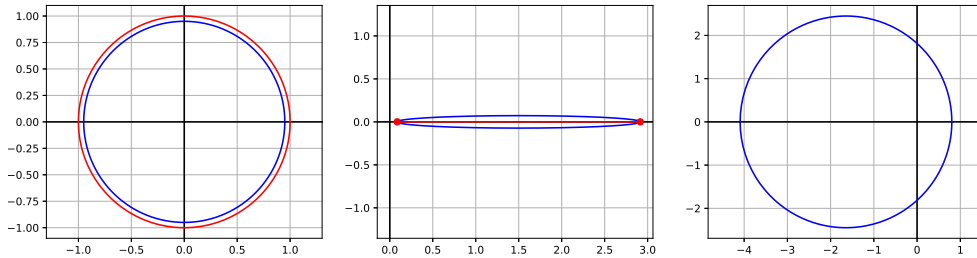


Figure 1: Computation of the Cauchy transform of an ellipse corresponding to a circle of radius $r_A = 0.95$ for the Marchenko-Pastur distribution with parameter $\lambda = 0.5$ (see Example 3.3).

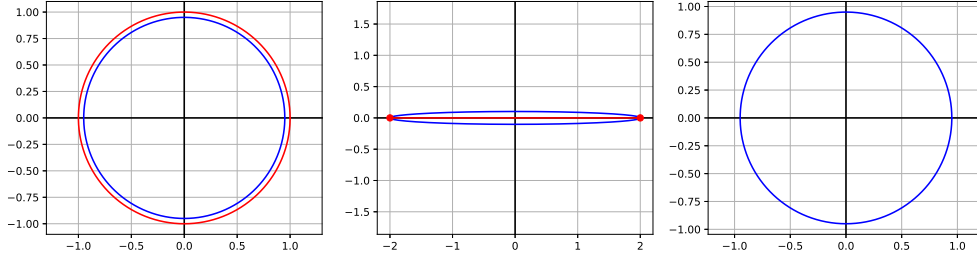


Figure 2: Computation of the Cauchy transform of an ellipse corresponding to a circle of radius $r_A = 0.95$ for the semicircle distribution (see Example 3.3).

3.2 Step 2: Computing the support of the free sum

The second point in Theorem 3.2 gives a clear characterization of the support of $\mu = \mu_1 \boxplus \mu_2$ we use a result from [8]. Note that

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

so it can be approximated numerically using (6) and (5) when w is inside both Γ_1 and Γ_2 . In our algorithm, we use bisection to numerically find a zero of $g'(w)$ in the interval $((1 - \varepsilon) \max\{c_{N/2}^{(1)}, c_{N/2}^{(2)}\}, 0)$ and one in the interval $(0, (1 - \varepsilon) \min\{c_0^{(1)}, c_0^{(2)}\})$, where the parameter ε is needed because when w is very close to either Γ_1 or Γ_2 the computations become unstable; see Section 5 for a better insight on this problem. We stop the binary search when the size of the interval is smaller than a fixed tolerance. We denote by ξ_a and ξ_b the zeros that we found numerically. Then, we compute $a \approx g(\xi_a)$ and $b \approx g(\xi_b)$, and these are (up to numerical errors) the extrema of the support of $\mu = \mu_1 \boxplus \mu_2$.

3.3 Step 3: Computation of the Cauchy transform of μ

The next step in our algorithm is finding a circle of radius r_B that is entirely contained in $G_\mu(\mathbb{C} \cup \{\infty\} \setminus [a, b])$. We want r_B to be as large as possible, for reasons that will be clarified in Section 5. An upper bound for r_B is the radius of the largest circle that is contained inside the intersection of Γ_1 and Γ_2 . Given a point w inside this intersection, the criterion from Theorem 3.2 allows us to check whether w is in the image of G_μ or not. Therefore, we can perform a binary search on the radius and find a suitable value of r_B .

Consider the curve $\Gamma := g(r_B \partial \mathbb{D}) = G_\mu^{-1}(r_B \partial \mathbb{D})$. Leveraging the analyticity of G_μ , for a point z inside Γ we can write $G_\mu(z)$ using the Cauchy integral formula

$$G_\mu(z) = \frac{1}{2\pi i} \int_{\Gamma} \frac{G_\mu(s)}{s - z} ds = \frac{r_B^2}{2\pi i} \int_{\partial \mathbb{D}} \frac{g'(r_B v)v}{g(r_B v) - z} dv.$$

where the second equality follows from the parametrization of Γ with $s = G_\mu^{-1}(r_B v)$ for $v \in \partial \mathbb{D}$. The trapezoidal quadrature rule for this integral gives

$$G_\mu(z) \approx \frac{r_B^2}{N} \sum_{j=0}^{N-1} \frac{g'(r_B \xi_N^j) \xi_N^{2j}}{g(r_B \xi_N^j) - z}. \quad (7)$$

We now choose a circle of radius r_C which is contained inside $J_{[a,b]}^-(\Gamma)$, and we evaluate $\mathcal{G}(v) = G(J_{[a,b]}(v))$ numerically on the points $r_C \xi_M^j$, for $j = 0, 1, \dots, M$. Let us denote by h_0, h_1, \dots, h_M the values obtained by the quadrature formula (7). We will use these to recover an approximation of the measure μ in the next section.

Example 3.4. In Figure 3, we show, in the middle block, the numerically computed extrema a and b (the red dots) of the support of $\mu = \mu_1 \boxplus \mu_2$, where μ_1 and μ_2 are the Marchenko-Pastur distribution with parameter $\lambda = 0.5$ and the semicircle distribution from Example 3.3, respectively. The circle of radius r_B is drawn in orange in the right block. The curve Γ is the orange curve in the middle block, and the orange curve in the left block is $J_{[a,b]}^-(\Gamma)$. The circle of radius r_C is in green in the left block, its image via $J_{[a,b]}$ is the green ellipse in the middle block and its image via \mathcal{G}_μ is the green curve in the right block.

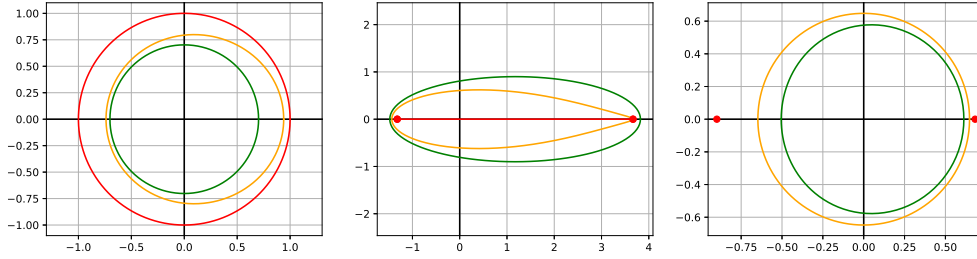


Figure 3: Illustration of Step 3 of the proposed algorithm for free additive convolution, see Example 3.4.

3.4 Step 4: Recovering the density of μ from its Cauchy transform

From the previous step of the algorithm, we have the (approximate) value of \mathcal{G}_μ in M equispaced points on a circle of radius $r_C < 1$. If we truncate the power series (1) to the first M terms, we obtain the relation

$$\begin{bmatrix} \mathcal{G}_\mu(r_C \xi_M^0) \\ \mathcal{G}_\mu(r_C \xi_M^1) \\ \vdots \\ \mathcal{G}_\mu(r_C \xi_M^{M-1}) \end{bmatrix} = \begin{bmatrix} \xi_M^0 & & & \\ & \xi_M^1 & & \\ & & \ddots & \\ & & & \xi_M^{M-1} \end{bmatrix} \cdot F \cdot \begin{bmatrix} g_1 r_C \\ g_2 r_C^2 \\ \vdots \\ g_M r_C^M \end{bmatrix}, \quad (8)$$

where F is the Fourier matrix of size $M \times M$. Therefore, we can recover the (approximate) coefficients g_1, \dots, g_M by performing an IFFT to the vector $[h_0, h_1, \dots, h_{M-1}]^T$ that approximates the left-hand-side and then dividing the j th entry by r_C^j for $j = 1, \dots, M$. Note that, when r_C^j is very small, we cannot hope that the corresponding coefficient of the power series is numerically accurate. Hence, we truncate the series to the first m terms, for some $m < M$. Now using Theorem A.1 we take the following approximation for the density $d\mu(x) = f(x)dx$:

$$f(J_{[a,b]}(\exp(i\theta))) \approx -\frac{1}{\pi} \operatorname{Im} \left(\sum_{j=1}^m g_j \exp(ij\theta) \right),$$

which can be written nicely in matrix form:

$$\begin{bmatrix} f(J_{[a,b]}(\xi_m^0)) \\ f(J_{[a,b]}(\xi_m^1)) \\ \vdots \\ f(J_{[a,b]}(\xi_m^{m-1})) \end{bmatrix} \approx -\frac{1}{\pi} \text{Im} \left(\begin{bmatrix} \xi_m^0 & & & \\ & \xi_m^1 & & \\ & & \ddots & \\ & & & \xi_m^{m-1} \end{bmatrix} \cdot F \cdot \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix} \right). \quad (9)$$

By padding the vector $[g_1 \ g_2 \ \cdots \ g_m]^T$ with zeros, we get a discretization of μ in any number of points.

We remark that (9) only makes sense, theoretically, when the power series corresponding to \mathcal{G}_μ also converges on the unit circle. This is the case when μ has sqrt-behavior at the boundary, and its density is analytic when non-zero. While Theorem 3.2 only states that $\mu = \mu_1 \boxplus \mu_2$ has sqrt-behavior at the boundary, the result from [1, Theorem 2.2 and Lemma 2.7] states that under mild assumptions μ has the analyticity property as well.

Example 3.5. In Figure 4 we plot, in the right block, the approximation of the density of $\mu_1 \boxplus \mu_2$ for the input measures from Examples 3.3 and 3.4, together with the densities of the input measures. In the left block, we compare the obtained approximation of $d\mu(x)$ with the empirical eigenvalue distribution of a 5000×5000 matrix resulting from the sum of two random matrices sampled accordingly to Example 2.2 and 2.3.

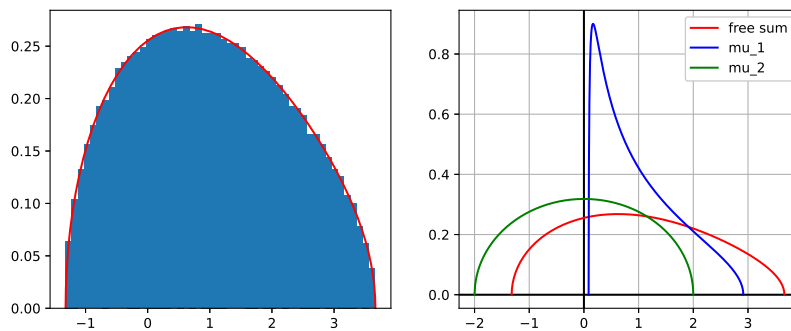


Figure 4: Illustration of Step 4 of the proposed algorithm for free additive convolution, see Example 3.5.

Remark 3.6 (Comparison with [8]). The work [8] also proposes an algorithm to compute the free convolution of measures. In the algorithm from [8], in the sqrt-behavior case, they explicitly compute the inverse of the Cauchy transform. However, our approach via the Cauchy integral theorem allows us to compute the R -transform for any measure with compact support, and not only the ones with sqrt-behavior at the boundary. This makes sense to consider, because the measure μ_2 may not have sqrt-behavior at the boundary. In addition, our approach for evaluating \mathcal{G}_μ on a circle in Step 3 simplifies the approach used in [8] to recover the resulting measure from its inverse Cauchy transform.

Algorithm 1 Computation of the free sum of two measures

Require: Measures μ_1 and μ_2 , number of quadrature points N , number of Fourier coefficients m , tolerance ε , radius r_A , tolerance **tol**. Optional: Cauchy transform, inverse Cauchy transform, the derivative of Cauchy transform, for one or both measures.

Ensure: Discretization of the density of the measure $\mu := \mu_1 \boxplus \mu_2$.

```
1: % Step 1: Set up the computation of the R-transform for both input measures
2: for  $k = 1, 2$  do
3:   if (Inverse) Cauchy transform of  $\mu_k$  is unknown then
4:     Compute values of  $d_j^{(k)}, J'_{[a_k, b_k]}(\xi_N^j), c_j^{(k)}$  appearing in Eq. (6) using (5).
5:   end if
6: end for
7: % Step 2: Find support of  $\mu$ 
8: Find  $\xi_a \in (\max\{\mathcal{G}_{\mu_1}(-r_A + \varepsilon), \mathcal{G}_{\mu_2}(-r_A + \varepsilon)\}, 0)$  such that  $g'(\xi_a) = 0$  ( $g(z)$  from (2)).
9: Find  $\xi_b \in (0, \min\{\mathcal{G}_{\mu_1}(r_A - \varepsilon), \mathcal{G}_{\mu_2}(r_A - \varepsilon)\})$  such that  $g'(\xi_b) = 0$  ( $g(z)$  from (2))
10: The support of  $\mu$  is  $[a, b] := [g(\xi_a), g(\xi_b)]$ .
11: % Step 3: Compute the Cauchy transform of  $\mu$ 
12: Find a (large, if possible) circle inside  $G_\mu(\mathbb{C} \cup \{\infty\} \setminus [a, b])$  by binary search on the radius, using Theorem 3.2 to check whether the circle is inside the wanted region or not. Call  $r_B$  its radius and  $\Gamma := g(r_B \partial \mathbb{D})$ .
13: Let  $r_C$  be (slightly) smaller than the radius of the largest circle inscribed in  $J_{[a, b]}^-(\Gamma)$ .
14: Evaluate  $\mathcal{G}_\mu$  on  $M = \max\{100, \lceil 16 \log_{1/r_C} 10 \rceil\}$  equispaced points on the circle of radius  $r_C$ , using the Cauchy integral theorem + trapezoidal quadrature rule (7).
15: % Step 4: Recover the density of  $\mu$ 
16: Approximate first  $m$  coefficients of power series of  $\mathcal{G}_\mu(z)$  via IFFT (8).
17: Return the approximation of the density of  $\mu$  by (9).
```

4 An algorithm for free multiplicative convolution

The case of free *multiplicative* convolution is very similar to the additive case, so we will skip the details. We present an algorithm for two input measures μ_1 and μ_2 that satisfy the following set of assumptions.

Assumptions 4.1. *We assume that 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.

Under these assumptions (which are more restrictive than the Assumptions 3.1), a theorem similar to Theorem 3.2 holds. The proof is in Appendix B.

Theorem 4.2. *Under the Assumptions 4.1, let us define*

$$t(w) := \frac{w}{1+w} T_1^{-1}(w) T_2^{-1}(w), \quad (10)$$

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

- $\mu := \mu_1 \boxtimes \mu_2$ 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:

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

In summary, the algorithm again has four steps:

1. Set up the computation of the S-transform for μ_1 and μ_2 .
2. Compute the support of μ
3. Compute the T-transform of μ on a suitable set of points.
4. Recover the density of μ from its T-transform.

Step 1. Following the same procedure as in Section 3.1 we can find an approximation of $T(z)$ by rewriting the integral that defines T on the unit circle and then discretizing it using the trapezoidal quadrature rule:

$$T(z) \approx \frac{\pi}{N} \sum_{k=0}^N \frac{b-a}{2} \sin\left(\frac{2\pi k}{N}\right) \frac{f\left(\frac{b-a}{2} \cos\left(\frac{2\pi k}{N}\right) + \frac{b+a}{2}\right)}{z - \left(\frac{b-a}{2} \cos\left(\frac{2\pi k}{N}\right) + \frac{b+a}{2}\right)}.$$

In order to be able to evaluate the S-transform (or the inverse of the T-transform), we will need a contour on which to apply the Cauchy integral theorem. In our algorithm, we compute \mathcal{T}_{μ_1} and \mathcal{T}_{μ_2} in N equispaced points on a circle of radius $r_A < 1$ (close to 1). Analogously, we can compute the derivatives of the T-transform in these same points. Let us denote the approximations by

$$c_j^{(k)} \approx T_k(J_{[a_k, b_k]}(\xi_N^j)), \quad d_j^{(k)} \approx T'_k(J_{[a_k, b_k]}(\xi_N^j)) \quad (11)$$

for $j = 0, 1, \dots, N$ and for $k = 1, 2$.

As the S-transform is analytic in a neighborhood of zero, we can write the Cauchy integral formula for any z inside the curve $\Gamma_1 = \mathcal{T}_{\mu_1}(r_A \partial \mathbb{D})$ (the procedure is analogous for Γ_2):

$$S_{\mu_1}(w) = \frac{1}{2\pi i} \int_{\Gamma_1} \frac{S_{\mu_1}(s)}{w-s} ds.$$

Using the parametrization of Γ_1 from the circle of radius r_A , and then discretizing the circle in N equispaced points using the trapezoidal quadrature rule, we obtain a formula for approximating $S_{\mu_1}(w)$ (and therefore also $T_{\mu_1}^{-1}(w)$) for any w inside Γ_1 :

$$S_{\mu_1}(w) \approx \frac{r_A}{N} \sum_{j=1}^N \frac{\xi_N^j d_j^{(1)} J'_{[a_1, b_1]}(r_A \xi_N^j) (1 + c_j^{(1)})}{c_j^{(1)} J_{[a_1, b_1]}(r_A \xi_N^j) (w - c_j^{(1)})}, \quad T_{\mu_1}^{-1}(w) = \frac{1+w}{w \cdot S_{\mu_1}(w)}.$$

Algorithm 2 Computation of the free product of two measures

Require: Measures μ_1 and μ_2 , number of quadrature points N , number of Fourier coefficients m , tolerance ε , radius r_A , tolerance **tol**. Optional: T-transform, S-transform, derivative of T-transform, for one or both measures.

Ensure: Discretization of the density of the measure $\mu := \mu_1 \boxtimes \mu_2$.

- 1: % *Step 1: Set up the computation of the S-transform for both measures*
 - 2: **for** $k = 1, 2$ **do**
 - 3: **if** (Inverse) S-transform of μ_k is unknown **then**
 - 4: Compute values of $d_j^{(k)}$, $J'_{[a_k, b_k]}(\xi_N^j)$, $c_j^{(k)}$ appearing in Eq. (11).
 - 5: **end if**
 - 6: **end for**
 - 7: % *Step 2: Compute the support of μ*
 - 8: Find $\xi_a \in (\max\{\mathcal{T}_{\mu_1}(-r_A + \varepsilon), \mathcal{T}_{\mu_2}(-r_A + \varepsilon)\}, 0)$ such that $t'(\xi_a) = 0$ ($t(z)$ from (10)).
 - 9: Find $\xi_b \in (0, \min\{\mathcal{T}_{\mu_1}(r_A - \varepsilon), \mathcal{T}_{\mu_2}(r_A - \varepsilon)\})$ such that $t'(\xi_b) = 0$ ($t(z)$ from (10))
 - 10: The support of μ is $[a, b] := [t(\xi_a), t(\xi_b)]$.
 - 11: % *Step 3: Compute the T-transform of μ*
 - 12: Find a (large, if possible) circle inside $T_\mu(\mathbb{C} \cup \{\infty\} \setminus [a, b])$ by binary search on the radius, using the criterion from Theorem 4.2 to check whether the circle is inside the wanted region or not. Let r_B be the radius of the chosen circle and let $\Gamma := t(r_B \partial \mathbb{D})$ be the image of the chosen circle via t .
 - 13: Let r_C be (slightly) smaller than the radius of the largest circle inscribed in $J_{[a, b]}^-(\Gamma)$
 - 14: Evaluate \mathcal{T}_μ on $M = \max\{100, \lceil 16 \log_{1/r_C} 10 \rceil\}$ equispaced points on a circle of radius r_C , which is slightly smaller than the largest circle inscribed in Γ , using the Cauchy integral formula + trapezoidal quadrature rule (12).
 - 15: % *Step 4: Recover the density of μ from its T-transform.*
 - 16: Approximate first m coefficients of power series of $\mathcal{T}(z)$ via IFFT (8).
 - 17: Return the approximation of the density of μ by (13).
-

Step 2. Being able to evaluate $T_{\mu_1}^{-1}$ and $T_{\mu_2}^{-1}$, we can find the support $[a, b]$ of $\mu := \mu_1 \boxtimes \mu_2$ using Theorem 4.2 (by bisection).

Step 3. Theorem 4.2 also gives a criterion to find a circle of radius r_B inside the image of T_μ . Analogously to Section 3.3, we can now evaluate T_μ in any point inside $T_\mu^{-1}(r_B \partial \mathbb{D})$ using the Cauchy integral formula. The discretization by the trapezoidal quadrature rule reads

$$T_\mu(z) \approx \frac{r_B^2}{N} \sum_{j=0}^{N-1} \frac{(J_{[a, b]}^-)'(t(r_B \xi_N^j)) t'(r_B \xi_N^j) \xi_N^{2j}}{J_{[a, b]}^-(t(r_B \xi_N^j)) - z}. \quad (12)$$

We use this formula to approximate $\mathcal{T}_\mu(z)$ on a circle of radius r_C contained inside $\mathcal{T}_\mu^{-1}(r_B \partial \mathbb{D})$.

Step 4. From the approximate values of T_μ , we find a truncated series representation of $\mathcal{T}_\mu(v) \approx \sum_{j=1}^m t_j v^j$ using (8) (which also holds when replacing \mathcal{G} with \mathcal{T}). Finally, we recover the measure μ from the following relation:

$$\begin{bmatrix} f(J_{[a, b]}(\xi_m^0)) J_{[a, b]}(\xi_m^0) \\ f(J_{[a, b]}(\xi_m^1)) J_{[a, b]}(\xi_m^1) \\ \vdots \\ f(J_{[a, b]}(\xi_m^{m-1})) J_{[a, b]}(\xi_m^{m-1}) \end{bmatrix} \approx -\frac{1}{\pi} \text{Im} \left(\begin{bmatrix} \xi_m^0 & & & \\ & \xi_m^1 & & \\ & & \ddots & \\ & & & \xi_m^{m-1} \end{bmatrix} \cdot F \cdot \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{bmatrix} \right). \quad (13)$$

The proposed method for free multiplicative convolution is summarized in Algorithm 2.

5 Error analysis of free convolutions

The discussion of this section focuses on free additive convolution; the proof for free multiplicative convolution is exactly the same. The aim is to explain, for each step of our proposed algorithm, what sources of error are present, and how they can be bounded under suitable assumptions on the measures μ_1 and μ_2 and the other parameters involved in Algorithm 1.

Important for our analysis is the well-known fact that the convergence of the trapezoidal quadrature rule is exponential when the function to be integrated is analytic in a suitable region of the complex plane [2]. We recall a version of this result from [13].

Theorem 5.1 (Exponential convergence of trapezoidal rule for analytic functions). *Suppose u is analytic and satisfies $|u(s)| \leq C$ in an annulus $\rho^{-1} < |s| < \rho$ of the complex plane, for some $\rho > 1$. Let $N > 1$ and consider the approximation I_N of the integral $I := \int_{\partial\mathbb{D}} u(s)ds$ using N equispaced quadrature points on the unit circle and the trapezoidal quadrature rule. Then*

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

This immediately implies the following result on the error in the computation of the Cauchy transforms for measures with sqrt-behavior at the boundary.

Corollary 5.2. *Let μ be a measure with sqrt-decay behavior at the boundary, and further assume that the function $\psi(J_{[a,b]}(v))$ (from Definition 2.1) is analytic in an annulus $\mathcal{A}_{\rho_0} = \{\rho_0^{-1} < |v| < \rho_0\}$ around the unit circle. Let $z \in \mathbb{C} \setminus [a, b]$ and let c be the approximation of $G(z)$ obtained by the trapezoidal quadrature rule (5) with N points. Then for any $\rho < \min(\rho_0, |J_{[a,b]}^+(z)|)$ we have*

$$|G(z) - c| \leq \frac{4\pi}{\rho^N - 1} \cdot \|\psi \circ J_{[a,b]}\|_{\mathcal{A}_\rho} \frac{(b-a)^2}{8} \rho \left(\rho + \frac{1}{\rho}\right)^2 \cdot \frac{1}{\text{dist}(z, J_{[a,b]}(\mathcal{A}_\rho))},$$

where $\|\psi \circ J_{[a,b]}\|_{\mathcal{A}_\rho}$ denotes the maximum modulus of the function $\psi \circ J_{[a,b]}$ on the set \mathcal{A}_ρ .

Proof. If μ has sqrt-behavior at the boundary, we can further manipulate the expression (4) and get

$$\begin{aligned} G(z) &= -\frac{i(b-a)^2}{4} \int_0^\pi \sin^2(\theta) \frac{\psi\left(\frac{b-a}{2} \cos(\theta) + \frac{b+a}{2}\right)}{z - \left(\frac{b-a}{2} \cos(\theta) + \frac{b+a}{2}\right)} d\theta \\ &= \frac{1}{2} \frac{i(b-a)^2}{4} \int_0^{2\pi} \sin^2(\theta) \frac{\psi\left(\frac{b-a}{2} \cos(\theta) + \frac{b+a}{2}\right)}{z - \left(\frac{b-a}{2} \cos(\theta) + \frac{b+a}{2}\right)} d\theta. \end{aligned}$$

Here, the second equality follows from symmetry properties of $\sin^2 \theta$ and $\cos \theta$. Changing back the variable to $v = \exp(i\theta)$ we get

$$G(z) = -\frac{1}{2} \int_{\partial\mathbb{D}} \frac{(b-a)^2}{4} \frac{1}{v} \left(v - \frac{1}{v}\right)^2 \frac{\psi(J_{[a,b]}(v))}{z - J_{[a,b]}(v)} dv, \quad (14)$$

with $\partial\mathbb{D}$ in anti-clockwise direction. Discretizing the integral (14) using the trapezoidal quadrature rule with $2N$ equispaced points is equivalent to discretizing (4) in $N+1$ equispaced quadrature points in $[-\pi, 0]$ with the trapezoidal quadrature rule. Let us consider the function

$$u(v) := -\frac{1}{2} \frac{(b-a)^2}{4} \frac{1}{v} \left(v - \frac{1}{v}\right)^2 \frac{\psi(J_{[a,b]}(v))}{z - J_{[a,b]}(v)}$$

that appears when computing the Cauchy transform. Note that $\frac{1}{v} (v - \frac{1}{v})^2$ is analytic in $\mathbb{C} \setminus \{0\}$ and $\frac{1}{z - J_{a,b}(v)}$ is analytic in the annulus $|J_{a,b}^-(z)| < |v| < |J_{a,b}^+(z)|$. The statement of the corollary, therefore, follows from Theorem 5.1. \square

A result for the derivative $G'(z)$ can be obtained in a similar way.

In our algorithm, we use the approximation (5) for points on the Joukowski transform of a circle of radius $r_A < 1$, therefore $\rho = \min(\rho_0, r_A^{-1})$ in this case. This means that the closer r_A gets to 1, the slower the convergence becomes. Note that (5) can also be used for measures that do not have the sqrt-behavior at the boundary. However, we have no guarantees on the fast convergence of the trapezoidal quadrature rule.

Theorem 5.3 (Error in the computation of the R-transform). *Let μ be a measure with support on $[a, b]$ and let w be a point inside the curve $\Gamma = G(r\partial\mathbb{D})$ for some $r \in (0, 1)$ for which G is invertible inside the disk of radius r . Let $\rho < \min\{r^{-1}, \text{dist}(G^{-1}(w), r\partial\mathbb{D})\}$. Let $\xi_N = \exp(2\pi i/N)$. Assume we have computed approximations $c_j \approx \mathcal{G}(r\xi_N^j)$ and $d_j \approx G'(J(r\xi_N^j))$ such that $|\mathcal{G}(r\xi_N^j) - c_j| \leq \varepsilon$ and $|G'(J(r\xi_N^j)) - d_j| \leq \varepsilon$. Let $m_1 = \min\{\|\mathcal{G}\|_\Gamma, c_0, \dots, c_{N-1}\}$. Let m_2 be the distance of w from Γ . Let us discretize the integral (3) with the trapezoidal quadrature rule (6)*

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

Then

$$|R_N(w) - R(w)| \leq \frac{4\pi}{\rho^N - 1} \max_{\rho^{-1} \leq |v| \leq \rho} |u(v)| + \varepsilon r \|J'\|_{r\partial\mathbb{D}} \left(\frac{\|J\|_{r\partial\mathbb{D}}}{m_2} + \frac{1}{m_1 m_2} + (\varepsilon + \|G'\|_{J(r\partial\mathbb{D})}) \frac{1}{m_2^2} \left(\|J\|_{r\partial\mathbb{D}} + \frac{|w|}{m_1^2} + \frac{2}{m_1} \right) \right). \quad (15)$$

Proof. Informally, we have two sources of error:

- the quadrature error – as the R-transform is analytic, Theorem 5.1 can be applied;
- the fact that the function to be integrated cannot be evaluated exactly – each point requires quadrature as well.

More precisely, by triangular inequality we have

$$|R(w) - R_N(w)| \leq \left| \frac{1}{2\pi i} \int_\Gamma \frac{R(s)}{w - s} - \frac{r}{N} \sum_{j=1}^N \xi_N^j \cdot G'(J(r\xi_N^j)) \cdot J'(\xi_N^j) \cdot \frac{J(r\xi_N^j) - 1/\tilde{G}(r\xi_N^j)}{\tilde{G}(r\xi_N^j) - w} \right| + \frac{r}{N} \sum_{j=1}^N |\xi_N^j J'(r\xi_N^j)| \cdot \left| d_j \frac{J(r\xi_N^j) - 1/c_j}{c_j - w} - G'(J(r\xi_N^j)) \frac{J(r\xi_N^j) - 1/\mathcal{G}(r\xi_N^j)}{\mathcal{G}(r\xi_N^j) - w} \right|.$$

We can directly use Theorem 5.1 to bound the first term by $\frac{4\pi}{\rho^N - 1} \max\{|u(v)| : \rho^{-1} \leq |v| \leq \rho\}$, where $u(v)$ is the function to be integrated, defined in (3). Let us analyze each term in the sum of the second part. To simplify notation, we drop the dependence of the functions on $r\xi_N^j$ and

denote $J(r\xi_N^j) = J_j$, $G'(J(r\xi_N^j)) = G'_j$, $\mathcal{G}(r\xi_N^j) = \mathcal{G}_j$. We have

$$\begin{aligned} \left| d_j \frac{J_j - 1/c_j}{c_j - w} - G'_j \frac{J_j - 1/\mathcal{G}_j}{\mathcal{G}_j - w} \right| &\leq |G'_j - d_j| \cdot \left| \frac{J_j - 1/\mathcal{G}_j}{\mathcal{G}_j - w} \right| + |d_j| \cdot \left| \frac{J_j - 1/c_j}{c_j - w} - \frac{J_j - 1/\mathcal{G}_j}{\mathcal{G}_j - w} \right| \\ &= |G'_j - d_j| \cdot \left| \frac{J_j - 1/\mathcal{G}_j}{\mathcal{G}_j - w} \right| + |d_j| \cdot \left| \frac{J_j(\mathcal{G}_j - c_j) + w(1/c_j - 1/\mathcal{G}_j) + (c_j/\mathcal{G}_j - \mathcal{G}_j c_j)}{(c_j - w)(\mathcal{G}_j - w)} \right|. \end{aligned}$$

We can bound the terms with $|G'_j - d_j| \cdot \left| \frac{J_j - 1/\mathcal{G}_j}{\mathcal{G}_j - w} \right| \leq \frac{\varepsilon}{m_2} \left(\|J\|_{r\partial\mathbb{D}} + \frac{1}{m_1} \right)$, $|1/c_j - 1/\mathcal{G}_j| = \left| \frac{\mathcal{G}_j - c_j}{c_j \mathcal{G}_j} \right| \leq \frac{\varepsilon}{m_1^2}$, $|c_j/\mathcal{G}_j - \mathcal{G}_j/c_j| \leq \frac{2\varepsilon}{m_1}$, $|(c_j - w)(\mathcal{G}_j - w)| \geq m_2^2$, therefore obtaining (15). \square

Theorem (5.3) basically states that the quadrature error in the computation of the R-transform decreases exponentially with the number of quadrature points, and the error due to the inexact evaluation of G and G' is not amplified significantly. Note that this result does not need the assumption that μ has sqrt-behavior at the boundary.

Remark 5.4 (Error in the computation of the support of $\mu_1 \boxplus \mu_2$). *Let us denote by $g'_N(w)$ the approximation of the function $g'(w)$ obtained using N quadrature points for all the involved quadrature rules. When $N \rightarrow \infty$, the functions $g'_N(w)$ are holomorphic and they converge to $g'(w)$ uniformly on compact subsets of the intervals $(\max\{\mathcal{G}_{\mu_1}(-r_A + \varepsilon), \mathcal{G}_{\mu_2}(-r_A + \varepsilon)\}, 0)$ and $(0, \min\{\mathcal{G}_{\mu_1}(r_A - \varepsilon), \mathcal{G}_{\mu_2}(r_A - \varepsilon)\})$. By Hurwitz's theorem, we can conclude that for sufficiently large values of N , the functions $g'_N(w)$ have exactly one simple zero in each of the intervals, and these zeros converge to the zeros ξ_a, ξ_b of $g'(w)$. Therefore, when the number of quadrature points is sufficiently large, the computed values of the support $[a, b]$ of μ are accurate.*

What we said until now means that we are able to evaluate the R-transforms of μ_1 and μ_2 accurately, together with the support of $\mu = \mu_1 \boxplus \mu_2$. The next step in Algorithm 1 is the evaluation of $\mathcal{G}_\mu(v)$ on a suitable circle of radius $r_C \in (0, 1)$ (see line 14).

Remark 5.5 (Error in the computation of the Cauchy transform of $\mu_1 \boxplus \mu_2$). *Let $r = r_C$ and let us discuss the sources of error in the evaluation of $\mathcal{G}(v)$. Let us denote by $\tilde{g}_j, \tilde{g}'_j$ the approximations of $g(r\xi_N^j)$ and $g'(r\xi_N^j)$, respectively, obtained by numerical quadrature. For any point v inside the curve $J_{[a,b]}^-(\Gamma)$ defined in Section 3.3, letting \tilde{a}, \tilde{b} be the approximated extrema of the support of μ , the approximation of $\mathcal{G}(v)$ obtained from (7) would be*

$$\tilde{\mathcal{G}}_v := \frac{r^2}{N} \sum_{j=0}^{N-1} \frac{\tilde{g}'_j \xi_N^{2j}}{\tilde{g}_j - J_{[\tilde{a}, \tilde{b}]}(v)}.$$

We can split the error $|\mathcal{G}(v) - \tilde{\mathcal{G}}_v|$ into three sources by triangular inequality:

$$\begin{aligned} |\mathcal{G}(v) - \tilde{\mathcal{G}}_v| &\leq \left| G(J_{[a,b]}(v)) - G(J_{[\tilde{a}, \tilde{b}]}(v)) \right| + \left| G(J_{[\tilde{a}, \tilde{b}]}(v)) - \frac{r^2}{N} \sum_{j=0}^{N-1} \frac{g'(r\xi_N^j) \xi_N^{2j}}{g(r\xi_N^j) - J_{[\tilde{a}, \tilde{b}]}(v)} \right| \\ &\quad + \frac{r^2}{N} \sum_{j=0}^{N-1} \left| \frac{g'(r\xi_N^j)}{g(r\xi_N^j) - J_{[\tilde{a}, \tilde{b}]}(v)} - \frac{\tilde{g}'_j}{\tilde{g}_j - J_{[\tilde{a}, \tilde{b}]}(v)} \right|. \end{aligned}$$

We refrain from outlining a full error analysis of this step, but we will briefly comment on the three terms in the above error bound. The first term can be expected to be small if $[\tilde{a}, \tilde{b}]$ is accurate.

The second term, the error in the trapezoidal quadrature rule, can be bounded by Theorem 5.1 because the function to be integrated (the Cauchy transform G) is analytic in an annulus around $\partial\mathbb{D}$; therefore, this part decreases exponentially fast in the number of the quadrature points N . Finally, the third term can be bounded similarly to what is done in Theorem 5.3 if we know that $\tilde{g}_j, \tilde{g}_j', \tilde{a}, \tilde{b}$ have been computed accurately.

The last part of the algorithm is about recovering the density of the measure $\mu = \mu_1 \boxplus \mu_2$ from the (approximate) values of the Cauchy transform $G = G_\mu$ on equispaced points on the circle of radius $r = r_C$.

Theorem 5.6 (Error in the computation of the coefficients of \mathcal{G}). *Let μ be a measure with sqrt-behavior, at the boundary, with support $[a, b]$ and density $d\mu(x) = \sqrt{b-x}\sqrt{x-a}\psi(x)dx$, where $\psi(x)$ is Lipschitz-continuous. Assume that we have a vector $[\tilde{\mathcal{G}}_0 \ \tilde{\mathcal{G}}_1 \ \dots \ \tilde{\mathcal{G}}_{M-1}]^T$ that approximates, up to entrywise error ε , the vector $[\mathcal{G}(r\xi_M^0) \ \mathcal{G}(r\xi_M^1) \ \dots \ \mathcal{G}(r\xi_M^{M-1})]^T$ for some $M \in \mathbb{N}$ and some $r < 1$, that is, $|\tilde{\mathcal{G}}_j - \mathcal{G}(r\xi_M^j)| \leq \varepsilon$ for $j = 0, 1, \dots, M-1$. Let*

$$\begin{bmatrix} \tilde{g}_1 \\ \tilde{g}_2 \\ \vdots \\ \tilde{g}_M \end{bmatrix} := \begin{bmatrix} r^{-1} & & & \\ & r^{-2} & & \\ & & \ddots & \\ & & & r^{-M} \end{bmatrix} F^{-1} \begin{bmatrix} \xi_M^{-0} & & & \\ & \xi_M^{-1} & & \\ & & \ddots & \\ & & & \xi_M^{-M+1} \end{bmatrix} \begin{bmatrix} \tilde{\mathcal{G}}_0 \\ \tilde{\mathcal{G}}_1 \\ \vdots \\ \tilde{\mathcal{G}}_{M-1} \end{bmatrix}.$$

Then

$$\left\| \begin{bmatrix} \tilde{g}_1 \\ \tilde{g}_2 \\ \vdots \\ \tilde{g}_M \end{bmatrix} - \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_M \end{bmatrix} \right\| \leq \begin{bmatrix} r^{-1} \\ r^{-2} \\ \vdots \\ r^{-M} \end{bmatrix} \frac{\varepsilon}{\sqrt{M}} + \frac{r^M}{1-r^M} \|\psi\|_{[a,b]} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},$$

where the symbol $|\cdot|$ means that the inequality holds entry-wise and g_1, \dots, g_M are the exact coefficients of the series expansion (1).

Proof. First of all, we can write

$$\begin{bmatrix} \mathcal{G}(r\xi_M^0) \\ \mathcal{G}(r\xi_M^1) \\ \vdots \\ \mathcal{G}(r\xi_M^{M-1}) \end{bmatrix} = \begin{bmatrix} \xi_M^0 & & & \\ & \xi_M^1 & & \\ & & \ddots & \\ & & & \xi_M^{M-1} \end{bmatrix} \begin{bmatrix} F & F & F & \dots \end{bmatrix} \begin{bmatrix} g_1 r \\ g_2 r^2 \\ g_3 r^3 \\ \vdots \end{bmatrix},$$

where we have an infinite matrix and an infinite vector (it is alright to write this because the series converges absolutely). Using triangular inequality (entry-wise), we have

$$\begin{aligned} \left\| \begin{bmatrix} \tilde{g}_1 \\ \tilde{g}_2 \\ \vdots \\ \tilde{g}_M \end{bmatrix} - \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_M \end{bmatrix} \right\| &\leq \left\| \begin{bmatrix} r^{-1} & & & \\ & r^{-2} & & \\ & & \ddots & \\ & & & r^{-M} \end{bmatrix} F^{-1} \begin{bmatrix} \xi_M^{-0} & & & \\ & \xi_M^{-1} & & \\ & & \ddots & \\ & & & \xi_M^{-M+1} \end{bmatrix} \left(\begin{bmatrix} \tilde{\mathcal{G}}_0 \\ \tilde{\mathcal{G}}_1 \\ \vdots \\ \tilde{\mathcal{G}}_{M-1} \end{bmatrix} - \begin{bmatrix} \mathcal{G}(r\xi_M^0) \\ \mathcal{G}(r\xi_M^1) \\ \vdots \\ \mathcal{G}(r\xi_M^{M-1}) \end{bmatrix} \right) \right\| \\ &+ \left\| \begin{bmatrix} r^{-1} & & & \\ & r^{-2} & & \\ & & \ddots & \\ & & & r^{-M} \end{bmatrix} F^{-1} \begin{bmatrix} \xi_M^{-0} & & & \\ & \xi_M^{-1} & & \\ & & \ddots & \\ & & & \xi_M^{-M+1} \end{bmatrix} \begin{bmatrix} \xi_M^0 & & & \\ & \xi_M^1 & & \\ & & \ddots & \\ & & & \xi_M^{M-1} \end{bmatrix} \begin{bmatrix} F & F & F & \dots \end{bmatrix} \begin{bmatrix} g_1 r \\ g_2 r^2 \\ g_3 r^3 \\ \vdots \end{bmatrix} - \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_M \end{bmatrix} \right\|. \end{aligned}$$

The first term is bounded by

$$\left\| \begin{bmatrix} r^{-1} & & & \\ & r^{-2} & & \\ & & \ddots & \\ & & & r^{-M} \end{bmatrix} \cdot \left\| F^{-1} \begin{bmatrix} \xi_M^{-0} & & & \\ & \xi_M^{-1} & & \\ & & \ddots & \\ & & & \xi_M^{-M+1} \end{bmatrix} \right\|_2 \cdot \begin{bmatrix} \varepsilon \\ \varepsilon \\ \vdots \\ \varepsilon \end{bmatrix} \right\| \leq \begin{bmatrix} r^{-1} \\ r^{-2} \\ \vdots \\ r^{-M} \end{bmatrix} \frac{\varepsilon}{\sqrt{N}}$$

because F^{-1} has norm $1/\sqrt{M}$ and all the ξ_M^{-j} have modulus 1. The second term simplifies to

$$\left\| \begin{bmatrix} g_{M+1}r^M + g_{2M+1}r^{2M} + g_{3M+1}r^{3M} + \dots \\ g_{M+2}r^M + g_{2M+2}r^{2M} + g_{3M+2}r^{3M} + \dots \\ \vdots \\ g_{2M}r^M + g_{3M}r^{2M} + g_{4M}r^{3M} + \dots \end{bmatrix} \right\| \leq \frac{r^M}{1-r^M} \left(\max_{k \geq M+1} |g_k| \right) \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Due to the connections with expansions of \mathcal{G} in the Chebyshev series (see Remark 2.9), all coefficients g_k have modulus bounded by $\|\psi\|_{[a,b]}$, which concludes the proof of the theorem. \square

Theorem 5.6 tells us that whenever ε/r^j is large, we cannot expect the approximation of g_j to be reliable. This is why we are only keeping the first few approximations of the series coefficients. We also need to ensure that $r^M = r_C^M$ is small so that the second error term is small (but this is less problematic): for this reason, in our code, we set $M = \max \left\{ 100, \left\lceil 16 \log_{1/r_C} 10 \right\rceil \right\}$.

Remark 5.7 (Error in the recovery of the density of $\mu_1 \boxplus \mu_2$). *Let us denote by*

$\begin{bmatrix} \tilde{f}_0 & \dots & \tilde{f}_{m-1} \end{bmatrix}^T$ *the approximation obtained from (9) by plugging in the vector $\begin{bmatrix} \tilde{g}_1 & \dots & \tilde{g}_m \end{bmatrix}^T$. Using the triangular inequality and the fact that F is unitary up to a \sqrt{m} scaling factor, we get that*

$$\left\| \begin{bmatrix} f(J_{[a,b]}(\xi_m^0)) \\ f(J_{[a,b]}(\xi_m^1)) \\ \vdots \\ f(J_{[a,b]}(\xi_m^{m-1})) \end{bmatrix} - \begin{bmatrix} \tilde{f}_0 \\ \tilde{f}_1 \\ \vdots \\ \tilde{f}_{m-1} \end{bmatrix} \right\| \leq \frac{\sqrt{m}}{\pi} \left\| \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix} - \begin{bmatrix} \tilde{g}_1 \\ \tilde{g}_2 \\ \vdots \\ \tilde{g}_m \end{bmatrix} \right\| + \left\| \begin{bmatrix} f(J_{[a,b]}(\xi_m^0)) \\ f(J_{[a,b]}(\xi_m^1)) \\ \vdots \\ f(J_{[a,b]}(\xi_m^{m-1})) \end{bmatrix} + \frac{1}{\pi} \operatorname{Im} \left(\begin{bmatrix} \xi_m^0 & & & \\ & \xi_m^1 & & \\ & & \ddots & \\ & & & \xi_m^{m-1} \end{bmatrix} \cdot F \cdot \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix} \right) \right\|.$$

The first error term can be bounded with Theorem 5.6. The second error term depends on how fast the series (1) converges to $\mathcal{G}_\mu(v)$, which, in turn, depends on the regularity of the density function associated with μ (see Remark 2.9).

6 Numerical experiments

6.1 Exponential convergence of trapezoidal quadrature rule for the pointwise evaluation of the Cauchy transform

First of all, we consider the evaluation of the Cauchy transform in some points outside the support for the semicircle, uniform (with $m = 3$), and Marchenko-Pastur (with $\lambda = 0.7$) distributions.

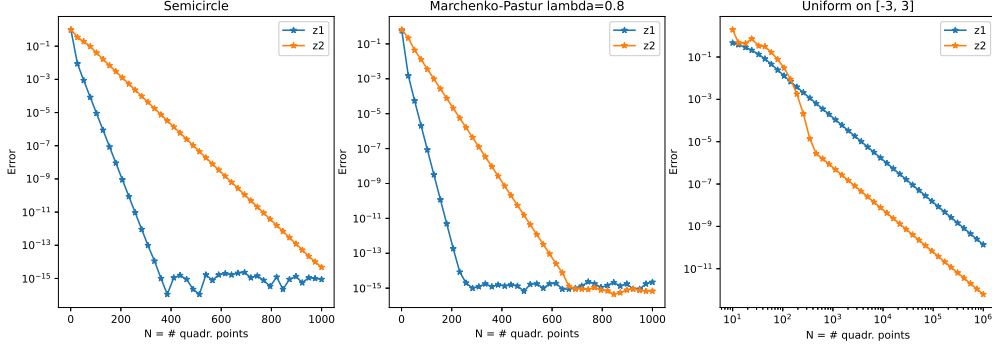


Figure 5: Evaluation of the Cauchy transform of the points $z_1 = 1.001 \cdot b$ and $z_2 = (0.8 + 0.01i) \cdot b$ for the semicircle distribution (left), the Marchenko-Pastur with parameter 0.8 (middle), and the uniform distribution on $[-3, 3]$ (right). The x-axis is the number of quadrature points, and the y-axis is the error of the trapezoidal quadrature rule.

The results are reported in Figure 5. For each measure, we consider the point $z_1 = 1.001 \cdot b$ (blue) and $z_2 = (0.8 + 0.01i) \cdot b$ (orange). The x-axis is the number of quadrature points, and the y-axis is the error of the trapezoidal quadrature rule. As expected, the error decreases exponentially for the semicircle and Marchenko-Pastur. The same cannot be said for the uniform distribution (note that the graph is log-log in this case).

6.2 Recovery of measures from the Cauchy transform

We test numerically the recovery of the spectral measure μ from the knowledge of the (exact) $\mathcal{G}(z)$ on a circle of radius $R = 0.9$ in $N = 3000$ points coming from equispaced points on a circle, using $m = 40$ coefficients in the power series (1) for the Cauchy transform. The results are shown in Figure 6. As expected, we can recover the measures with sqrt-behavior at the boundary very well.

6.3 Free additive and multiplicative convolution of distributions

In this section, we test Algorithm 1 on various input measures. In some cases, we know the analytic form of the support of the free sum or the analytic form of the free sum itself. This allows us to test the accuracy of our algorithm on some examples. In the other cases, we will compare the approximate density returned by our algorithm with the empirical eigenvalue distribution of the sum of two random matrices that are chosen according to the input distributions.

How to read the plots. Let us explain what is plotted in each figure. The first (top) block row is about the measure μ_1 , the second block row is about μ_2 , and the third block row is about the free convolution $\mu_1 \boxplus \mu_2$. We pass from the first (left) column to the second (middle) column with the Joukowski transform, and from the second column to the third (right) column with the Cauchy/T-transform. The red circle in the left column is always the unit circle. The red line in the second column indicates the support of the corresponding measure. The blue curve in the (1,3) block (the top right block) is Γ_1 and the blue curve in the (2,3) block (the top middle block) is Γ_2 . These same curves are reported in the (4,3) block, together with an orange circle

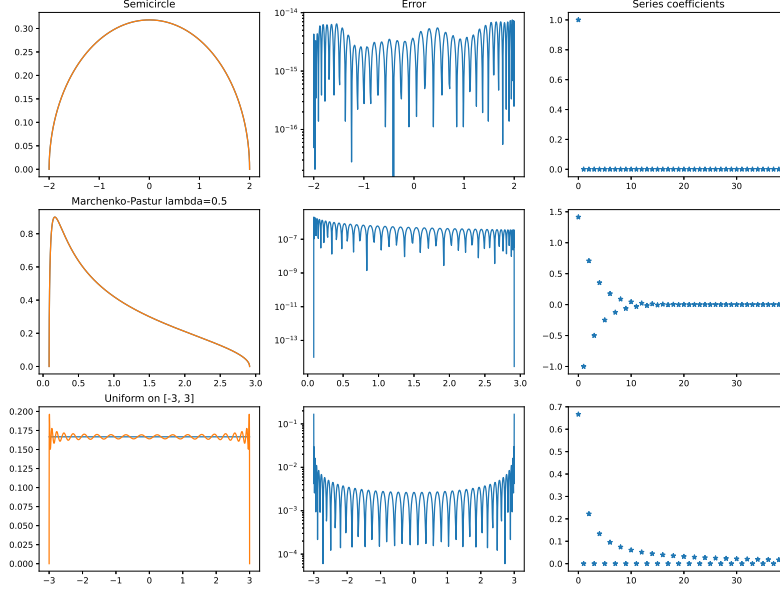


Figure 6: Recovering the measure from 40 Fourier coefficients of the Cauchy transform. From top to bottom, we consider the semicircle distribution, the Marchenko-Pastur law with parameter $\lambda = 0.5$, and the uniform distribution on the interval $[-3, 3]$. In the left block column, the blue curve is the density function, and the orange curve is the reconstruction of the measure. In the middle block column, we plot the pointwise error $|f(x) - \{f\}(x)|$. The right block column illustrates the magnitude of the approximate coefficients g_1, g_2, \dots, g_{10} .

that is in the intersection of the two curves and inside the image of $\mathbb{C} \cup \{\infty\} \setminus [a, b]$. The same orange circle is drawn in the (1,3), (2,3), and (3,3) blocks, and the inverses via G_1, G_2, G are in the (1,2), (2,2), (3,2) blocks, respectively. Finally, the circle \mathcal{A} is represented in green in the (3,1) block, and its image via J and then G/T is also in green in the (3,2) and (3,3) blocks. In the (4,2) block, we plot the densities of the measures μ_1, μ_2 , and μ . In the (4,1) block, we (usually) plot the eigenvalue distribution of one instance of a random matrix of size 5000×5000 obtained as a sum/product of matrices with limit eigenvalue distributions μ_1 and μ_2 .

Implementation details. Unless indicated otherwise, we set the following parameters for Algorithm 1:

- For computing the Cauchy transform in a point, we use 400 quadrature points for measures with sqrt-behavior at the boundary (semicircle, Marchenko-Pastur) and 4000 points for other measures.
- We use 400 quadrature points to discretize the integrals corresponding to the Cauchy integral formulae for computing the R-transform of μ_1 and μ_2 as well as the Cauchy transform of μ .

- The number of points on the circle of radius r_C is chosen as $\max \left\{ 100, \left\lfloor 16 \log_{1/r_C} 10 \right\rfloor \right\}$ (this is to ensure that the second term in the error of Theorem 5.6 is negligible). Then we truncate the series to the first $m = 20$ terms.
- We choose a parameter $\varepsilon = 0.05$ that quantifies how “far from the border” are the curves and intervals we take. More specifically:
 - we set $r_A = 1 - \varepsilon$;
 - we use this value of ε in lines 8–9 of Algorithm 1;
 - when looking for the radius of the largest circle inside

$$G_{\mu_1}(\mathbb{C} \cup \{\infty\} \setminus [a_1, b_1]) \cap G_{\mu_2}(\mathbb{C} \cup \{\infty\} \setminus [a_2, b_2]) \cap G_{\mu}(\mathbb{C} \cup \{\infty\} \setminus [a, b])$$

we start looking at the circle of radius $(1 - \varepsilon)r$, where r is the radius of the largest circle inscribed in $G_{\mu_1}(\mathbb{C} \cup \{\infty\} \setminus [a_1, b_1]) \cap G_{\mu_2}(\mathbb{C} \cup \{\infty\} \setminus [a_2, b_2])$;

- when choosing the radius r_C we multiply by $(1 - \varepsilon)$ the radius of the largest circle inscribed in $J_{[a,b]}^-(\Gamma)$.
- The tolerance for stopping the binary search for the computation of the support of μ is set to 10^{-12} . The tolerance for stopping the binary search for the computation of r_B is 10^{-3} .

For Algorithm 1, we use the same values of the parameters in the corresponding parts of the code.

Code availability. A Python implementation of Algorithms 1 and 2 is available at <https://github.com/Alice9> together with scripts to reproduce all the numerical examples in this paper.

Example 6.1. In Figure 7, we compute the free additive convolution of two semicircle distributions. The result is known analytically, and it is a rescaled semicircle distribution, with support in $[-2\sqrt{2}, 2\sqrt{2}]$ and density function $f(x) = \frac{1}{4\pi} \sqrt{8 - x^2}$. Our algorithm is able to compute the free additive convolution with an accuracy of 10^{-14} . We only used $m = 10$ coefficients of the series (1) in this example.

Example 6.2. In Figure 8, we compute the free additive convolution of the semicircle distribution with the uniform distribution on the interval $[-4, 4]$. In this case, we do not know the analytic form of the solution, but we can compute the support of the free sum, which is

$$\left[-\frac{1}{4} \log \left(4 + \sqrt{17} \right) - \sqrt{17}, \frac{1}{4} \log \left(4 + \sqrt{17} \right) + \sqrt{17} \right].$$

The supports that we computed numerically agree with the theoretical one up to an absolute error of $3.6 \cdot 10^{-6}$. The uniform distribution does not have an sqrt-behavior at the boundary; hence the trapezoidal quadrature rule for computing the Cauchy transform converges more slowly. For this reason, we use 4000 quadrature points for this example. Note that we could have used an explicit expression of the Cauchy transform because it is known analytically, but we wanted to see how our algorithm performed without this knowledge.

Example 6.3. In Figure 9, we consider the free additive convolution of the uniform distribution on the interval $[-2, 2]$ and the Marchenko-Pastur distribution with $\lambda = 0.7$.

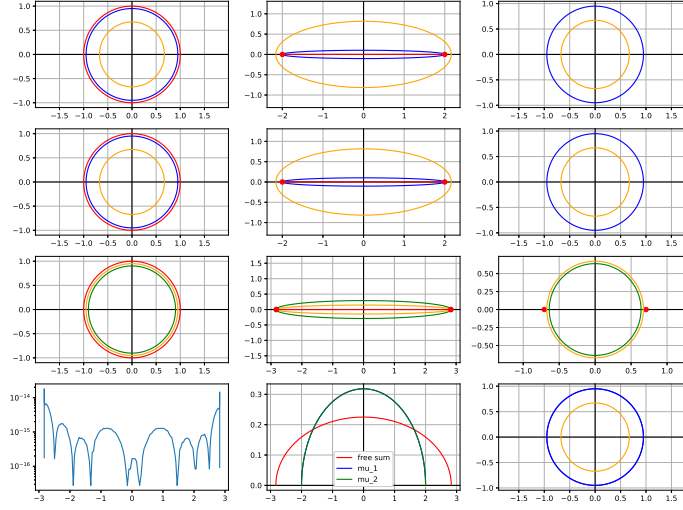


Figure 7: Free sum of semicircle and semicircle. Only 400 quadrature points. The (4, 2) subplot is the error of the computed free sum. See Example 6.1 for a precise description.

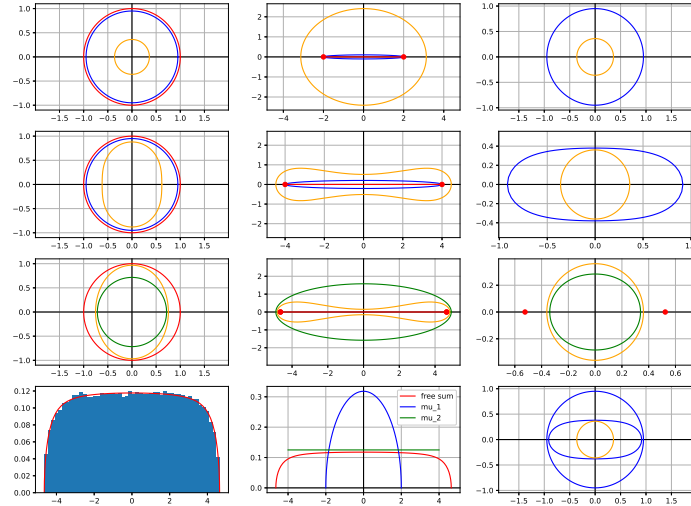


Figure 8: Free sum of semicircle and uniform on $[-4, 4]$. See Example 6.2.

Example 6.4. In Figure 10, we consider the free multiplicative convolution of a shifted semicircle distribution (with support in $[1, 5]$ instead of $[-2, 2]$) and a Marchenko-Pastur distribution with $\lambda = 0.2$.

Example 6.5. In Figure 11, we consider the free multiplicative convolution of two identical

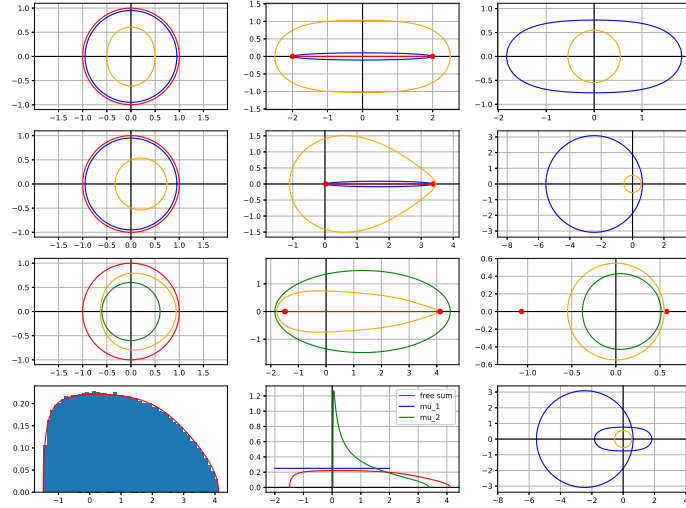


Figure 9: Free sum of uniform on $[-2, 2]$ and Marchenko-Pastur ($\lambda = 0.7$). See Example 6.3.

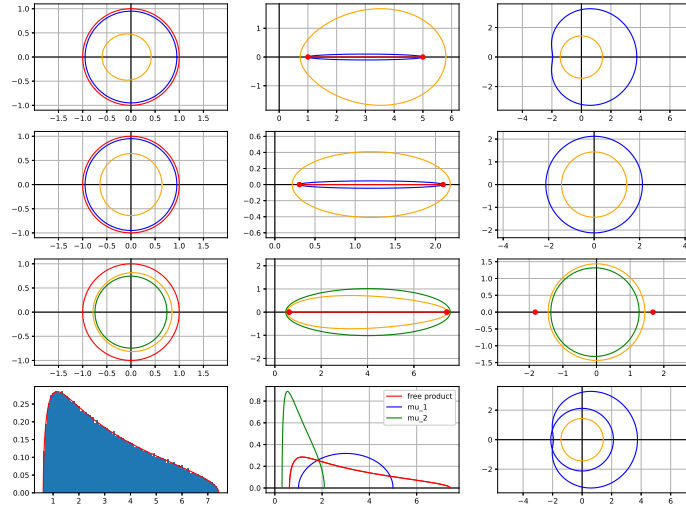


Figure 10: Free multiplication of shifted semicircle and Marchenko-Pastur distribution. See Example 6.4.

shifted semicircle distributions.

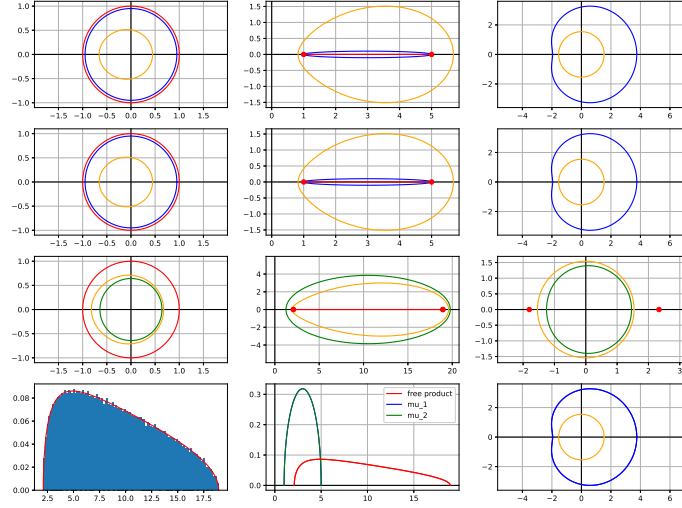


Figure 11: Free multiplication of two shifted semicircle distributions. See Example 6.5.

6.4 Strengths and limitations of our algorithms

The numerical examples above show that Algorithms 1 and 2 give a good approximation of the free convolution of measures in a variety of situations. In the following four examples, we show that our algorithm can give good insights even in cases in which the input measure does not satisfy the assumptions of Theorem 3.2.

Example 6.6. In Figure 12, we consider the free additive convolution of a uniform distribution on the interval $[-1, 1]$ and a uniform distribution on the interval $[-2, 2]$. This combination does not fall into the theory for our algorithm, but their free additive convolution has an *sqrt*-behavior at the boundary (see, e.g., [1]).

Example 6.7. In Figure 13 we consider the free multiplicative convolution of the Marchenko-Pastur distribution with $\lambda = 0.1$ and the uniform distribution on the interval $[1, 3]$.

Example 6.8. In Figure 14, we consider the free convolution of the semicircle distribution with the distribution μ_2 that has support in $[-\sqrt{3}, \sqrt{3}]$ and has a density

$$d\mu_2(x) = \frac{5\sqrt{3}}{144}(x^2 + 1)^2.$$

This has the property that $G'_{\mu_2}(i) = G'_{\mu_2}(-i) = 0$ (and it does not have *sqrt*-behavior at the boundary), therefore its Cauchy transform is not invertible. The image via \mathcal{G}_{μ_2} of the blue circle of radius r_A is a curve with two self-intersections. This is not a problem for the application of the Cauchy integral formula, as long as we consider a point that is inside the curve. The disadvantage of this example is that we need r_B – the radius of the orange circle – to be quite small, and this means that we need to truncate the power series corresponding to $\mathcal{G}_{\mu_1 \boxplus \mu_2}$ to the

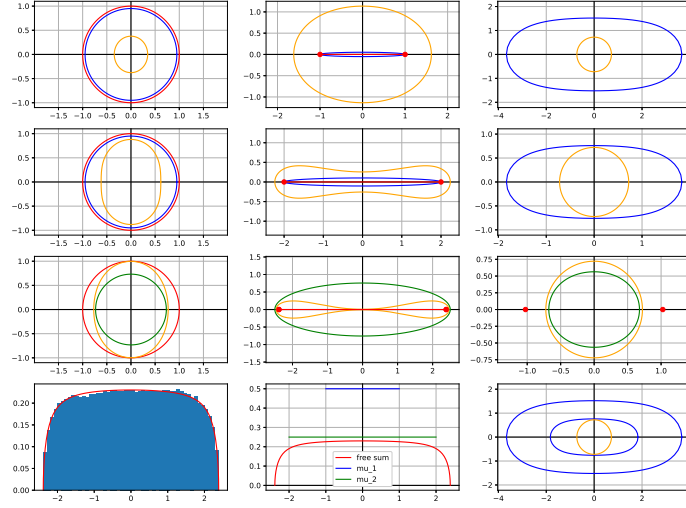


Figure 12: Free sum of uniform distributions on $[-1, 1]$ and $[-2, 2]$. See Example 6.6.

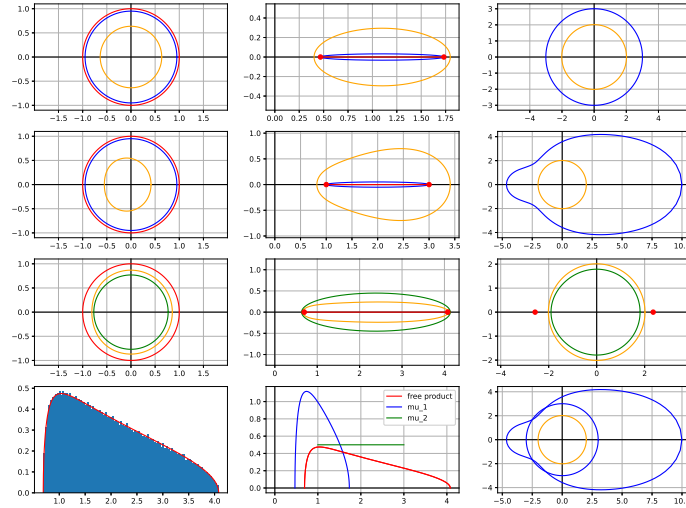


Figure 13: Free multiplication of Marchenko-Pastur and uniform distribution. See Example 6.7.

first few terms. The numerical approximation that we get closely agrees with the histogram of the eigenvalues of the sum of a symmetric random matrix and a random matrix whose eigenvalues are samples from the weird distribution. We used 4000 quadrature points for the discretization of the integral of the Cauchy transform of μ in this example.

Although the theory does not extend to this case, we can also consider measures that have

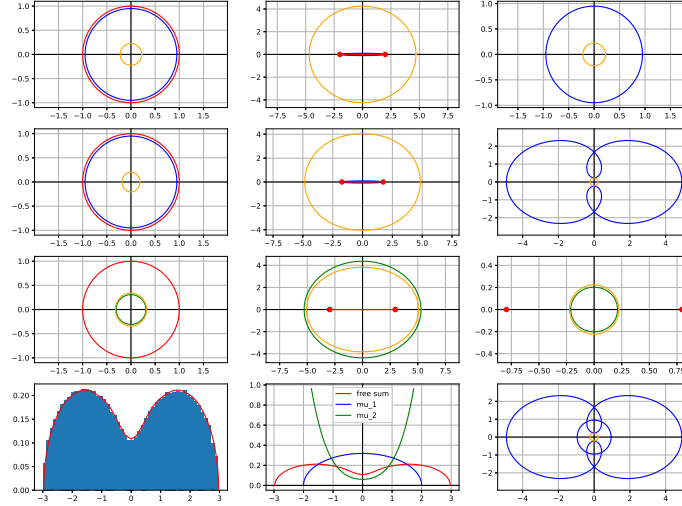


Figure 14: Free sum of semicircle law and the distribution from Example 6.8, with $m = 20$.

atoms. For a point measure, it is easy to explicitly write its Cauchy transform (which is not invertible when there are more than two atoms). In our code, we use an explicit expression for the part of the Cauchy transform corresponding to the atoms, and quadrature for the rest. When considering the free convolution μ of a measure with sqrt-behavior at the boundary with a discrete measure, it can happen that the support of μ splits into two or more distinct pieces; when this is *not* the case, our algorithm can successfully approximate the density of μ , as in the next example.

Example 6.9. In Figure 15, we consider the free multiplicative convolution of a shifted semicircle distribution with a discrete measure that has masses $\frac{1}{7}$ in the points $1, \frac{3}{2}, 2, \frac{5}{2}, 3, \frac{7}{2}, 4$. In this case, the T -transform is easy to compute directly, so we do not use the trapezoidal quadrature rule for this part of the algorithm.

We comment on a warning on the choice of the parameters (r_A and ε) in our algorithms. On the one hand, when r_A is (relatively) far from 1 and ε is (relatively) large, the trapezoidal rules for the computation of $g(z)$ on the two intervals $(\max\{\mathcal{G}_{\mu_1}(-r_A + \varepsilon), \mathcal{G}_{\mu_2}(-r_A + \varepsilon)\}, 0)$ and $(0, \min\{\mathcal{G}_{\mu_1}(r_A - \varepsilon), \mathcal{G}_{\mu_2}(r_A - \varepsilon)\})$ converge very fast, allowing for a small value of quadrature points N . On the other hand, Theorem 3.2 ensures the existence of a zero of g' in the *larger* intervals $(\max(G_{\mu_1}(a_1), G_{\mu_2}(a_2)), 0)$ and $(0, \min(G_{\mu_1}(b_1), G_{\mu_2}(b_2)))$. Therefore, an easy choice of r_A and ε may mean that we cannot successfully find the support of μ . In our implementation, we give a warning when this happens, and then proceed to choose $\xi_a = \max\{\mathcal{G}_{\mu_1}(-r_A + \varepsilon), \mathcal{G}_{\mu_2}(-r_A + \varepsilon)\}$ or $\xi_b = \min\{\mathcal{G}_{\mu_1}(r_A - \varepsilon), \mathcal{G}_{\mu_2}(r_A - \varepsilon)\}$, which still gives visually good results, see the next example.

Example 6.10. We consider the free additive convolution of a semicircle distribution with the uniform distribution on the interval $[-10, 10]$. We can compute the support of the result analytically and it is $[-10.3497, 10.3497]$. The choice of $\varepsilon = 0.05$ is not small enough to allow us to find

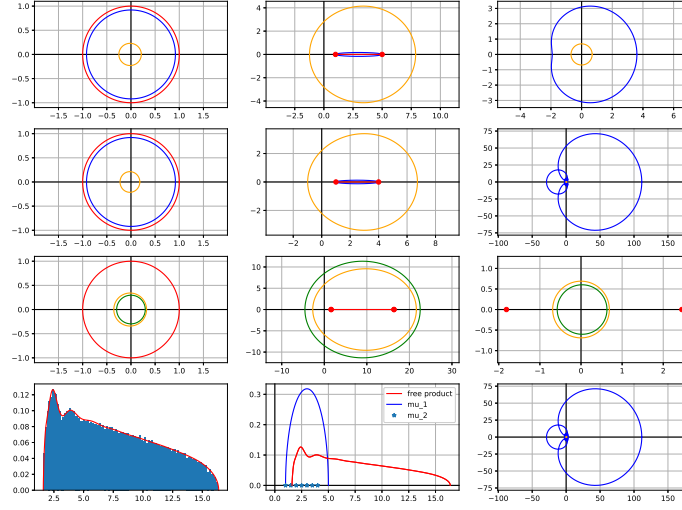


Figure 15: Free multiplication of shifted semicircle distribution with a discrete distribution. See Example 6.9.

the zeros of the derivative of $g(z)$ as defined in (2). Choosing a smaller value $\varepsilon = 0.02$ results in an accurate approximation; see Figure 16.

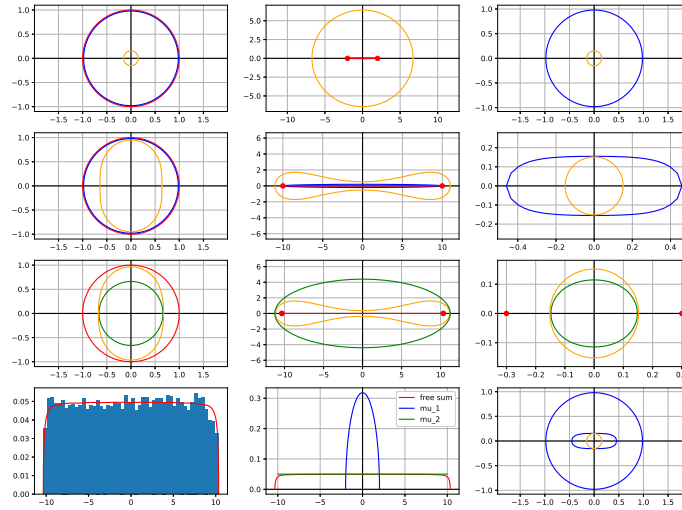


Figure 16: Free addition of semicircle distribution with uniform distribution on $[-20, 20]$. See Example 6.10.

7 Conclusions

In this paper, we have proposed an algorithm that allows us to approximately compute the density of a measure with an sqrt-behavior at the boundary that results from the free (additive or multiplicative) convolution of two measures with compact support. Our methods use numerical quadrature, and the accuracy improves quickly when increasing the number of quadrature points.

We mention three possible directions in which our work could be extended. First, one could find more general assumptions under which the strategy of Theorem 3.2 allows us to compute the support of μ , or prove Theorem 4.2 under the less restrictive assumptions of Theorem 3.2. Indeed, the numerical experiments seem to indicate that Algorithm 2 also works under the more general assumption that one of the input measures has sqrt-behavior at the boundary and the other one is a more general Jacobi measure. Second, it would be nice to extend the method to the case where the support is on more than one interval, for example, when considering the free convolution of a semicircle or Marchenko-Pastur distribution with a discrete measure. Finally, it would be interesting to see whether this approach can be extended to operator-valued free probability theory (see, e.g., [6]), as this would allow for an extension of the algorithm for free additive convolution to an algorithm for the computation of rational functions in free random variables.

Acknowledgements. The authors would like to thank Emmanuel Candès, Jorge Garza Vargas, and Iain Johnstone for inspiring discussions on this work.

A Proofs of the results in Section 2

Proof of Theorem 2.7. Given $z \in \mathbb{C} \setminus [a, b]$ we are going to prove that G admits a series expansion in a neighborhood of z . Let us consider points $w \in \mathbb{C}$ such that

$$|z - w| < \frac{1}{2} \text{dist}(z, [a, b]). \quad (16)$$

Then, for all $z \in [a, b]$ we have $|w - z|/|x - z| < \frac{1}{2}$. Therefore, the series $\sum_{n \geq 0} \left(\frac{w - z}{x - z} \right)^n$ converges uniformly to $\frac{x - z}{x - w}$ in the neighborhood (16). By rearranging the terms, we get that

$$(w - x)^{-1} = - \sum_{n \geq 0} (x - z)^{-n-1} (w - z)^n.$$

Therefore, by Fubini-Tonelli's theorem, we can exchange sum and integral in the definition of the Cauchy transform and we get

$$G(z) = \sum_{n \geq 0} \left[- \int_a^b (x - z)^{-(n+1)} d\mu(x) \right] (w - z)^n$$

for all w in the neighborhood (16). This proves that G is analytic on $\mathbb{C} \setminus [a, b]$.

To prove that G is analytic at ∞ , we need to show that $g(z) := G\left(\frac{1}{z}\right)$ is analytic in zero. We have that

$$g(z) = \int_a^b \frac{z}{1 - zx} d\mu(x).$$

The series $z \cdot \sum_{n \geq 0} (zx)^n$ converges uniformly to $\frac{z}{1-zx}$ for $|z| < \frac{1}{2} \min \left\{ \frac{1}{|a|}, \frac{1}{|b|} \right\}$ (with the convention that $\frac{1}{0} = +\infty$). Therefore, we can exchange sum and integral and we get

$$g(z) = \sum_{n \geq 1} \left(\int_a^b x^{n-1} d\mu(x) \right) z^n = z + \sum_{n \geq 2} \left(\int_a^b x^{n-1} d\mu(x) \right) z^n$$

in the region where $|z| < \frac{1}{2} \min \left\{ \frac{1}{|a|}, \frac{1}{|b|} \right\}$, which proves that G is also analytic in $z = \infty$. \square

The proof of Corollary 2.12 follows directly from the following theorem, which is a slight generalization of Theorem 2.11.

Theorem A.1. *Assume $d\mu(x) = f(x)dx$ for a continuous function f . For any $c < d \in (a, b)$ we have that*

$$-\frac{1}{\pi} \lim_{\substack{z \rightarrow 0 \\ z \in \mathbb{C}^+}} \int_c^d \operatorname{Im}(G(t+z)) dt = \mu([c, d]).$$

Proof. We denote $z = \varepsilon + iy$, with $\varepsilon \in \mathbb{R}$ and $y \in \mathbb{R}^+$. We have that

$$\operatorname{Im}(G(t+z)) = \int_a^b \operatorname{Im} \left(\frac{1}{t + \varepsilon - x + iy} \right) d\mu(x) = \int_a^b \frac{-y}{(t + \varepsilon - x)^2 + y^2} d\mu(x).$$

Thanks to Fubini-Tonelli's theorem, we can write

$$\begin{aligned} \int_c^d \operatorname{Im}(G(t+z)) dt &= \int_a^b \left(\int_c^d -\frac{y}{(t + \varepsilon - x)^2 + y^2} dt \right) d\mu(x) \\ &= - \int_a^b \left(\int_{(c+\varepsilon-x)/y}^{(d+\varepsilon-x)/y} \frac{1}{1 + \sqcup^2} d\sqcup \right) d\mu(x) \\ &= - \int_a^b \left(\arctan \left(\frac{d + \varepsilon - x}{y} \right) - \arctan \left(\frac{c + \varepsilon - x}{y} \right) \right) d\mu(x), \end{aligned}$$

where we made the change of variables $\sqcup = \frac{t+\varepsilon-x}{y}$. Now consider the function $F((\varepsilon, y), x) := \arctan \left(\frac{d+\varepsilon-x}{y} \right) - \arctan \left(\frac{c+\varepsilon-x}{y} \right)$. We have that $|F((\varepsilon, y), x)| \leq 2\pi$ and that $F((\varepsilon, y), x) \rightarrow_{(\varepsilon, y) \rightarrow (0, 0^+)} \mathcal{F}(x)$ where

$$\mathcal{F}(x) = \begin{cases} 0 & \text{if } x \notin [c, d]; \\ \pi & \text{if } x \in (c, d), \end{cases}$$

for all x except for $x \in \{c, d\}$. The function \mathcal{F} is measurable, and it agrees μ -almost everywhere with the μ -almost everywhere existing pointwise limit. Hence we can apply the dominated convergence theorem. This, together with the fact that $d\mu(\{c, d\}) = 0$, implies

$$\lim_{(\varepsilon, y) \rightarrow (0, 0^+)} \int_c^d \operatorname{Im}(G(t + \varepsilon + iy)) dt = - \int_a^b \mathcal{F}(x) d\mu(x) = -\pi \mu([c, d]). \quad \square$$

Proof of Theorem 2.14. Theorem 17 in Chapter 3 of [7] states that if the support is contained in an interval $[-r, r]$, then $R(z)$ is analytic in a disk of radius $\frac{1}{6r}$. Now recall that the function $G(z)$

is analytic on $\mathbb{C} \setminus [a, b]$; by the inverse function theorem for holomorphic functions (Theorem B.1), it is invertible in a neighborhood of a point $z \in \mathbb{C} \setminus [a, b]$ whenever $G'(z) \neq 0$. We have that

$$G'(z) = - \int_a^b \frac{1}{(x-z)^2} d\mu(x),$$

which is always nonzero for $z \in \mathbb{R} \setminus [a, b]$, from which we can conclude the second part of the theorem. \square

B Proof of Theorem 4.2

Theorem B.1 (Inverse function theorem for holomorphic functions [4, Theorem 7.5]). *Let U be an open set in \mathbb{C} , let $f : U \rightarrow \mathbb{C}$ be holomorphic, and let $p \in U$ be such that $f'(p) \neq 0$. Then there exists an open neighborhood V of p such that $f : V \rightarrow f(V)$ is a biholomorphism.*

Consequence: Under the assumptions of Theorem B.1, if $f(p) \in \mathbb{R}$ then there exists precisely one curve in V passing through p on which f is real-valued.

Theorem B.2 (Complex Morse lemma [14]). *Let U be an open set in \mathbb{C} , let $f : U \rightarrow \mathbb{C}$ be a holomorphic function, and let $p \in U$ be such that $f'(p) = 0$ and $f''(p) \neq 0$ (a non-degenerate critical point of f). Then there exist neighborhoods \tilde{U} of p and V of 0, and a bijective holomorphic function $\varphi : V \rightarrow \tilde{U}$ such that $\varphi(0) = p$ and $f(\varphi(w)) = f(p) + w^2$.*

Consequence: Under the assumptions of Theorem B.2, if additionally $f(p) \in \mathbb{R}$, there exist *exactly two* curves passing through p on which f is real-valued. These curves are the images via φ of the intersection of V with the x-axis and y-axis.

Lemma B.3. *Let μ be a measure with sqrt-behavior at the boundary. For z in a neighborhood of 0 we have*

$$T_\mu^{-1}(z) = \frac{1}{z} \mathbb{E}[X_\mu] + \text{analytic function.}$$

Proof. We have that

$$\begin{aligned} T_\mu \left(\frac{1}{z} \right) &= \int_a^b \frac{x}{1/z - x} d\mu(x) = \int_a^b \frac{zx}{1 - zx} d\mu(x) \\ &= \int_a^b \left(\sum_{n=1}^{+\infty} (zx)^n d\mu(x) \right) = \sum_{n=1}^{+\infty} z^n \int_a^b x^n d\mu(x) = z \mathbb{E}[X_\mu] + z^2 f(z), \end{aligned}$$

for some holomorphic function $f(z)$. The fourth inequality follows Fubini-Tonelli's theorem because the series converges absolutely in a neighborhood of zero. Let us apply the inverse function theorem (Theorem B.1) to $\varphi(z) := T_\mu \left(\frac{1}{z} \right)$ in a neighborhood of 0, where $\varphi(0) = 0$ and $\varphi'(0) = \mathbb{E}[X_\mu] > 0$. We get that $\varphi^{-1}(z) = \frac{z}{\mathbb{E}[X_\mu]} + \tilde{\varphi}(z)$ for a holomorphic $\tilde{\varphi}(z)$, therefore

$$T_\mu^{-1}(z) = \frac{\mathbb{E}[X_\mu]}{z} \left(\frac{1}{1 + z\tilde{\varphi}(z)\mathbb{E}[X_\mu]} \right) = \frac{\mathbb{E}[X_\mu]}{z} + \frac{\tilde{\varphi}(z)}{1 + z\varphi(z)\mathbb{E}[X_\mu]},$$

where the function $\frac{\tilde{\varphi}(z)}{1 + z\varphi(z)\mathbb{E}[X_\mu]}$ is holomorphic in a neighborhood of zero. \square

Proposition B.4. *Assume that μ satisfies the Assumptions 4.1. Then $T_\mu([a, b])$ is a bounded continuous curve.*

Proof. There is a similar result for the additive case [8, Proposition A.8]; Proposition B.4 is obtained by substituting $\psi(x)$ with $x\psi(x)$ in all the integrals in the proof of [8, Proposition A.8]. \square

Proposition B.5. *If μ is a measure with sqrt-behavior at the boundary and the support $[a, b]$, then T_μ^{-1} is analytic in a neighborhood of $T_\mu(a)$ and $T_\mu(b)$, with zero derivative and nonzero second derivative.*

Proof. Note that $T_\mu(z) = cG_\nu(z)$ for a suitable constant $c := \int_a^b x d\mu(x) > 0$ and a suitable measure ν defined by $d\nu(x) = \frac{1}{c}x d\mu(x)$. If μ is a measure with sqrt-behavior at the boundary, so is ν . Therefore, we can apply [8, Proposition A.6] to ν and get the stated result. \square

We are now ready to prove Theorem 4.2 by following similar steps to the proof of the additive case [8, Theorem 2.2].

Proof of Theorem 4.2. The set $T_{\mu_1}(\mathbb{C}^-) \cap T_{\mu_2}(\mathbb{C}^-)$ is bounded and has a continuous border by Proposition B.4. Let ν be on the border, and assume without loss of generality that $\nu \in \partial T_{\mu_1}(\mathbb{C}^-)$. This implies that $T_{\mu_1}^{-1}(\nu) \in \mathbb{R}$.

Claim #1: $\mathcal{I}t(\nu) > 0$.

Proof of claim #1: Since $T_{\mu_1}^{-1}(\nu) \in \mathbb{R}$ we need to show that $\mathcal{I}\left(\frac{\nu T_{\mu_2}^{-1}(\nu)}{1+\nu}\right) > 0$.

The function $f(y) := \frac{1}{y} + \frac{1}{y T_{\mu_2}(y)}$ is holomorphic for $y \in \mathbb{C}^-$ (this is true since T_{μ_2} is holomorphic and nonzero here), therefore the function $\tilde{f}(y) := \mathcal{I}(f(y))$ is harmonic. If $y \notin [a, b]$ we have $\tilde{f}(y) = 0$; if $y \in \mathbb{R} \setminus [a, b]$ we have, by the Stiltjes inversion theorem (Theorem 2.11),

$$\lim_{\varepsilon \rightarrow 0^-} \mathcal{I}(T_{\mu_2}(y + \varepsilon I)) = \pi y d\mu_2(y) > 0,$$

therefore $\tilde{f}(y) \leq 0$. By the modulus maximization property of harmonic functions, we have that $\tilde{f}(y) < 0$ for all $y \in \mathbb{C}^-$. Substituting $y = T_{\mu_2}^{-1}(\nu)$ we get $\mathcal{I}\left(\frac{1+\nu}{\nu T_{\mu_2}^{-1}(\nu)}\right) < 0$

and therefore $\mathcal{I}\left(\frac{\nu T_{\mu_2}^{-1}(\nu)}{1+\nu}\right) > 0$, which proves the claim.

By Lemma B.3, we have that, in a neighborhood of zero,

$$t(z) = \frac{z}{1+z} \frac{1}{z^2} \mathbb{E}[X_{\mu_1}] \mathbb{E}[X_{\mu_2}] + f(z) = \frac{C}{z} + f(z) \quad (17)$$

for a positive real constant C and an analytic function $f(z)$, therefore $\mathcal{I}(t(z)) < 0$ in a neighborhood of zero. It follows that there exists a point $\zeta \in T_{\mu_1}^{-1}(\mathbb{C}^-) \cap T_{\mu_2}^{-1}(\mathbb{C}^-)$ in which $t(\zeta) \in \mathbb{R}$. By the consequence of Theorem B.1 there exists a curve Γ in \mathbb{C}^- on which t is real-valued. This curve cannot intersect itself (otherwise, we would have a loop and, by analyticity, the function t would be a constant), is bounded, and cannot touch the part of the boundary in the complex plane. Therefore, it will connect with the real axis; there needs to be exactly one point to the

left of 0 and one point to the right of 0, which we will call ξ_a and ξ_b , respectively. Indeed, if there were two points on the same side of zero, we would get a closed loop on which t is real-valued in a region where t is analytic, leading to t being a constant. For the same reason, Γ is the only curve inside $T_{\mu_1}^{-1}(\mathbb{C}^-) \cap T_{\mu_2}^{-1}(\mathbb{C}^-)$ where t is real-valued. We have proved that Γ divides $T_{\mu_1}^{-1}(\mathbb{C}^-) \cap T_{\mu_2}^{-1}(\mathbb{C}^-)$ into an interior region D where $\mathcal{I}(t(z)) > 0$ and an exterior region such that $\mathcal{I}(t(z)) < 0$.

Claim #2: Let $\xi_b^1 := T_{\mu_1}(b_1)$ and $\xi_b^2 := T_{\mu_2}(b_2)$. Then $\xi_b < \min\{\xi_b^1, \xi_b^2\}$.

Claim #3: $t'(z) > 0$ in a left neighborhood of $\min\{\xi_b^1, \xi_b^2\}$ on the real axis (which includes the point $\min\{\xi_b^1, \xi_b^2\}$).

Proof of claim #3: For $z \in (0, \min\{\xi_b^1, \xi_b^2\})$ we have that

$$t'(z) = \frac{z}{1+z}(T_{\mu_1}^{-1})'(z)T_{\mu_2}^{-1}(z) + \frac{z}{1+z}T_{\mu_1}^{-1}(z)(T_{\mu_2}^{-1})'(z) + \frac{1}{(1+z)^2}T_{\mu_1}^{-1}(z)T_{\mu_2}^{-1}(z).$$

Without loss of generality, we can assume that $\xi_b^1 < \xi_b^2$. In this case $(T_{\mu_1}^{-1})'(\xi_b^1) = 0$ (thanks to Proposition B.5), therefore we need to show that

$$t'(\xi_b^1) = \frac{T_{\mu_1}^{-1}(\xi_b^1)}{1 + \xi_b^1} \left(\frac{\xi_b^1}{T_{\mu_2}'(T_{\mu_2}^{-1}(\xi_b^1))} + \frac{T_{\mu_2}^{-1}(\xi_b^1)}{1 + \xi_b^1} \right) > 0.$$

The first factor, $\frac{T_{\mu_1}^{-1}(\xi_b^1)}{1 + \xi_b^1}$, is positive. Let us look at the second factor. Letting $w = T_{\mu_2}^{-1}(\xi_b^1)$, this is equivalent to show that

$$\frac{T_{\mu_2}(w)}{T_{\mu_2}'(w)} + \frac{w}{1 + T_{\mu_2}(w)} = \frac{T_{\mu_2}(w) + T_{\mu_2}^2(w) + wT_{\mu_2}'(w)}{(1 + T_{\mu_2}(w))T_{\mu_2}'(w)} > 0. \quad (18)$$

- We have $1 + T_{\mu_2}(w) = 1 + \xi_b^1 > 0$.
- We have $T_{\mu_2}'(w) = -\int_{a_2}^{b_2} \frac{x}{(w-x)^2} dx < 0$
- We have

$$\begin{aligned} & T_{\mu_2}(w) + T_{\mu_2}^2(w) + wT_{\mu_2}'(w) \\ &= \int_{a_2}^{b_2} \frac{x}{w-x} dx + \left(\int_{a_2}^{b_2} \frac{x}{w-x} dx \right)^2 - \int_{a_2}^{b_2} \frac{xw}{(w-x)^2} dx \\ &< \int_{a_2}^{b_2} \frac{x}{w-x} dx + \int_{a_2}^{b_2} \frac{x^2}{(w-x)^2} dx - \int_{a_2}^{b_2} \frac{xw}{(w-x)^2} dx = 0, \end{aligned}$$

where we used Jensen's inequality for the convex function $x \mapsto \frac{x}{w-x}$.

Hence, (18) holds, proving the claim.

Proof of claim #2: The function t is analytic in a neighborhood of ξ_b . We have that $t'(z)$ goes to $-\infty$ in a (right) neighborhood of zero (because of (17)), and $t'(z)$ is positive in a (left) neighborhood of $\min\{\xi_b^1, \xi_b^2\}$ because of Claim #3. Therefore,

there exists at least one point in the interval $(0, \xi_b)$ where t' is zero. In each of these points, p the consequence of Morse Lemma (Theorem B.2) tells us that there are two curves emanating from p on which t is real-valued; however, the only real-valued curves are the x -axis and Γ . Therefore there can be only one such point, and it has to be ξ_b .

With a similar argument, defining $\xi_a^1 := T_{\mu_1}(a_1)$ and $\xi_a^2 := T_{\mu_2}(a_2)$, one can show that $\xi_a > \max\{\xi_a^1, \xi_a^2\}$.

Now, $t(z) = T_{\mu}^{-1}(z)$ in a neighborhood of 0, and $t(z)$ is a single-valued function in D , therefore $t(z) = T_{\mu}^{-1}(z)$ inside all of D (this step is because single-valuedness implies that two functions which coincide in an open set coincide everywhere). Since the only real-valued curve contained in $T_{\mu}(\mathbb{C}^-)$ is the boundary, Γ is the boundary of $T_{\mu}(\mathbb{C}^-)$.

We want to use the inverse function theorem on t inside \bar{D} (the closure), except for the points ξ_a and ξ_b . We check that the derivative of $t(z)$ is never zero on this set: we have $t'(z) = \frac{1}{T_{\mu}'(T_{\mu}^{-1}(z))}$ when $z \in D$ (open) and $\mathcal{I}(z) \neq 0$; moreover if $t'(z)$ had other zeros in the segment (ξ_a, ξ_b) or on Γ we would have another real-valued curve and this is impossible. The inverse function theorem (Theorem B.1) implies that t^{-1} exists and is analytic everywhere in $t(D)$, on (a, b) where $a := t(\xi_a)$ and $b := t(\xi_b)$, and on Γ (except for, in general, the points a and b). Note that we can conclude that t^{-1} is also analytic on (a, b) because t is defined, analytic, and invertible also on the segments $(\xi_a, 0)$ and $(0, \xi_b)$ and in a domain D' which is symmetric of D with respect to the real axis.

Let us look at the density function $f(x)$ of μ . We have that

$$xf(x) = \frac{1}{\pi} \lim_{\varepsilon \rightarrow 0^+} \mathcal{I}(t^{-1}(x - i\varepsilon)) \text{ for } x \in (a, b),$$

therefore $f(x)$ is analytic in (a, b) . Let us look at point a (the argument for b is entirely analogous). There are two real-valued curves passing from the point ξ_a , therefore the function t has the form

$$t(z) = a + c_2(z - \xi_a)^2 + (z - \xi_a)^3 h(z)$$

in a neighborhood of ξ_a , for some analytic function $h(z)$, and similarly for ξ_b . By Theorem B.2, there exists a local biholomorphism φ such that $\varphi(0) = \xi_a$ and¹ $t(\varphi(z)) = a - z^2$. Therefore, by setting $w = a - z^2$ we get the expression

$$t^{-1}(w) = T_{\mu}(w) = \varphi(\sqrt{a - w}) = c_0 + c_{1/2}\sqrt{a - w} + c_1(a - w) + c_{3/2}(a - w)^{3/2} + \dots,$$

where all the coefficients are real because $T_{\mu}(w)$ is real for $w \in \mathbb{R} \setminus [a, b]$. Therefore, for $x \in (a, b)$ sufficiently close to a we have

$$\begin{aligned} xf(x) &= \frac{1}{\pi} \lim_{\varepsilon \rightarrow 0^+} \mathcal{I}(T_{\mu}(x - i\varepsilon)) \\ &= \frac{1}{\pi} \lim_{\varepsilon \rightarrow 0^+} \mathcal{I}(\sqrt{a - x + i\varepsilon}(c_{1/2} + c_{3/2}(a - w) + c_{5/2}(a - w)^2 + \dots)) \\ &= \frac{1}{\pi} \lim_{\varepsilon \rightarrow 0^+} \sqrt{x - a} \cdot h(x), \end{aligned}$$

¹The minus sign arises because we want to choose φ in such a way that a right neighborhood of ξ_a on the real axis is sent into a left neighborhood of a on the real axis, and a piece of Γ close to ξ_a is sent to a right neighborhood of a on the real axis.

for a function $h(x)$ that is analytic in a right neighborhood of a . The same argument holds in a neighborhood of b . Therefore, μ is a measure with sqrt-behavior at the boundary, that is, with the form $d\mu(x) = \sqrt{x-a}\sqrt{b-x}\psi(x)$ for an analytic function $\psi(x)$, with an invertible T-transform. This proves Theorem 4.2 and allows us to reconstruct the measure using the truncated series expansion. \square

References

- [1] Z. Bao, L. Erdős, and K. Schnelli. On the support of the free additive convolution. *J. Anal. Math.*, 142(1):323–348, 2020.
- [2] P. J. Davis. On the numerical integration of periodic analytic functions. *On numerical approximation*, pages 21–23, 1959.
- [3] E. Dobriban. Efficient computation of limit spectra of sample covariance matrices. *Random Matrices Theory Appl.*, 4(4):1550019, 36, 2015.
- [4] K. Fritzsche, H. Grauert, and H. Grauert. *From holomorphic functions to complex manifolds*, volume 213. Springer, 2002.
- [5] J. W. Helton, T. Mai, and R. Speicher. Applications of realizations (aka linearizations) to free probability. *J. Funct. Anal.*, 274(1):1–79, 2018.
- [6] T. Mai and R. Speicher. Free probability, random matrices, and representations of non-commutative rational functions. In *Computation and combinatorics in dynamics, stochasticity and control*, volume 13 of *Abel Symp.*, pages 551–577. Springer, Cham, 2018.
- [7] J. A. Mingo and R. Speicher. *Free probability and random matrices*, volume 35 of *Fields Institute Monographs*. Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2017.
- [8] S. Olver and R. R. Nadakuditi. Numerical computation of convolutions in free probability theory. *arXiv preprint arXiv:1203.1958*, 2012.
- [9] N. R. Rao and A. Edelman. The polynomial method for random matrices. *Found. Comput. Math.*, 8(6):649–702, 2008.
- [10] J. W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, 55(2):331–339, 1995.
- [11] J. W. Silverstein and S.-I. Choi. Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.*, 54(2):295–309, 1995.
- [12] L. N. Trefethen. *Approximation Theory and Approximation Practice, Extended Edition*. SIAM, 2019.
- [13] L. N. Trefethen and J. A. C. Weideman. The exponentially convergent trapezoidal rule. *SIAM Rev.*, 56(3):385–458, 2014.
- [14] H. Zoladek. *The monodromy group*, volume 67. Springer Science & Business Media, 2006.