MONOTONE PROBABILITY THEORY

TAKAHIRO HASEBE

Abstract. This is an introduction to monotone probability theory, a kind of noncommutative probability based on the notion of monotone independence for noncommutative random variables. The first part, Sections 1–4, introduces basic materials including noncommutative probability spaces, monotone independence, sums and products of independent random variables, monotone cumulants, the central limit theorem, and Cauchy transform. The second part, Section 5–8, focuses on more advanced topics including a detailed study of monotone convolution, noncommutative stochastic processes, connections to dynamics of holomorphic self-maps of the upper half-plane (in particular, Loewner theory), and applications to outliers of random matrices.

Contents

Preface

Noncommutative probability theory is developing a kind of probability theory by regarding operators as random variables. The mathematical basis of noncommutative probability theory comes from quantum physics, where physical observables are modeled as self-adjoint operators on a Hilbert space and the probability distribution of an observable is defined through the spectral measure evaluated by a state.

An example of noncommutative random variable is a random matrix, i.e., a matrix that has random variables as its entries. Two significant contributions have been made in an early stage of research: Wishart in 1920's applied random matrices as the estimator of the covariance matrix of iid data of normally distributed random vectors; Wigner in 1950's introduced random matrices to model the energy levels of nucleons of nuclei. Since then random matrices have discovered numerous connections and applications to other fields: connections to other mathematics include the distribution of prime numbers, integrable systems (e.g. the Peinlevé equations) [112], and random Young diagrams [37, 95, 128]; applications beyond mathematics include log-gas systems [64], wireless communications [52], quantum information [48] and deep learning [131].

Significant progress in noncommutative probability has been made in the context of operator algebras: Voiculescu initiated free probability theory in 1980's motivated by free group factors. The central idea of free probability is "free independence" for noncommutative random variables [150]. A parallelism lies between probability theory and free probability theory, and various concepts are defined accordingly, e.g., free entropy, free convolution, free central limit theorem, freely infinitely divisible distributions and free cumulants. In addition to operator algebras, free probability has found various connections and applications to other fields including random matrices, representation theory, combinatorics, complex analysis, Hopf algebras and quantum groups. In particular, through random matrices, free probability has been applied beyond mathematics to other fields, see e.g. [48, 52, 63].

Fock spaces form an important aspect of noncommutative probability from physical, probabilistic, and operator-algebraic perspectives. Numerous generalized Fock spaces have been proposed so far. In the 1990's, De Giosa, Lu and Muraki introduced monotone Fock spaces, creation and annihilation operators on them, and a monotone Brownian motion as the sum of these operators [58, 108, 119]. Muraki identified the concept of "monotone probability theory" as implicit in these operators [120, 121]. The building block is "monotone independence" for noncommutative random variables. Again, a parallelism exists between monotone probability and free or classical probability. Progress in this field has uncovered connections to various fields as well. The monotone convolution of probability measures is characterized by the composition of holomorphic self-maps of the complex upper half-plane, thus finding a connection to (holomorphic) dynamical systems. In particular, certain noncommutative stochastic processes correspond to dynamics of holomorphic functions called Loewner chains. As for random matrices, Cébron, Dahlqvist and Gabriel recently applied monotone independence to the analysis of eigenvalues of large random matrices with perturbation [41]. Monotone independence also appears in some graph product [1]. Given increasing new aspects of monotone independence, it is now an appropriate time to offer a detailed expository article to these subjects.

Assuming the familiarity with graduate-level probability theory, functional analysis, complex analysis, and ordinary differential equations, we have tried to make the presentation as self-contained as possible, with minimal references to the literature. The exceptions include the existence of a solution to Hamburger's moment problem (Theorem A.1), Carleman's condition (A.2) for the determinacy of moment sequences, the Lebesgue–Besicovitch differentiation theorem (Lemma 6.10), and Weingarten calculus on the unitary group (Section 8.1).

The structure of this exposition is as follows. Section 1 offers the definition of noncommutative probability spaces with examples, the definition of monotone independence and the calculation of the distribution of the sum and product of monotonically independent random variables.

Section 2 presents a canonical construction of monotonically independent subalgebras on the free product algebra. Some properties of monotone independence, in particular, positivity and associativity, are proved. We also offer a universal construction of monotone independence in the context of graph theory. There are numerous binary operations on graphs, called graph products. One of them is called the comb product and it gives rise to monotone independence.

Section 3 introduces the concept of monotone cumulants of random variables and demonstrates how to calculate the monotone cumulants from the moments of random variables. The monotone cumulants turn out to be useful for investigating the distribution of iid sums of random variables. As applications, the central limit theorem and Poisson's law of small numbers for monotonically iid random variables are established. In particular, the arcsine distribution appears in the monotone CLT instead of the normal distribution in the classical CLT.

Section 4 introduces tools from measure theory and complex analysis. First we establish an integral formula for Nevanlinna functions, i.e., holomorphic functions on the upper half-plane taking values with nonnegative imaginary part. Then the Cauchy transform and several other transforms of probability measures on the real line are investigated. In general, when showing convergence of a sequence of holomorphic functions or measures, we heavily use the technique of "compactness argument". This technique is explained in Sections 4.1 and 4.2.

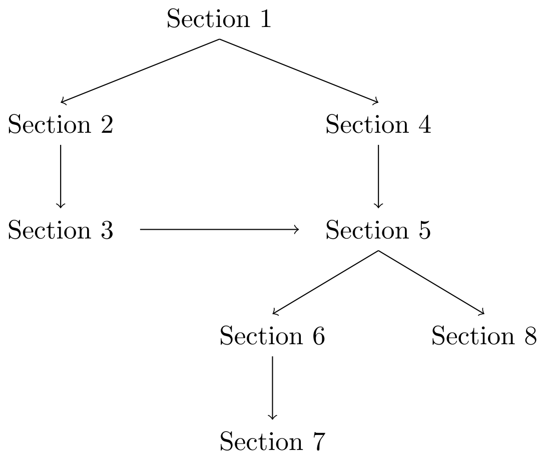

FIGURE 1. Connections of sections

Section 5 investigates additive and multiplicative monotone convolutions for arbitrary probability measures on the basis of complex-analytic methods. We then study various aspects of additive monotone convolution, including the support and moments, convolution semigroups and infinite divisibility.

Section 6 discusses monotone convolution hemigroups that describe the marginal distributions of noncommutative stochastic processes with monotonically independent increments, called monotone additive processes. Their relation to Loewner theory is studied in details. We derive an integral equation and an integro-differential equation satisfied by a Loewner chain.

Section 7 is devoted to constructions of monotone additive processes as operator processes on Hilbert spaces. We present two constructions: one is the sum of three kinds of operators on the monotone Fock space, which generalizes Lu and Muraki's construction of monotone Brownian motion. This construction heavily depends on the integral equation for Loewner chains developed in Section 6. We also offer quite a different construction based on certain classical Markov processes.

Lastly, Section 8 addresses applications of monotone independence to random matrices. For a large square random matrix, a small number of eigenvalues located away from the other eigenvalues are called outliers. Existence or non-existence of outliers for some random matrix models are analyzed by using monotone independence.

Appendices include supplementary results that would be too heavy to place in the main text. The materials here include determinacy of moment sequences, the relationship of convergence of moments and weak convergence of measures, the compositional inverse of reciprocal Cauchy transform, one-parameter semigroups of holomorphic self-maps, and Cauchy's functional equation $f(x+y) = f(x) + f(y)$.

The logical dependence of sections is summarized in Figure 1. Some of the dependence is weak. In particular, Section 5 studies a connection of monotone convolution semigroups to monotone cumulants developed in Section 3, but a considerable part of Section 5 is independent of Section 3. Section 8 uses some results on monotone convolutions in Section 5 but results up to Section 5.1 suffice.

Some topics that are not included but deserve to be mentioned are: a Hopf-algebraic approach to cumulants; refined limit theorems that lead to deep dynamics of iteration of holomorphic self-maps; connections of monotone probability and free probability; C^*-algebras related to monotone probability. Some references to these subjects are provided in Notes 3.6, 3.6, 5.4 and 7.3 respectively. Finally, it is worth noting that the monograph [90] and the expository article [138] are also valuable sources on monotone and other notions of independence written from different perspectives.

I am grateful to Dan Voiculescu who encouraged me to write an introductory exposition of monotone probability theory. I greatly appreciate stimulating discussions with Guillaume Cébron and Katsunori Fujie, and suggestions from Wojciech Młotkowski, Janusz Wysoczański and Franz Lehner.

Sapporo, November 2025
Takahiro Hasebe

Notation

$\mathbb{N}$	the set of positive integers $\{1,2,3,\dots\}$
$\mathbb{N}_0$	the set of nonnegative integers $\{0,1,2,3,\dots\}$
$[n]$	the set $\{1,2,\dots,n\}$ for any $n\in\mathbb{N}$
$\mathbf{1}_A$	the unit of a unital algebra A
$\mathrm{Sp}(a)$	the spectrum $\{z\in\mathbb{C} : z\mathbf{1}_A - a \text{ is not invertible}\}$ of an element a of a unital C^*-algebra
$\langle S\rangle$	the subalgebra generated by a subset S of an algebra
$\mathrm{C}^*\langle S\rangle$	the C^*-subalgebra generated by a subset S of a C^*-algebra
$\overrightarrow{\prod_{t\in T}} a_t$	the ordered product $a_{t_1}a_{t_2}\cdots a_{t_n}$ for a totally ordered set $T=\{t_1<t_2<\cdots<t_n\}$ and elements a_t of an associative algebra A; the ordered product is $\mathbf{1}_A$ if $T=\emptyset$
$\mathbb{B}(H)$	the set of bounded linear operators on a Hilbert space H
$\mathrm{M}_N(\mathbb{C})$	the set of $N\times N$ matrices with complex entries
$C(X)$	the set of $\mathbb{C}$-valued continuous functions on a topological space X
$\mathcal{B}(X)$	the set of the Borel sets of a topological space X
$\mathbb{D}$	the complex unit disk $\{z\in\mathbb{C} : \lvert z\rvert<1\}$
$\mathbb{C}^+$	the complex upper half-plane $\{z\in\mathbb{C} : \Im(z)>0\}$
μ_a	the analytic distribution of a real random variable a
G_μ	the Cauchy transform $\displaystyle\int_{\mathbb{R}}\frac{1}{z-t}\,\mu(\mathrm{d}t)$
F_μ	the reciprocal Cauchy transform $1/G_\mu(z)$
ψ_μ	the moment generating function $\displaystyle\int_{\mathbb{R}}\frac{zt}{1-zt}\,\mu(\mathrm{d}t)=\frac{1}{z}G_\mu\left(\frac{1}{z}\right)-1$
η_μ	the η-transform $\displaystyle\frac{\psi_\mu(z)}{1+\psi_\mu(z)}=1-zF_\mu\left(\frac{1}{z}\right)$
$\arg$	the argument function defined on $\mathbb{C}\setminus[0,+\infty)$ so that $\arg z\in(0,2\pi)$
∇_γ	the sector domain $\{z\in\mathbb{C}^+ : \gamma\lvert\Re(z)\rvert<\Im(z)\}$, $\gamma>0$
χ_B	the characteristic function of a subset B, i.e., $\chi_B(x):=\begin{cases}1, & x\in B,\\ 0, & \text{otherwise}\end{cases}$
$\mathrm{m}_n(\mu)$	the nth moment $\displaystyle\int_{\mathbb{R}} t^n\,\mu(\mathrm{d}t)$ of a Borel measure μ on $\mathbb{R}$, if it exists
$\mathrm{Var}(\mu)$	the variance of a probability measure μ on $\mathbb{R}$
$\triangle$	the set $\{(s,t)\in\mathbb{R}^2 : 0\le s\le t<+\infty\}$
$\mathrm{A}(m,v)$	the arcsine law with mean m and variance $v>0$, i.e., $\displaystyle\frac{1}{\pi\sqrt{2v-(x-m)^2}}\,\chi_{(m-\sqrt{2v},m+\sqrt{2v})}(x)\,\mathrm{d}x$
$\mathrm{S}(m,v)$	the semicircle law with mean m and variance $v>0$, i.e., $\displaystyle\frac{1}{2\pi v}\sqrt{4v-(x-m)^2}\,\chi_{(m-2\sqrt{v},m+2\sqrt{v})}(x)\,\mathrm{d}x$

1. Monotone independence

The standard Kolmogorov's formulation of probability theory builds upon a probability space that is a triplet of a set Ω, a σ-field $\mathcal{F} \subseteq 2^\Omega$ and a probability measure $\mathbb{P}$ defined as a function on $\mathcal{F}$. The first step toward noncommutative probability theory is to shift the focus from the probability space to a function space over it, say $L^\infty := L^\infty(\Omega, \mathcal{F}, \mathbb{P})$. We can regard the expectation $\mathbb{E}$ as a linear functional on L^∞. The central idea is to generalize L^∞ to a possibly noncommutative algebra and $\mathbb{E}$ to a linear functional on it, which leads to the concept of noncommutative probability space. Elements of the algebra are called random variables. In contrast, the usual probability theory is often called classical probability theory, just as mechanics in physics before quantum theory is referred to as classical mechanics.

Numerous notions in probability theory can naturally be formulated in the noncommutative setting. In this section, we will consider the distribution of a random variable, independence of random variables, and the sum and product of independent random variables. A striking feature is that the notion of independence is not unique. What we focus on is the one called "monotone independence".

1.1. Noncommutative probability spaces.

We start by collecting basic materials from functional analysis. Most of the facts on C^*-algebras can be found in standard textbooks, e.g. in [51].

Definition 1.1. Let A be an associative algebra over $\mathbb{C}$, possibly not having a unit.

(i) A is called a $*$-algebra if there exists a map $A \ni a \mapsto a^* \in A$ that satisfies $(a^*)^* = a$, $(ab)^* = b^*a^*$ and $(\lambda a + \mu b)^* = \overline{\lambda}a^* + \overline{\mu}b^*$ for all $\lambda, \mu \in \mathbb{C}$ and all $a, b \in A$.

(ii) A is called a Banach algebra if A is a Banach space with respect to a norm $\|\cdot\|$ and the inequality $\|ab\| \le \|a\|\|b\|$ holds for all $a, b \in A$.

(iii) A is called a C^*-algebra if A is a $*$-algebra and also a Banach algebra such that $\|a^*a\| = \|a\|^2$ holds for all $a \in A$.

For an algebra A, especially when it is non-unital, we define its **unitization** $\tilde{A} := \mathbb{C} \oplus A$ with unit $\mathbf{1}_{\tilde{A}} := (1, 0)$ and multiplication $(\lambda, a)(\mu, b) := (\lambda\mu, \lambda b + \mu a + ab)$. If A is a $*$-algebra then $\tilde{A}$ is a unital $*$-algebra with involution $(\lambda, a)^* := (\overline{\lambda}, a^*)$.

Definition 1.2. (i) A pair (A, φ) of an associative algebra A over $\mathbb{C}$ and a linear functional $\varphi\colon A \to \mathbb{C}$ is called a **noncommutative probability space** (nc-probability space for short). If A is a unital algebra and φ is a unital linear functional, i.e., $\varphi(\mathbf{1}_A) = 1$, then (A, φ) is called a unital nc-probability space.

(ii) Let A be a unital $*$-algebra. A linear functional $\varphi\colon A \to \mathbb{C}$ is called a **state** if $\varphi(\mathbf{1}_A) = 1$ and φ is positive, i.e., $\varphi(a^*a) \ge 0$ for all $a \in A$. Such a pair (A, φ) is called a **unital $*$-probability space**.

(iii) Let A be a (possibly non-unital) $*$-algebra. A linear functional $\varphi\colon A \to \mathbb{C}$ is called a **restricted state** if the extended linear functional $\tilde{\varphi}\colon \tilde{A} \to \mathbb{C}$ with $\tilde{\varphi}(\mathbf{1}_{\tilde{A}}) := 1$ is a state in the sense of (ii). A pair (A, φ) of a $*$-algebra and a restricted state on it is called a **$*$-probability space**.

(iv) Let A be a unital C^*-algebra. If a linear functional φ on A is a state in the sense of (ii) above, then the pair (A, φ) is called a **unital C^*-probability space.** It is known that φ automatically becomes continuous with norm 1.

In any setting above, an element $a \in A$ is called a **random variable** and $\varphi(a)$ is called the **expectation** of a. In a $*$-probability space, an element a with $a^* = a$ is called a **real random variable**.

Remark 1.3. a) On a unital $*$-probability space (A, φ), φ is self-adjoint, i.e., $\varphi(a^*) = \overline{\varphi(a)}$ holds for all $a \in A$. Moreover, the Cauchy–Schwarz inequality

$$|\varphi(a^*b)| \le \sqrt{\varphi(a^*a)}\sqrt{\varphi(b^*b)}$$

holds for all $a, b \in A$. These can be proved from the fact that the quadratic function $\mathbb{C} \ni \lambda \mapsto \varphi((a + \lambda b)^*(a + \lambda b))$ is nonnegative.

b) On a $*$-algebra A, the mere positivity $\varphi(a^*a) \ge 0$, $a \in A$, does not imply that φ is a restricted state. For example, let $\mathbb{C}_0[x]$ be the $*$-algebra of polynomials without constant terms, equipped with the involution $(a_1x + a_2x^2 + \cdots a_nx^n)^* := \overline{a_1}x + \overline{a_2}x^2 + \cdots + \overline{a_n}x^n$. Let $\varphi\colon \mathbb{C}_0[x] \to \mathbb{C}$ be defined linearly by $\varphi(x) := \alpha$ and $\varphi(x^n) := 0$ for all $n \ge 2$. For any $\alpha \in \mathbb{C}$ the positivity condition $\varphi(a^*a) \ge 0$ holds, but φ fails to be a restricted state as soon as $\alpha \neq 0$ because $\tilde{\varphi}((\lambda 1 + x)^*(\lambda 1 + x)) = |\lambda|^2 + 2\alpha\Re(\lambda)$, which fails to be nonnegative for some $\lambda \in \mathbb{C}$.

c) Let A be a unital $*$-algebra and $\varphi\colon A \to \mathbb{C}$ be a unital linear functional. Then φ is a state if and only if φ is a restricted state. First, it is obvious that if φ is a restricted state then it is a state. Conversely, suppose that φ is a state. Let $\mathbf{1}_{\tilde{A}}$ denote the unit of $\tilde{A}$ and $\mathbf{1}_A$ denote that of A. For the unital extension $\tilde{\varphi}\colon \tilde{A} \to \mathbb{C}$ and $a \in A$ we have

$$\tilde{\varphi}((\lambda\mathbf{1}_{\tilde{A}} + a)^*(\lambda\mathbf{1}_{\tilde{A}} + a)) = |\lambda|^2 + \lambda\varphi(a^*) + \overline{\lambda}\varphi(a) + \varphi(a^*a).$$

This is exactly the same as $\varphi((\lambda\mathbf{1}_A + a)^*(\lambda\mathbf{1}_A + a))$, which is nonnegative.

d) We could also define non-unital C^*-probability spaces by requiring φ to be a positive continuous linear functional with norm 1, but we will not need this general setting.

Fundamental examples of noncommutative probability spaces are provided below. The first three correspond to classical probability theory.

Example 1.4. (a) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathbb{E}$ denote the expectation, i.e., the linear functional $X \mapsto \int_\Omega X(\omega)\,\mathbb{P}(d\omega)$. Then the pair $(L^\infty(\Omega, \mathcal{F}, \mathbb{P}), \mathbb{E})$ is a unital C^*-probability space with unit χ_Ω, involution $X \mapsto X^*$ defined by $X^*(\omega) := \overline{X(\omega)}$, and the L^∞ norm $\|\cdot\|$. Alternatively, if we consider the larger space $L^{\infty-} := \bigcap_{1 \le p < +\infty} L^p(\Omega, \mathcal{F}, \mathbb{P})$, then $(L^{\infty-}, \mathbb{E})$ is a unital $*$-probability space.

(b) Let Ω be a compact topological space and $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{B}(\Omega))$. The set $C(\Omega)$ of the $\mathbb{C}$-valued continuous functions on Ω is a unital C^*-algebra equipped with the same unit and involution as above and the supremum norm $\|\cdot\|$. The pair $(C(\Omega), \mathbb{E})$ is a unital C^*-probability space.

(c) Let $\mathbb{C}[x]$ be the polynomial algebra containing the unit. Let $(\alpha_n)_{n\geq 0}$ be any sequence of complex numbers. The linear functional $\varphi\colon \mathbb{C}[x] \to \mathbb{C}$ defined by $\varphi(x^n) := \alpha_n$, $n \in \mathbb{N}_0$, provides a nc-probability space $(\mathbb{C}[x], \varphi)$. Moreover, we introduce an involution on $\mathbb{C}[x]$ by $(x^n)^* := x^n$ and extending it by antilinearity. Let μ be a probability measure on $\mathbb{R}$ that has finite moments of all orders. Then the linear function $\varphi_\mu\colon \mathbb{C}[x] \to \mathbb{C}$,

$$\varphi_\mu(P(x)) := \int_{\mathbb{R}} P(t)\,\mu(\mathrm{d}t)$$

gives a unital $*$-probability space $(\mathbb{C}[x], \varphi_\mu)$.

(d) Let H be a Hilbert space and $\xi \in H$ be a unit vector, i.e., $\|\xi\| = 1$. Let $\mathbb{B}(H)$ be the C^*-algebra of bounded linear operators on H equipped with operator norm and involution being the adjoint. Let $\varphi_\xi(a) := \langle \xi, a\xi\rangle$ called the vector state. Then $(\mathbb{B}(H), \varphi_\xi)$ is a unital C^*-probability space.

(e) Let $\mathrm{M}_N(\mathbb{C})$ be the unital $*$-algebra of $N \times N$ matrices of complex numbers with involution being the conjugate transpose. Let Tr be the canonical trace, i.e., $\mathrm{Tr}(a)$ is the sum of the diagonal entries of a. Then $(\mathrm{M}_N(\mathbb{C}), \frac{1}{N}\mathrm{Tr})$ is a unital $*$-probability space. We can also make it a unital C^*-probability space by naturally identifying $\mathrm{M}_N(\mathbb{C})$ with $\mathbb{B}(\mathbb{C}^N)$.

(f) We generalize example (e) to random matrices. Let $\mathrm{M}_N(L^{\infty-})$ be the unital $*$-algebra of $N \times N$ matrices with entries in $L^{\infty-}$ defined in (a). With the unit $\mathbf{1}_N := (\delta_{i,j})_{i,j\in[N]}$ and involution $a = (X_{i,j})_{i,j\in[N]} \mapsto a^* := (X_{j,i}^*)_{i,j\in[N]}$, the pair $(\mathrm{M}_N(L^{\infty-}), \frac{1}{N}\mathbb{E}\circ\mathrm{Tr})$ becomes a unital $*$-probability space.

1.2. **Distributions of random variables.** In probability theory, for an $\mathbb{R}$-valued random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the distribution of X is the probability measure μ_X on $\mathbb{R}$ defined by

$$\mu_X(B) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}), \qquad B \in \mathcal{B}(\mathbb{R}),$$

where $\mathcal{B}(\mathbb{R})$ is the set of the Borel subsets of $\mathbb{R}$. By the change-of-variable formula, for any bounded continuous function $f\colon \mathbb{R} \to \mathbb{C}$, we have

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\mu_X(\mathrm{d}x).$$

We take this formula as a starting point to define a distribution of a real random variable in the noncommutative setting.

1.2.1. *The case of C^*-probability spaces.* Let a be a real random variable in a unital C^*-probability space (A, φ). Recall that the spectrum

$$\mathrm{Sp}(a) := \{z \in \mathbb{C} : z\mathbf{1}_A - a \text{ is not invertible}\}$$

is a compact subset and is contained in the interval $[-\|a\|, \|a\|]$, and a admits continuous functional calculus, i.e., there is an isometric unital $*$-homomorphism $F_a\colon C(\mathrm{Sp}(a)) \to A$ such that $F_a(P) = P(a)$ for any polynomial P, where $C(\mathrm{Sp}(a))$ is the C^*-algebra endowed with supremum norm. The notation $f(a) := F_a(f)$ is used for all $f \in C(\mathrm{Sp}(a))$.

Proposition 1.5. *Let (A, φ) be a unital C^*-probability space and $a \in A$ be a real random variable. There exists a unique probability measure μ_a on $\mathrm{Sp}(a)$ such that*

$$\varphi(f(a)) = \int_{\mathrm{Sp}(a)} f(x)\mu_a(\mathrm{d}x)$$

for all continuous functions $f\colon \mathrm{Sp}(a) \to \mathbb{C}$. The probability measure μ_a is called the ***distribution*** *of a. Sometimes we call it the* ***analytic distribution*** *of a to distinguish it from the algebraic one in Definition 1.7 below.*

Proof. This is a consequence of Riesz–Markov–Kakutani's theorem applied to the continuous positive linear functional $f \mapsto \varphi(f(a))$. □

Since μ_a is supported on the compact set $\mathrm{Sp}(a)$, μ_a is a unique probability measure on $\mathbb{R}$ such that

$$\varphi(a^n) = \int_{\mathbb{R}} x^n\,\mu_a(\mathrm{d}x), \qquad n \in \mathbb{N}, \tag{1.1}$$

see Proposition A.4.

Example 1.6. (a) Let $a \in \mathrm{M}_N(\mathbb{C})$ be Hermitian. Then $a = udu^*$ for some unitary matrix u and the real diagonal matrix $d = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ of eigenvalues of a, so that

$$\frac{1}{N}\mathrm{Tr}(a^n) = \frac{1}{N}\mathrm{Tr}(d^n) = \frac{1}{N}\sum_{i=1}^N \lambda_i^n = \int_{\mathbb{R}} x^n\,\mu_a(\mathrm{d}x), \qquad n \in \mathbb{N},$$

where $\mu_a := \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$ is called the empirical eigenvalue distribution of a.

(b) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We consider the unital C^*-probability space $(\mathbb{B}(H), \varphi_{\chi_\Omega})$, where $H := L^2(\Omega, \mathcal{F}, \mathbb{P})$ and φ_{χ_Ω} is the vector state. Let X be a bounded real random variable. Let $m_X \in \mathbb{B}(H)$ denote the multiplication operator $f \mapsto Xf$ on H. Then

$$\varphi_{\chi_\Omega}(m_X^n) = \langle \chi_\Omega, m_X^n \chi_\Omega\rangle = \int_\Omega X(\omega)^n\,\mathbb{P}(\mathrm{d}\omega) = \int_{\mathbb{R}} x^n\,\mu_X(\mathrm{d}x), \qquad n \in \mathbb{N},$$

where $\mu_X(\cdot) := \mathbb{P}(X \in \cdot)$ is the distribution of X. Therefore, μ_{m_X} coincides with μ_X.

(c) Suppose that (A, φ) is a unital C^*-probability space and $a \in A$ is a real random variable. If φ is a homomorphism on $\langle a \rangle$ (the algebra generated by a), then the analytic distribution μ_a is the delta measure $\delta_{\varphi(a)}$, as the nth moment of μ_a equals $\varphi(a)^n$.

1.2.2. *The cases of $*$- and nc-probability spaces.* We now turn to the setting of a $*$-probability space (A, φ). Let $(\tilde{A}, \tilde{\varphi})$ be its unital extension. Let $a \in A$ be a real random variable. Then the sequence $s_n := \tilde{\varphi}(a^n)$, $n = 0, 1, 2, \dots$ is positive semi-definite, i.e., for every $n \in \mathbb{N}_0$ and $c_0, c_1, \dots, c_n \in \mathbb{R}$ we have

$$\sum_{i,j=0}^{n} c_i c_j s_{i+j} = \tilde{\varphi}\left(\left(\sum_{i=0}^{n} c_i a^i\right)^* \left(\sum_{j=0}^{n} c_j a^j\right)\right) \geq 0.$$

This ensures the existence of a probability measure μ_a on $\mathbb{R}$ having finite moments of all orders such that (1.1) holds; see Theorem A.1. In general, however, the probability measure μ_a is not unique, see Example A.3. For this reason, we will call the sequence $(\varphi(a^n))_{n\in\mathbb{N}}$ itself the distribution of a instead of μ_a. More generally, we extend this term to elements of a nc-probability space.

Definition 1.7. Let (A, φ) be a nc-probability space. For any $a \in A$, the sequence $(\varphi(a^n))_{n\in\mathbb{N}}$ is called the **distribution** of a. Each number $\varphi(a^n)$ is called the n**th moment of** a**.**

Relying on this definition, we can say that a and b are **identically distributed**, or have an identical distribution, if $\varphi(a^n) = \varphi(b^n)$ for all $n \in \mathbb{N}$. More generally, given two nc-probability spaces (A, φ) and (B, ψ), we may say that $a \in A$ and $b \in B$ are identically distributed if $\varphi(a^n) = \psi(b^n)$ for all $n \in \mathbb{N}$.

1.2.3. *Spectral measures and analytic distributions.* For bounded self-adjoint operators on complex Hilbert spaces, we can show that the analytic distribution is the evaluation of the spectral measure by a state. Although this fact is not essential in the following, this is a core idea of noncommutative probability and of quantum physics, and so it is worth noting here. Let us first recall the concept of spectral measure.

Definition 1.8. Let a be a bounded self-adjoint operator on a complex Hilbert space H. The **spectral measure of** a is a function E_a defined on $\mathcal{B}(\mathrm{Sp}(a))$ taking values in the set of orthogonal projections on H, such that

(i) $E_a(\mathrm{Sp}(a)) = \mathrm{id}_H$,

(ii) for disjoint sets $B_n \in \mathcal{B}(\mathrm{Sp}(a))$, $n \in \mathbb{N}$, the identity

$$E_a\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} E_a(B_n)$$

holds, where the infinite sum converges in the sense of strong operator topology,

(iii) it holds for all $\xi \in H$ that

$$\langle \xi, a\xi \rangle = \int_{\mathrm{Sp}(a)} x \, \mathrm{d}\langle \xi, E_a(x)\xi \rangle,$$

where the right-hand side is the integral against the Borel measure $B \mapsto \langle \xi, E_a(B)\xi \rangle$. This formula is often written more simply as

$$a = \int_{\mathrm{Sp}(a)} x \, E_a(\mathrm{d}x)$$

and is called the spectral decomposition of a.

It is known that a spectral measure exists; the reader is referred to e.g. [139, Theorem 5.1 and Proposition 5.10] or [51, Theorem IX, 2.2].

The spectral measure offers Borel functional calculus: for any bounded (Borel) measurable function $f\colon \mathrm{Sp}(a) \to \mathbb{C}$ we can define $f(a)$ by the **spectral integral**

$$f(a) := \int_{\mathrm{Sp}(a)} f(x) \, E_a(\mathrm{d}x). \tag{1.2}$$

This satisfies several properties e.g., $\chi_{\mathrm{Sp}(a)}(a) = \mathrm{id}_H$, $(fg)(a) = f(a)g(a)$, and

$$\langle \xi, f(a)\xi \rangle = \int_{\mathrm{Sp}(a)} f(x) \, \mathrm{d}\langle \xi, E_a(x)\xi \rangle, \qquad \xi \in H. \tag{1.3}$$

One can also interpret that (1.3) is the definition of the spectral integral, i.e., the spectral integral $f(a)$ in (1.2) is a unique bounded operator such that (1.3) holds; recall the well known fact that $\langle \xi, b\xi \rangle = 0$ for all $\xi \in H$ implies $b = 0$ if H is a complex Hilbert space.

Proposition 1.9. *Let a be a bounded self-adjoint operator on a Hilbert space H and $\xi \in H$ be a unit vector. Let E_a be the spectral measure of a. Then the analytic distribution of a in the C^*-probability space $(\mathbb{B}(H), \varphi_\xi)$ is given by $\mu_a = \varphi_\xi \circ E_a$.*

Proof. The definition of spectral measure implies that $\mu := \varphi_\xi \circ E_a$ is a probability measure on $\mathbb{R}$. By Borel functional calculus we have

$$\varphi_\xi(a^n) = \langle \xi, a^n \xi \rangle = \int_{\mathrm{Sp}(a)} x^n \, \mathrm{d}\langle \xi, E_a(x)\xi \rangle = \int_{\mathrm{Sp}(a)} x^n \, \mu(\mathrm{d}x), \qquad n \in \mathbb{N}_0,$$

which shows $\mu = \mu_a$. ∎

Remark 1.10. The spectral measure exists also for unbounded self-adjoint operators on Hilbert spaces. Then we can define the distribution of the operator to be the evaluation of the spectral measure by a state. This is a standard method in noncommutative probability to handle arbitrary probability measures on the real line.

The analytic distributions in Example 1.6 can be understood via spectral measures.

Example 1.11. In the setting of Example 1.6 (a), we identify $\mathrm{M}_N(\mathbb{C})$ with $\mathbb{B}(\mathbb{C}^N)$. Let $\lambda'_1, \lambda'_2, \dots, \lambda'_M$ $(1 \le M \le N)$ be the eigenvalues of a without counting multiplicities. The spectrum of a equals the finite set

$$\mathrm{Sp}(a) = \{\lambda'_1, \dots, \lambda'_M\}.$$

Let E_i be the orthogonal projection from $\mathbb{C}^N$ onto the eigenspace of the eigenvalue λ'_i, i.e.,

$$E_i \xi = \sum_{j=1}^{m_i} \langle u_{i,j}, \xi \rangle u_{i,j},$$

where $\{u_{i,1}, u_{i,2}, \dots, u_{i,m_i}\}$ is an orthonormal basis of the eigenspace. We show that

$$E(B) := \sum_{i \colon \lambda'_i \in B} E_i, \qquad B \subseteq \mathrm{Sp}(a)$$

is the spectral measure of a. First, as is well known in linear algebra, the vectors $\{u_{i,j} : 1 \le j \le m_i,\ 1 \le i \le M\}$ form a basis of $\mathbb{C}^N$. Since every vector $\xi \in \mathbb{C}^N$ can be expressed as the linear combination

$$\xi = \sum_{i=1}^{M} \sum_{j=1}^{m_i} \langle u_{i,j}, \xi \rangle u_{i,j},$$

one can easily see that

$$E(\mathrm{Sp}(a)) = \sum_{i=1}^{M} E_i = \mathrm{id}_{\mathbb{C}^N}.$$

Second, observing $\langle \xi, E(B)\xi \rangle = \sum_{i \colon \lambda'_i \in B} \sum_{j=1}^{m_i} |\langle \xi, u_{i,j} \rangle|^2$ we see that the Borel measure $\langle \xi, E(\cdot)\xi \rangle$ can be expressed as $\langle \xi, E(\cdot)\xi \rangle = \sum_{i=1}^{M} (\sum_{j=1}^{m_i} |\langle \xi, u_{i,j} \rangle|^2) \delta_{\lambda'_i}$. We thus obtain

$$\begin{aligned} \int_{\mathrm{Sp}(a)} x \, \mathrm{d}\langle \xi, E(x)\xi \rangle &= \sum_{i=1}^{M} \left(\sum_{j=1}^{m_i} |\langle \xi, u_{i,j} \rangle|^2 \right) \lambda'_i = \sum_{i=1}^{M} \sum_{j=1}^{m_i} \langle \xi, a u_{i,j} \rangle \langle u_{i,j}, \xi \rangle \\ &= \sum_{i=1}^{M} \sum_{j=1}^{m_i} \langle a^* \xi, u_{i,j} \rangle \langle u_{i,j}, \xi \rangle = \langle a^* \xi, \xi \rangle = \langle \xi, a \xi \rangle, \end{aligned}$$

where we used the basic relations $a u_{i,j} = \lambda'_i u_{i,j}$ and $\langle \xi, \eta \rangle = \sum_{k=1}^{N} \langle \xi, v_k \rangle \langle v_k, \eta \rangle$ for any orthonormal basis $(v_k)_{k \in [N]}$. Therefore we have verified that E is the spectral measure.

Finally, the analytic distribution μ_a in the C^*-probability space $(\mathrm{M}_N(\mathbb{C}), \frac{1}{N}\mathrm{Tr})$ can be recovered as

$$\mu_a(B) = \frac{1}{N} \mathrm{Tr}(E(B)) = \frac{1}{N} \sum_{i \colon \lambda'_i \in B} \mathrm{Tr}(E_i) = \frac{1}{N} \sum_{i \colon \lambda'_i \in B} m_i = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}(B).$$

Example 1.12. In the setting of Example 1.6 (b), observe first that the spectrum of m_X is exactly the support S of μ_X (see Section 4.1 for the definition of support). We show that the spectral measure of the multiplication operator m_X is given by

$$E(B) := m_{\chi_B(X)}, \qquad B \in \mathcal{B}(S).$$

First, it is obvious that $E(S) = \mathrm{id}_H$ as $\mathbb{P}(X \in S) = 1$. Second, $E(B)$ is an orthogonal projection as one can easily check $E(B) = E(B)^2 = E(B)^*$. Third, for disjoint Borel subsets $B_n \in \mathcal{B}(S)$ and $\xi \in H$ we have

$$E\left(\bigcup_{n=1}^{\infty} B_n \right) \xi = \chi_{\bigcup_{n \in \mathbb{N}} B_n}(X) \xi = \sum_{n=1}^{\infty} \chi_{B_n}(X) \xi = \sum_{n=1}^{\infty} E(B_n) \xi,$$

where the second equality holds (in the L^2 sense) by the dominated convergence theorem. Finally, for $\xi \in H$, the measure $B \mapsto \langle \xi, E(B)\xi \rangle$ is given by $\mathbb{E}[\chi_B(X)|\xi|^2] = \mathbb{E}[\delta_X(B)|\xi|^2]$. If $f = \chi_B$ with $B \in \mathcal{B}(S)$, then by the definition of integral

$$\int_S f(x) \, \mathrm{d}\langle \xi, E(x)\xi \rangle = \mathbb{E}[\chi_B(X)|\xi|^2] = \mathbb{E}[f(X)|\xi|^2] = \langle \xi, f(X)\xi \rangle.$$

For a nonnegative bounded Borel measurable function f, we approximate it by simple functions from below. By the monotone convergence theorem, the same formula $\int_S f(x) \, \mathrm{d}\langle \xi, E(x)\xi \rangle = \langle \xi, f(X)\xi \rangle$ still holds. By linearity, the same holds for any bounded Borel measurable function f. In particular, selecting $f(x) = x$ we conclude that E is indeed the spectral measure of m_X. Note that the above arguments actually show that $f(m_X) = m_{f(X)}$.

Using the spectral measure, we recover the analytic distribution of m_X in Example 1.6 (b)

$$\mu_{m_X}(B) = \varphi_{\chi_\Omega}(E(B)) = \langle \chi_\Omega, \chi_B(X)\chi_\Omega \rangle = \mathbb{P}(X \in B),$$

so that μ_{m_X} coincides with the distribution of X in the sense of classical probability.

1.3. **Independence of two subalgebras.** On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the independence of two sub-σ-fields $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ is defined by the condition

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \qquad A \in \mathcal{G},\ B \in \mathcal{H}.$$

This is also equivalent to the condition that

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y], \qquad X \in L^\infty(\Omega, \mathcal{G}, \mathbb{P}),\ Y \in L^\infty(\Omega, \mathcal{H}, \mathbb{P}). \tag{1.4}$$

The latter condition can be regarded as a certain relation between the subalgebras $L^\infty(\Omega, \mathcal{G}, \mathbb{P})$ and $L^\infty(\Omega, \mathcal{H}, \mathbb{P})$ of $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$.

Let us generalize the independence of the form (1.4) to a nc-probability space (A, φ). From the above observation, it is natural that subalgebras of A should play the role of sub-σ-fields in probability theory. Let B, C be subalgebras of A. Because random variables are allowed to be noncommuting, focusing only on the expectation $\varphi(bc)$, $b \in B$, $c \in C$, is not sufficient; it is natural to discuss the more general quantities

$$\begin{aligned} &\varphi(b_1c_1b_2\cdots c_{n-1}b_n), && \varphi(b_1c_1b_2\cdots c_{n-1}b_nc_n), \\ &\varphi(c_0b_1c_1b_2\cdots c_{n-1}b_n), && \varphi(c_0b_1c_1b_2\cdots c_{n-1}b_nc_n), \qquad b_i \in B,\ c_i \in C, \end{aligned}$$

which we call **mixed moments of** B **and** C. To discuss the above four kinds of alternating words in a unified way, we consider $\varphi(c_0b_1c_1\cdots c_{n-1}b_nc_n)$ allowing $c_0 = \mathbf{1}_{\tilde{A}}$ or $c_n = \mathbf{1}_{\tilde{A}}$ in the unitized algebra $\tilde{A}$ (if A is unital then we can just take $c_0 = \mathbf{1}_A$ or $c_n = \mathbf{1}_A$). Inspired by (1.4), we consider a "universal rule" for calculating the mixed moments of B and C, and call it an independence of B and C in noncommutative probability theory.

A direct generalization of the classical stochastic independence (1.4) to a nc-probability space is as follows.

Definition 1.13. Let (A, φ) be a nc-probability space. Subalgebras $B, C \subseteq A$ are called **tensor independent** if

$$\varphi(c_0b_1c_1b_2c_2\cdots b_nc_n) = \varphi(b_1b_2\cdots b_n)\varphi(c_0c_1c_2\cdots c_n)$$

holds for all $n \geq 1$, $b_1, b_2, \dots, b_n \in B$, $c_1, c_2, \dots, c_{n-1} \in C$ and $c_0, c_n \in C \cup \{\mathbf{1}_{\tilde{A}}\}$. In case $n = 1$ and $c_0 = c_n = \mathbf{1}_{\tilde{A}}$ we understand $\varphi(c_0c_n) = 1$.

Two subsets $S, T \subseteq A$ are said to be tensor independent if the subalgebras $B := \langle S \rangle$ and $C := \langle T \rangle$ are tensor independent. If (A, φ) is a $*$-probability space, then two subsets $S, T \subseteq A$ are called $*$-tensor independent if the $*$-subalgebras $B := \langle b, b^* : b \in S \rangle$ and $C := \langle c, c^* : c \in T \rangle$ are tensor independent.

Example 1.14 (A canonical model for the tensor independence)**.** Let (A_i, φ_i), $i = 1, 2$, be two nc-probability spaces. Let $A := A_1 \otimes A_2$ and $\varphi := \varphi_1 \otimes \varphi_2$. Then the subalgebras $B := A_1 \otimes \mathbf{1}_{A_2}$ and $C := \mathbf{1}_{A_1} \otimes A_2$ are tensor independent in (A, φ). To see this, for example for $b_i = x_i \otimes \mathbf{1}_{A_2}$, $i = 1, 2$, and $c_i = \mathbf{1}_{A_2} \otimes y_i$, $i = 1, 2, 3$, we have

$$c_0b_1c_1b_2c_2 = (x_1x_2) \otimes (y_0y_1y_2)$$

and hence

$$\varphi(c_0b_1c_1b_2c_2) = \varphi_1(x_1x_2)\varphi_2(y_0y_1y_2) = \varphi(b_1b_2)\varphi(c_0c_1c_2).$$

Here we introduce monotone independence as another factorization formula for mixed moments.

Definition 1.15. Let (A, φ) be a nc-probability space. Subalgebras $B, C \subseteq A$ are called **monotonically independent** if

$$\varphi(c_0b_1c_1b_2c_2\cdots b_nc_n) = \varphi(b_1b_2\cdots b_n)\varphi(c_0)\varphi(c_1)\varphi(c_2)\cdots\varphi(c_n) \tag{1.5}$$

for all $n \geq 1$, $b_1, b_2, \dots, b_n \in B$ and $c_1, c_2, \dots, c_{n-1} \in C$ and $c_0, c_n \in C \cup \{\mathbf{1}_{\tilde{A}}\}$. Here we understand $\varphi(\mathbf{1}_{\tilde{A}}) = 1$.

Two subsets $S, T \subseteq A$ are called monotonically independent if the subalgebras $B := \langle S \rangle$ and $C := \langle T \rangle$ are monotonically independent. If (A, φ) is a $*$-probability space, then subsets $S, T \subseteq A$ are called $*$-monotonically independent if the $*$-subalgebras $B := \langle b, b^* : b \in S \rangle$ and $C := \langle c, c^* : c \in T \rangle$ are monotonically independent.

Remark 1.16. a) Monotone independence has an "asymmetric" nature: B and C being monotonically independent does not imply that C and B are monotonically independent.

b) If B, C are monotonically independent and (A, φ) is unital, then one can show that B and $\langle \mathbf{1}_A, C \rangle$ are also monotonically independent, i.e., (1.5) holds for $c_0, c_1, \dots, c_n \in C \cup \{\mathbf{1}_A\}$ too. However, if one takes $b_i = \mathbf{1}_A$ for some i then (1.5) holds only in trivial cases. Suppose for example that $\varphi(c_0b_1c_1) = \varphi(b_1)\varphi(c_0)\varphi(c_1)$ were the case for $b_1 = \mathbf{1}_A$. Then we would have $\varphi(c_0c_1) = \varphi(c_0)\varphi(c_1)$ for all $c_0, c_1 \in C$, i.e., φ would be a homomorphism on C. In the setting of C^*-probability spaces, this would imply that any real random variable $c \in C$ has the trivial distribution $\delta_{\varphi(c)}$, see Example 1.6 (c).

c) The above b) suggests that the role of the unit $\mathbf{1}_A$ is rather different from classical probability theory. Here are two more remarks on the unit. In probability theory, a constant random variable is independent of any other random variable. In monotone probability, it is easy to see that a and $\mathbf{1}_A$ are monotonically independent for any $a \in A$; however, typically $\mathbf{1}_A$ and a are not monotonically independent. Also we can observe that the monotone independence of b and c does not imply that of $b + \lambda\mathbf{1}_A$ and c for $\lambda \in \mathbb{C} \setminus \{0\}$.

d) Suppose (A, φ) is a unital C^*-probability space. If B, C are $*$-subalgebras of A that are monotonically independent, then we can prove (1.5) for all $n \geq 1$, $b_1, b_2, \dots, b_n \in \mathrm{C}^*\langle B \rangle = \overline{B}$ and $c_0, c_1, \dots, c_{n-1} \in \mathrm{C}^*\langle \mathbf{1}_A, C \rangle$. Moreover, when $c_0 = c_n = \mathbf{1}_A$, the left- and right-most letters b_1 and b_n are allowed to be the unit, i.e., the formula

$$\varphi(b_1c_1b_2c_2\cdots c_{n-1}b_n) = \varphi(b_1b_2\cdots b_n)\varphi(c_1)\varphi(c_2)\cdots\varphi(c_{n-1}) \tag{1.6}$$

holds for all $n \geq 2$, $b_1, b_n \in \mathrm{C}^*\langle \mathbf{1}_A, B \rangle$, $b_2, \dots, b_{n-1} \in \mathrm{C}^*\langle B \rangle$ and $c_1, c_2, \dots, c_{n-1} \in \mathrm{C}^*\langle \mathbf{1}_A, C \rangle$.

Example 1.17 (A canonical model for monotone independence)**.** Let H_1, H_2 be Hilbert spaces. We fix arbitrary unit vectors ξ_i in H_i, $i = 1, 2$. Let $H := H_1 \otimes H_2$ be the tensor product Hilbert space, $\xi := \xi_1 \otimes \xi_2$, and $p \in \mathbb{B}(H_2)$ be the orthogonal projection onto the one-dimensional subspace $\mathbb{C}\xi_2$. We consider the unital C^*-probability space (A, φ), where $A = \mathbb{B}(H)$ and $\varphi(a) := \langle \xi, a\xi \rangle$, $a \in A$, and the $*$-subalgebras $B := \mathbb{B}(H_1) \otimes p$ and $C := \mathrm{id}_{H_1} \otimes \mathbb{B}(H_2)$ of A. Note that B is a $*$-subalgebra because of $p^2 = p = p^*$, and moreover, B, C are both C^*-subalgebras (i.e., closed with respect to the operator norm) because $\|x \otimes y\| = \|x\|\|y\|$ holds for all $x \in \mathbb{B}(H_1)$ and $y \in \mathbb{B}(H_2)$.

We show that B and C are monotonically independent in (A, φ). To see this, for example for $b_i = \tilde{b}_i \otimes p$, $i = 1, 2$, and $c_i = \mathrm{id}_{H_1} \otimes \tilde{c}_i$, $i = 1, 2, 3$, we have

$$c_0 b_1 c_1 b_2 c_2 = (\tilde{b}_1 \tilde{b}_2) \otimes (\tilde{c}_0 p \tilde{c}_1 p \tilde{c}_2)$$

and hence, with notation $\varphi_i(\cdot) = \langle \xi_i, \cdot\, \xi_i \rangle_{H_i}$,

$$\varphi(c_0 b_1 c_1 b_2 c_2) = \varphi_1(\tilde{b}_1 \tilde{b}_2) \varphi_2(\tilde{c}_0 p \tilde{c}_1 p \tilde{c}_2).$$

Note here that, since $p\xi_2 = p^*\xi_2 = \xi_2$, we have

$$\varphi_2(\tilde{c}_0 p \tilde{c}_1 p \tilde{c}_2) = \varphi_2(p \tilde{c}_0 p \tilde{c}_1 p \tilde{c}_2 p).$$

Straightforward calculations yield $p\tilde{c}p = \varphi_2(\tilde{c})p$ for any $\tilde{c} \in \mathbb{B}(H_2)$. Hence, we arrive at

$$\varphi_2(\tilde{c}_0 p \tilde{c}_1 p \tilde{c}_2) = \varphi_2(\tilde{c}_0)\varphi_2(\tilde{c}_1)\varphi_2(\tilde{c}_2) = \varphi(c_0)\varphi(c_1)\varphi(c_2)$$

and finally

$$\varphi(c_0 b_1 c_1 b_2 c_2) = \varphi(b_1 b_2)\varphi(c_0)\varphi(c_1)\varphi(c_2).$$

Example 1.18. A simpler example can be constructed on a single Hilbert space H equipped with a unit vector $\xi \in H$. Let p be the orthogonal projection onto $\mathbb{C}\xi$ and let $\varphi(\cdot) := \langle \xi, \cdot\, \xi \rangle$. We show that the sets $\{p\}$ and $\mathbb{B}(H)$ are monotonically independent in $(\mathbb{B}(H), \varphi)$. Since the $*$-algebra generated by p is just $\mathbb{C}p$, it suffices to compute $\varphi(a_1 p a_2 p \cdots p a_n)$, where $a_1, a_2, \dots, a_n \in \mathbb{B}(H)$. Using the relations $pap = \varphi(a)p$, $p\xi = \xi$ and $\varphi(p^n) = 1$, $n \in \mathbb{N}$, we can see that

$$\varphi(a_1 p a_2 p \cdots p a_n) = \varphi(p a_1 p a_2 \cdots a_n p) = \varphi(a_1)\varphi(a_2) \cdots \varphi(a_n)\varphi(p)$$

as desired.

1.4. **Independence of several subalgebras.** We extend the definition of tensor and monotone independence to the case of several subalgebras. For the tensor case, the definition comes from the following natural extension of Example 1.14.

Example 1.19. Let (A_i, φ_i), $i \in [N]$, be unital nc-probability spaces. Let

$$A := A_1 \otimes A_2 \otimes \cdots \otimes A_N,$$
$$\varphi := \varphi_1 \otimes \varphi_2 \otimes \cdots \otimes \varphi_N.$$

We consider the subalgebras

$$B_i := \mathbf{1}_{A_1} \otimes \mathbf{1}_{A_2} \otimes \cdots \otimes \mathbf{1}_{A_{i-1}} \otimes A_i \otimes \mathbf{1}_{A_{i+1}} \otimes \cdots \otimes \mathbf{1}_{A_N} \tag{1.7}$$

for $i \in [N]$. Then for any $i_1, i_2, \dots, i_n \in [N]$ and $b_1 \in B_{i_1}$, $b_2 \in B_{i_2}$, $\dots$, $b_n \in B_{i_n}$ we have

$$\varphi(b_1 b_2 \cdots b_n) = \varphi\left(\overrightarrow{\prod_{p:\, i_p = 1}} b_p\right) \varphi\left(\overrightarrow{\prod_{p:\, i_p = 2}} b_p\right) \cdots \varphi\left(\overrightarrow{\prod_{p:\, i_p = N}} b_p\right).$$

The above example can be abstracted to any family of subalgebras of any nc-probability space as follows.

Definition 1.20. Let (A, φ) be a nc-probability space. A family of subalgebras $(A_i)_{i \in I}$ of A is called **tensor independent** if for any $i_1, i_2, \dots, i_n \in I$ and $a_1 \in A_{i_1}$, $a_2 \in A_{i_2}$, $\dots$, $a_n \in A_{i_n}$, we have

$$\varphi(a_1 a_2 \cdots a_n) = \prod_{j \in I} \varphi\left(\overrightarrow{\prod_{p:\, i_p = j}} a_p\right).$$

Moreover, a family $(S_i)_{i \in I}$ of subsets of A is said to be tensor independent if so is $(A_i)_{i \in I}$, where $A_i = \langle S_i \rangle$ is the subalgebra generated by S_i. Independence of random variables $(x_i)_{i \in I}$ can be defined by regarding each x_i as the set of single element.

Example 1.17 can also be extended to an arbitrary number of subalgebras. Given Hilbert spaces H_i with unit vectors ξ_i $(1 \le i \le N)$, we set

$$H := H_1 \otimes H_2 \otimes \cdots \otimes H_N,$$
$$\xi := \xi_1 \otimes \xi_2 \otimes \cdots \otimes \xi_N,$$

and $A := \mathbb{B}(H)$, $\varphi(a) := \langle \xi, a\xi \rangle_H$, which form a unital C^*-probability space (A, φ). Let $p_i \in \mathbb{B}(H_i)$ be the orthogonal projection onto $\mathbb{C}\xi_i$. Then we consider the $*$-subalgebras

$$\begin{aligned} A_1 &:= \mathbb{B}(H_1) \otimes p_2 \otimes p_3 \otimes \cdots \otimes p_N, \\ A_2 &:= \mathrm{id}_{H_1} \otimes \mathbb{B}(H_2) \otimes p_3 \otimes \cdots \otimes p_N, \\ A_3 &:= \mathrm{id}_{H_1} \otimes \mathrm{id}_{H_2} \otimes \mathbb{B}(H_3) \otimes \cdots \otimes p_N, \\ &\ \ \vdots \\ A_N &:= \mathrm{id}_{H_1} \otimes \mathrm{id}_{H_2} \otimes \cdots \otimes \mathrm{id}_{H_{N-1}} \otimes \mathbb{B}(H_N). \end{aligned} \tag{1.8}$$

This operator model leads to the following definition; see also Example 1.23.

Definition 1.21. Let (A, φ) be a nc-probability space and I be a totally ordered set. A family of subalgebras $(A_i)_{i \in I}$ of A is called **monotonically independent** if for any $i_1, i_2, \dots, i_n \in I$ and $a_1 \in A_{i_1}$, $a_2 \in A_{i_2}$, $\dots$, $a_n \in A_{i_n}$, we have

$$\varphi(a_1 a_2 \cdots a_n) = \begin{cases} \varphi(a_k)\varphi(a_1 a_2 \cdots a_{k-1} a_{k+1} \cdots a_n) & \text{if } 2 \le k \le n-1 \text{ and } i_{k-1} < i_k > i_{k+1}, \\ \varphi(a_1)\varphi(a_2 a_3 \cdots a_n) & \text{if } i_1 > i_2, \\ \varphi(a_n)\varphi(a_1 a_2 \cdots a_{n-1}) & \text{if } i_{n-1} < i_n. \end{cases}$$

Moreover, a family $(S_i)_{i \in I}$ of subsets of A is called monotonically independent if so is $(A_i)_{i \in I}$, where $A_i = \langle S_i \rangle$ is the subalgebra generated by S_i. If (A, φ) is a $*$-probability space, then a family $(S_i)_{i \in I}$ of subsets of A is said to be $*$-monotonically independent if the $*$-subalgebras A_i generated by S_i are monotonically independent.

Remark 1.22. This is a recursive definition of monotone independence. Applying the definition repeatedly, the mixed moment $\varphi(a_1 a_2 \cdots a_n)$ will eventually be of the form

$$\prod_{j=1}^{m} \varphi(a_{k_1(j)} a_{k_2(j)} \cdots a_{k_{i(j)}(j)}),$$

where $k_1(j) < k_2(j) < \cdots < k_{i(j)}(j)$ for each j and $a_{k_1(j)}, a_{k_2(j)}, \dots$ belong to a common subalgebra $A_{p(j)}$; see also Example 1.24. Although writing down the general expression is complicated, we will later do so for $I = \{1, 2, 3\}$ in the proof of Proposition 2.1.

Example 1.23. The sequence $(A_i)_{i=1}^N$ of $*$-subalgebras of $\mathbb{B}(H)$ defined in (1.8) is monotonically independent. To see this, let $i_1, i_2, \dots, i_n \in [N]$ and $a_j \in A_{i_j}$, $j \in [n]$. Let $2 \le k \le n-1$ be such that $i_{k-1} < i_k > i_{k+1}$. We can see $a_{k-1} a_k a_{k+1} = \varphi(a_k) a_{k-1} a_{k+1}$ because of $p_i x p_i = \varphi_i(x) p_i$ for any $x \in \mathbb{B}(H_i)$, and hence

$$\varphi(a_1 a_2 \cdots a_n) = \varphi(a_k)\varphi(a_1 a_2 \cdots a_{k-1} a_{k+1} \cdots a_n).$$

The cases $i_1 > i_2$ and $i_{n-1} < i_n$ are handled with a trick: if $i_1 > i_2$ then we may insert the projection $p = p_1 \otimes p_2 \otimes \cdots \otimes p_N$ as

$$\varphi(a_1 a_2 \cdots a_n) = \langle \xi, a_1 a_2 \cdots a_n \xi \rangle = \langle p\xi, a_1 a_2 \cdots a_n \xi \rangle = \langle \xi, p a_1 a_2 \cdots a_n \xi \rangle,$$

where the last equality holds because of $p = p^*$. Then we can use the operator identity $p a_1 a_2 = \varphi(a_1) p a_2$ to show

$$\varphi(a_1 a_2 \cdots a_n) = \varphi(a_1)\varphi(a_2 \cdots a_n).$$

The case $i_{n-1} < i_n$ is similar.

Example 1.24. Suppose that subsets $\{a, a'\}, \{b, b'\}, \{c, c'\}$ are monotonically independent in (A, φ). Then

$$\begin{aligned} &\varphi(ab) = \varphi(ba) = \varphi(a)\varphi(b), \\ &\varphi(aba') = \varphi(aa')\varphi(b), \\ &\varphi(bab') = \varphi(b)\varphi(a)\varphi(b'), \\ &\varphi(abcb'c'a') = \varphi(abcb'a')\varphi(c') = \varphi(abb'a')\varphi(c)\varphi(c') = \varphi(aa')\varphi(bb')\varphi(c)\varphi(c'). \end{aligned}$$

The formula for $\varphi(bab')$ is already different from the case of tensor independence.

Proposition 1.25. *Let (A, φ) be a nc-probability space and I be a totally ordered set. Suppose that a family of subalgebras $(A_i)_{i \in I}$ of A is monotonically independent. Then the restriction of φ to the subalgebra $\langle A_i : i \in I \rangle$ is determined by $\varphi\restriction_{A_i}$, $i \in I$.*

Proof. This is a consequence of Remark 1.22. □

Remark 1.26. In free probability, unital $*$- or C^*-probability spaces (A, φ) often satisfy the conditions that

- φ is faithful, i.e., $\varphi(a^* a) = 0$ implies $a = 0$,
- φ is tracial, i.e., $\varphi(ab) = \varphi(ba)$ for $a, b \in A$.

In monotone probability, however, these conditions hold only in trivial cases. Let a, b be monotonically independent real random variables in a unital C^*-probability space (A, φ).

a) If φ is tracial then either $\mu_a = \delta_0$ or $\mu_b = \delta_{\varphi(a)}$. Suppose that $\mu_a \ne \delta_0$. We first observe that $\varphi((ba^2)b) = \varphi(a^2)\varphi(b)^2$, while $\varphi(b(ba^2)) = \varphi(a^2)\varphi(b^2)$. The traciality of φ therefore implies $\varphi(b^2) = \varphi(b)^2$ as $\varphi(a^2) > 0$. This means that the analytic distribution μ_b has vanishing variance, so that $\mu_b = \delta_{\varphi(b)}$.

b) If φ is faithful then again either $\mu_a = \delta_0$ or $\mu_b = \delta_{\varphi(b)}$. Since a and $b - \varphi(b)\mathbf{1}_A$ are monotonically independent, we may and do assume that $\varphi(b) = 0$ from the beginning. Because $\varphi((ab)^*(ab)) = \varphi(ba^2 b) = \varphi(a^2)\varphi(b)^2 = 0$, the faithfulness of φ implies $ab = 0$. If $\mu_a \ne \delta_0$ then $\varphi(a^2) > 0$ and so $0 = \varphi(a^2 b^n) = \varphi(a^2)\varphi(b^n)$, i.e., $\varphi(b^n) = 0$ for all $n \in \mathbb{N}$. This implies $\mu_b = \delta_0$.

1.5. **Additive monotone convolution.** For classically independent $\mathbb{R}$-valued random variables X, Y, the distribution of $X + Y$ is called the convolution of μ_X and μ_Y and is given by

$$(\mu_X * \mu_Y)(B) := \mu_{X+Y}(B) = \int_{\mathbb{R}^2} \chi_B(s+t)\mu_X(\mathrm{d}s)\mu_Y(\mathrm{d}t), \qquad B \in \mathcal{B}(\mathbb{R}).$$

It is well known that the exponential moment generating function (essentially equivalent to the characteristic function) is useful to calculate the convolution. For simplicity, assuming $X \in L^\infty$, let

$$E_X(z) := \mathbb{E}[e^{zX}] = \sum_{n \ge 0} \frac{\mathbb{E}[X^n]}{n!} z^n, \qquad z \in \mathbb{C}.$$

Due to the independence we have $E_{X+Y}(z) = \mathbb{E}[e^{zX}e^{zY}] = \mathbb{E}[e^{zX}]\mathbb{E}[e^{zY}] = E_X(z)E_Y(z)$.

Here we consider the distribution of $x+y$ when x and y are monotonically independent real random variables in a unital C^*-probability space (A,φ). Instead of the exponential moment generating function, a more useful function is the shifted moment generating function

$$M_x(z) := z\varphi((\mathbf{1}_A - zx)^{-1}) = \sum_{n\ge 0} \varphi(x^n)z^{n+1}, \qquad z\in\mathbb{C},\ |z| < 1/\|x\|,$$

where $1/0$ is to be interpreted as $+\infty$.

Theorem 1.27. *Let (A,φ) be a unital C^*-probability space. Suppose that $x,y\in A$ are monotonically independent. Then for all $z\in\mathbb{C}$ with $|z| < 1/(\|x\|+\|y\|)$ we have $|M_y(z)| < 1/\|x\|$ and*

$$M_{x+y}(z) = M_x(M_y(z)). \tag{1.9}$$

Proof. First we check that the functions $M_{x+y}(z)$ and $M_x(M_y(z))$ make sense. The assumption $|z| < 1/(\|x\|+\|y\|)$ implies $|z| < 1/\|x+y\|$ and $|z| < 1/\|y\|$ so that $M_{x+y}(z)$ and $M_y(z)$ are well defined. Moreover, one can check $|M_y(z)| < 1/\|x\|$ by using the estimate

$$\begin{aligned}|M_y(z)| &\le \sum_{n=0}^{\infty} |\varphi(y^n)||z|^{n+1} \le \sum_{n=0}^{\infty} \|y\|^n|z|^{n+1} = \frac{|z|}{1-\|y\||z|} \\ &< \frac{\frac{1}{\|x\|+\|y\|}}{1-\|y\|\frac{1}{\|x\|+\|y\|}} = \frac{1}{\|x\|}.\end{aligned}$$

Because x,y are noncommuting in general, the expansion of $(x+y)^n$ contains 2^n terms. The following expression is useful for us:

$$(x+y)^n = \sum_{\ell=0}^{n} \sum_{\substack{k_0,k_1,\dots,k_\ell\ge 0,\\ k_0+k_1+\cdots+k_\ell = n-\ell}} y^{k_0}xy^{k_1}x\cdots xy^{k_\ell}.$$

Evaluating this by φ and applying the definition of monotone independence together with Remark 1.16 b) yields

$$\varphi((x+y)^n) = \sum_{\ell=0}^{n} \sum_{\substack{k_0,k_1,\dots,k_\ell\ge 0,\\ k_0+k_1+\cdots+k_\ell = n-\ell}} \varphi(x^\ell)\varphi(y^{k_0})\varphi(y^{k_1})\cdots\varphi(y^{k_\ell}). \tag{1.10}$$

Then we can proceed as follows. First we perform formal calculations and later discuss analytic issues:

$$\begin{aligned}M_{x+y}(z) &= \sum_{n\ge 0} z^{n+1}\sum_{\ell=0}^{n} \sum_{\substack{k_0,k_1,\dots,k_\ell\ge 0,\\ k_0+k_1+\cdots+k_\ell = n-\ell}} \varphi(x^\ell)\varphi(y^{k_0})\varphi(y^{k_1})\cdots\varphi(y^{k_\ell}) &(1.11)\\ &= \sum_{\ell\ge 0}\sum_{n=\ell}^{\infty} \sum_{\substack{k_0,k_1,\dots,k_\ell\ge 0,\\ k_0+k_1+\cdots+k_\ell = n-\ell}} \varphi(x^\ell)\varphi(y^{k_0})z^{k_0+1}\varphi(y^{k_1})z^{k_1+1}\cdots\varphi(y^{k_\ell})z^{k_\ell+1} &(1.12)\\ &= \sum_{\ell\ge 0}\sum_{k_0,k_1,\dots,k_\ell\ge 0} \varphi(x^\ell)\varphi(y^{k_0})z^{k_0+1}\varphi(y^{k_1})z^{k_1+1}\cdots\varphi(y^{k_\ell})z^{k_\ell+1} &(1.13)\\ &= \sum_{\ell\ge 0}\varphi(x^\ell)M_y(z)^{\ell+1} &(1.14)\\ &= M_x(M_y(z)). &(1.15)\end{aligned}$$

Calculations (1.12)–(1.14) can be justified with Fubini's theorem because the sum (1.13) is absolutely convergent:

$$\begin{aligned}&\sum_{\ell\ge 0}\sum_{k_0,k_1,\dots,k_\ell\ge 0} \left|\varphi(x^\ell)\varphi(y^{k_0})z^{k_0+1}\varphi(y^{k_1})z^{k_1+1}\cdots\varphi(y^{k_\ell})z^{k_\ell+1}\right| \\ &\le \sum_{\ell\ge 0}\sum_{k_0,k_1,\dots,k_\ell\ge 0} |z|^{\ell+1}\|x\|^\ell(|z|\|y\|)^{k_0+k_1+\cdots+k_\ell} \\ &= \frac{|z|}{1-|z|\|y\|}\sum_{\ell\ge 0}\left(\frac{|z|\|x\|}{1-|z|\|y\|}\right)^\ell < +\infty.\end{aligned}$$

□

Suppose that x is a real random variable in a unital C^*-probability space. Since the moment sequence $\varphi(x^n)$, $n=1,2,3,\dots$, is encoded in $M_x(z)$ as the Taylor coefficients, the analytic distribution μ_x can be determined from M_x (later we show a more straightforward formula that recovers μ_x from M_x called the Stieltjes inversion, see Proposition 4.33). Conversely, M_x can be computed from μ_x by the formula

$$M_x(z) = \int_{\mathrm{Sp}(x)} \frac{z}{1-zt}\,\mu_x(\mathrm{d}t).$$

Thus, we can identify M_x with μ_x. Therefore, formula (1.9) gives a binary operation on the set of compactly supported probability measures, which is called **additive monotone convolution**. Later we extend additive monotone convolution to arbitrary probability measures on $\mathbb{R}$, see Theorem 5.1.

1.6. **Multiplicative monotone convolution.** In probability theory, multiplication of independent random variables is another natural operation. If X, Y are $\mathbb{R}$-valued random variables defined on a probability space, then the law μ_{XY} is called the multiplicative convolution of μ_X and μ_Y and is given by

$$\mu_{XY}(B) = \int_{\mathbb{R}^2} \chi_B(st)\mu_X(\mathrm{d}s)\mu_Y(\mathrm{d}t), \qquad B \in \mathcal{B}(\mathbb{R}).$$

In the particular case where X, Y are both positive, the use of $\log(XY) = \log X + \log Y$ allows us to reduce the calculation of multiplicative convolution to the additive convolution.

Here we will consider the multiplication of monotonically independent random variables. In the setting of unital C^*-probability space, for real random variables x, y, the product xy is not self-adjoint in general. To recover the self-adjointness, we consider $\sqrt{x}y\sqrt{x}$ or $\sqrt{y}x\sqrt{y}$ assuming x or y are positive. However, the result turns out to be rather trivial.

Proposition 1.28. *Let (A, φ) be a unital C^*-probability space. Let $x, y \in A$ be monotonically independent real random variables. Let $\alpha := \varphi(y)$.*

(i) *If $x \geq 0$ then $\mu_{\sqrt{x}y\sqrt{x}} = \mu_{\alpha x}$.*

(ii) *If $y \geq 0$ and $\alpha > 0$ then $\mu_{\sqrt{y}x\sqrt{y}} = (1-\beta)\delta_0 + \beta\mu_{\alpha x}$, where $\beta := \varphi(\sqrt{y})^2/\alpha$. If $\alpha = 0$ then $\mu_{\sqrt{y}x\sqrt{y}} = \delta_0$.*

Proof. (i) First note that $\mathrm{C}^*\langle x\rangle$ and $\mathrm{C}^*\langle y\rangle$ are monotonically independent, and that $\sqrt{x} \in \mathrm{C}^*\langle x\rangle$. We can therefore obtain

$$\varphi((\sqrt{x}y\sqrt{x})^n) = \varphi(\sqrt{x}yxyx\cdots y\sqrt{x}) = \varphi(\sqrt{x}x^{n-1}\sqrt{x})\varphi(y)^n = \varphi((\alpha x)^n).$$

(ii) If $\alpha = \varphi(y) > 0$ then for $n \geq 1$ we have

$$\varphi((\sqrt{y}x\sqrt{y})^n) = \varphi(\sqrt{y}xyx\cdots x\sqrt{y}) = \varphi(x^n)\varphi(\sqrt{y})^2\varphi(y)^{n-1} = \beta\varphi((\alpha x)^n).$$

Note that $\beta \leq 1$ holds by the Cauchy–Schwarz inequality. Moreover, because the analytic distribution of y is supported on $[0, +\infty)$ and not equal to δ_0, we have $\varphi(\sqrt{y}) = \int_0^\infty \sqrt{t}\,\mu_y(\mathrm{d}t) > 0$ and so $\beta > 0$. The conclusion follows by the fact

$$\int_{\mathbb{R}} t^n((1-\beta)\delta_0 + \beta\mu_{\alpha x})(\mathrm{d}t) = \beta\varphi((\alpha x)^n), \qquad n \geq 1.$$

If $\alpha = \varphi(y) = 0$ then the Cauchy–Schwarz inequality implies $\varphi(\sqrt{y}) = 0$, and so $\varphi((\sqrt{y}x\sqrt{y})^n) = \varphi(x^n)\varphi(\sqrt{y})^2\varphi(y)^{n-1} = 0$ for $n \geq 1$. □

A more nontrivial distribution of $\sqrt{x}y\sqrt{x}$ can be obtained by assuming the monotone independence of $x - \mathbf{1}_A$ and $y - \mathbf{1}_A$; recall from Remark 1.16 that this assumption is different from the monotone independence of x and y. Currently, in the literature, this is taken as the standard definition of multiplicative monotone convolution although the definition might look strange. There are reasons why we assume the independence of $x - \mathbf{1}_A$ and $y - \mathbf{1}_A$; one practical reason is that this is useful in a later application to random matrices, see Theorem 8.13. Another reason is that this multiplicative monotone convolution appears in free probability theory in the form of "subordination functions", see Notes 5.4.

To describe the multiplicative monotone convolution, useful transforms are the following **ψ-transform** (also called the **moment generating function**) and the **η-transform**

$$\psi_x(z) := \frac{1}{z}M_x(z) - 1 = \varphi(zx(\mathbf{1}_A - zx)^{-1}) = \sum_{n\geq 1}\varphi(x^n)z^n,$$

$$\eta_x(z) := \frac{\psi_x(z)}{1+\psi_x(z)},$$

which are holomorphic in a neighborhood of zero.

Theorem 1.29. *Let (A, φ) be a unital C^*-probability space. Let $x, y \in A$ be real random variables such that $x \geq 0$ and that $x - \mathbf{1}_A, y - \mathbf{1}_A$ are monotonically independent. Then for all $z \in \mathbb{C}$ sufficiently close to zero we have*

$$\eta_{xy}(z) = \eta_{yx}(z) = \eta_{\sqrt{x}y\sqrt{x}}(z) = \eta_x(\eta_y(z)).$$

Proof. Let $x_0 := x - \mathbf{1}_A$. Recall from Remark 1.16 d) that $\mathrm{C}^*\langle x_0\rangle$ and $\mathrm{C}^*\langle y\rangle$ are monotonically independent. We first expand $(\sqrt{x}y\sqrt{x})^n$ as

$$\begin{aligned}(\sqrt{x}y\sqrt{x})^n &= \sqrt{x}y(x_0+\mathbf{1}_A)y(x_0+\mathbf{1}_A)\cdots(x_0+\mathbf{1}_A)y\sqrt{x} \\ &= \sum_{k=0}^{n-1}\sum_{\substack{j_1,j_2,\dots,j_{k+1}\geq 1\\ j_1+j_2+\cdots+j_{k+1}=n}} \sqrt{x}y^{j_1}x_0y^{j_2}x_0\cdots x_0y^{j_{k+1}}\sqrt{x},\end{aligned}$$

where k stands for the number of x_0's selected from $(x_0+\mathbf{1}_A)$'s and $j_i - 1$ is the number of consecutive $\mathbf{1}_A$'s between the $(i-1)$th x_0 and ith x_0. Evaluating the above by φ yields

$$\varphi((\sqrt{x}y\sqrt{x})^n) = \sum_{k=0}^{n-1}\sum_{\substack{j_1,j_2,\dots,j_{k+1}\geq 1\\ j_1+j_2+\cdots+j_{k+1}=n}} \varphi(\sqrt{x}x_0^k\sqrt{x})\varphi(y^{j_1})\varphi(y^{j_2})\cdots\varphi(y^{j_{k+1}}) \tag{1.16}$$

$$= \sum_{k=0}^{n-1}\sum_{\substack{j_1,j_2,\dots,j_{k+1}\geq 1\\ j_1+j_2+\cdots+j_{k+1}=n}} \varphi(xx_0^k)\varphi(y^{j_1})\varphi(y^{j_2})\cdots\varphi(y^{j_{k+1}}). \tag{1.17}$$

When obtaining line (1.16), we have applied monotone independence of the form (1.6) thanks to the fact $\sqrt{x} \in \mathrm{C}^*\langle \mathbf{1}_A, x_0\rangle$. Note also that formula (1.17) holds because x_0 and $\sqrt{x}$ commute. Formula (1.17) leads to the following:

$$\begin{aligned}
\psi_{\sqrt{x}y\sqrt{x}}(z) &= \varphi(z\sqrt{x}y\sqrt{x}(\mathbf{1}_A - z\sqrt{x}y\sqrt{x})^{-1}) = \sum_{n=1}^\infty z^n \varphi((\sqrt{x}y\sqrt{x})^n) \\
&= \sum_{n=1}^\infty \sum_{k=0}^{n-1} \sum_{\substack{j_1,j_2,\dots,j_{k+1}\ge 1\\ j_1+j_2+\cdots+j_{k+1}=n}} z^n \varphi(xx_0^k)\varphi(y^{j_1})\varphi(y^{j_2})\cdots\varphi(y^{j_{k+1}}) \\
&= \sum_{k=0}^\infty \sum_{j_1,j_2,\dots,j_{k+1}\ge 1} \varphi(xx_0^k)\varphi((zy)^{j_1})\varphi((zy)^{j_1})\cdots\varphi((zy)^{j_{k+1}}) \qquad (1.18)\\
&= \sum_{k=0}^\infty \varphi(xx_0^k)\psi_y(z)^{k+1} = \varphi\left(x\sum_{k=0}^\infty (\psi_y(z)x_0)^k\right)\psi_y(z) \\
&= \varphi\left(x(\mathbf{1}_A - \psi_y(z)x_0)^{-1}\right)\psi_y(z) = \varphi\left(x\eta_y(z)(\mathbf{1}_A - \eta_y(z)x)^{-1}\right) \\
&= \psi_x(\eta_y(z)).
\end{aligned}$$

The expression in (1.18) is absolutely convergent for sufficiently small $|z|$, so that the above calculations can be justified by Fubini's theorem. The obtained formula $\psi_{\sqrt{x}y\sqrt{x}}(z) = \psi_x(\eta_y(z))$ is equivalent to the desired $\eta_{\sqrt{x}y\sqrt{x}}(z) = \eta_x(\eta_y(z))$ for small $|z|$. A slight modification of the above calculations of $\varphi((\sqrt{x}y\sqrt{x})^n)$ shows $\varphi((\sqrt{x}y\sqrt{x})^n) = \varphi((xy)^n) = \varphi((yx)^n)$. For example,

$$(xy)^n = xy(x_0 + \mathbf{1}_A)y(x_0+\mathbf{1}_A)\cdots(x_0+\mathbf{1}_A)y$$

can be used to show $\varphi((\sqrt{x}y\sqrt{x})^n) = \varphi((xy)^n)$. □

Remark 1.30. a) The attentive reader might have noticed that the assumption $x \ge 0$ is unnecessary to show the formulas $\eta_{xy}(z) = \eta_{yx}(z) = \eta_x(\eta_y(z))$.

b) One could also consider $\sqrt{y}x\sqrt{y}$ by assuming $y \ge 0$, which, however, would result in a more complicated formula; see [68, Section 9] and [69, Theorem 3.18].

Analogously to additive monotone convolution, Theorem 1.29 gives rise to a binary operation on probability measures with compact support (one is required to be supported on $[0,+\infty)$ as it comes from nonnegative elements $x \ge 0$). This operation is called **multiplicative monotone convolution** and it can also be generalized to probability measures with unbounded support, see Theorem 5.5.

1.7. **Notes.** Our definition of $*$-probability space in Definition 1.2 is used e.g. by Muraki [124], Gerhold [73], Gerhold, Hasebe and Ulrich [74] and Lachs [101]. The term "restricted state" is used in [73]. It is called "strongly positive linear functional" in [101] and simply "state" in [74]. Unital C^*-probability spaces and W^*-probability spaces are widely used in free probability [125, 149]. Hora and Obata's book [90] uses the setup of unital $*$-probability spaces and calls them algebraic probability spaces.

Muraki gave an abstract definition of monotone independence in [120] that had been implicit in earlier works on creation and annihilation operators on monotone Fock spaces [58, 108, 118, 119]. The original definition was slightly different from Definition 1.21. A definition equivalent to ours was given e.g. by Franz [66]. The operator model (1.8) for monotone independence is equivalent to the model given by Muraki [120] but the original model appeared more analogous to the operator model for free independence on the free product Hilbert space.

Rank-one perturbations of operators are intensively studied in mathematical physics, see e.g. [65, 142, 143]. Example 1.18 is connected to such works; the reader is referred to [79, Section 9] for further information. In general, higher-rank perturbations are not directly connected to monotone independence, but higher-rank perturbations and unitarily invariant random matrices show monotone independence asymptotically in the large size limit. This will be discussed in Section 8.

The formula for additive monotone convolution in Theorem 1.27 was given by Muraki [120]. Our proof is different and is adopted from [132, Theorem 3.2] and [87, Proposition 4.1]. The formula for multiplicative monotone convolution in Theorem 1.29 was given by Bercovici [31, Theorem 2.2] and Franz [68, Corollary 4.3]. One can also consider the multiplication of unitary elements that is omitted in this article; the interested reader is referred to [31, 68]. Commutators $\mathrm{i}(xy - yx)$ and anticommutators $xy + yx$ of monotonically independent random variables x, y were studied by Banna and Tseng [18].

Attempts are being made to unify or establish connections between different notions of independence. An incomplete list of those related to monotone independence is the following: Arizmendi, Mendoza and Vazquez-Becerra introduced "BMT independence" by naturally generalizing the operator models in Examples 1.19 and 1.23 [14]; Cébron, Dahlqvist, Gabriel and Gilliers found that monotone independence arises naturally from "cyclic-monotone independence" [41] and more generally from "cyclic-conditional freeness" [43]; Cébron, Dahlqvist and Male observed monotone independence in the context of "traffic independence" that captures asymptotic features of permutation-invariant random matrices [42]; Franz observed that "conditional freeness" of Bożejko, Leinert and Speicher contains monotone independence as a special case [68]; Hasebe constructed "conditionally monotone independence" with respect to two states [82] and a further generalization with respect to three states [81]; Jekel and Liu defined "tree independence" building upon the structure of trees [93]; Mingo and Tseng showed a construction of monotone independence within the framework of "infinitesimal freeness" [114]; Skoufranis derived monotone independence from "bi-free independence" of Voiculescu [144]; Wysochanski considered "bm-independence" for subalgebras indexed by partially ordered sets [156, 157, 158].

2. Universal construction of monotone independence

In probability theory, there is a canonical way to construct independent random variables. Let $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$, $i \in I$, be a family of probability spaces. We set $\Omega := \prod_{i\in I} \Omega_i$ be the product set, $\mathcal{F} := \bigotimes_{i\in I} \mathcal{F}_i$ be the product σ-field and $\mathbb{P} := \bigotimes_{i\in I} \mathbb{P}_i$ be the product measure. Given random variables $X_i \colon \Omega_i \to \mathbb{C}$ $(i \in I)$, we define $Y_i \colon \Omega \to \mathbb{C}$ by

$$Y_i(\omega_1, \omega_2, \dots) := X_i(\omega_i).$$

Then $(Y_i)_{i\in I}$ is an independent family of random variables defined in $(\Omega, \mathcal{F}, \mathbb{P})$ and the distribution of Y_i coincides with the distribution of given X_i.

A natural generalization of the above construction can be given for nc-probability spaces using the tensor product of algebras, which yields tensor independence. This is exactly Example 1.19, in which the index set was a finite set $I = \{1, 2, \dots, N\}$.

For monotone independence, a much bigger algebra, called the coproduct or the free product (without identification of units), is useful to define a canonical model of independent subalgebras. An advantage of free product algebra is that it has a universality property that allows us to construct other types of independence on the same algebra just by selecting different linear functionals, see Notes 2.5.

Notions of independence are also known to have universal realizations from the perspective of graph theory. Roughly speaking, the idea is to find a graph product whose adjacency matrix consists of operators of the form (1.7) in the tensor case, or (1.8) in the monotone case.

2.1. **Free product of algebras.** Let $(A_i)_{i\in I}$ be a family of algebras. Let A be the vector space over $\mathbb{C}$ defined by the algebraic direct sum

$$A = \bigoplus_{\substack{n\in\mathbb{N},\, i_1,i_2,\dots,i_n\in I,\\ i_1\neq i_2,\, i_2\neq i_3,\, \dots,\, i_{n-1}\neq i_n}} A_{i_1,i_2,\dots,i_n},$$

where $A_{i_1,i_2,\dots,i_n} := A_{i_1} \otimes A_{i_2} \otimes \cdots \otimes A_{i_n}$. In this vector space, we define a multiplication called the concatenation: for $a_1 \otimes a_2 \otimes \cdots \otimes a_n \in A_{i_1,i_2,\dots,i_n}$ and $b_1 \otimes b_2 \otimes \cdots \otimes b_m \in A_{j_1,j_2,\dots,j_m}$,

$$(a_1 \otimes a_2 \otimes \cdots \otimes a_n)(b_1 \otimes b_2 \otimes \cdots \otimes b_m) := \begin{cases} a_1 \otimes a_2 \otimes \cdots \otimes a_n \otimes b_1 \otimes b_2 \otimes \cdots \otimes b_m, & \text{if } i_n \neq j_1, \\ a_1 \otimes a_2 \otimes \cdots \otimes a_{n-1} \otimes (a_n b_1) \otimes b_2 \otimes \cdots \otimes b_m, & \text{if } i_n = j_1 \end{cases}$$

and then extend this definition to A by bilinearity. With this multiplication, simple tensors can be interpreted just as the multiplication of letters, so that, for example, we may simply write $a_1 a_2 \cdots a_n$ for $a_1 \otimes a_2 \otimes \cdots \otimes a_n$. This multiplication is associative and A becomes an algebra, which is denoted by $\bigsqcup_{i\in I} A_i$ and is called the **free product** or coproduct (see below). The algebra A contains each A_i as a direct summand, so that we can naturally interpret each A_i as a subalgebra of A.

The free product has a universality. Consider a family of algebras $(A_i)_{i\in I}$. An algebra A together with a family of homomorphisms $f_i \colon A_i \to A$, $i \in I$, is called a coproduct of $(A_i)_{i\in I}$ if for any family of homomorphisms g_i from A_i into an algebra B, $i \in I$, there exists a unique homomorphism $h \colon A \to B$ such that $h \circ f_i = g_i$, $i \in I$. A coproduct is unique up to isomorphisms. In fact, the free product $\bigsqcup_{i\in I} A_i$ together with the natural embeddings $\iota_i \colon A_i \to \bigsqcup_{i\in I} A_i$, satisfies the universality and hence is a coproduct.

2.2. **Monotone product of nc-probability spaces.** Given a family of nc-probability spaces $(A_i, \varphi_i)_{i\in I}$, where I is a totally ordered set, we set $A := \bigsqcup_{i\in I} A_i$. We aim to define a linear functional φ on A such that the subalgebras $(A_i)_{i\in I}$ are monotonically independent in (A, φ). We start from the case $I = \{1, 2\}$. The free product is then simpler:

$$A_1 \sqcup A_2 = \bigoplus_{n=1}^{\infty} \Big[\underbrace{(A_1 \otimes A_2 \otimes A_1 \otimes \cdots)}_{\text{length } n} \oplus \underbrace{(A_2 \otimes A_1 \otimes A_2 \otimes \cdots)}_{\text{length } n}\Big].$$

An advantage of the free product is that we can simply define φ to be the right-hand side of (1.5), i.e., for any $i_1, i_2, \dots, i_n \in \{1, 2\}$ with $i_1 \neq i_2$, $i_2 \neq i_3$, $\dots$, $i_{n-1} \neq i_n$ and any $a_1 a_2 \cdots a_n \in A_{i_1,i_2,\dots,i_n}$, we set

$$\varphi(a_1 a_2 \cdots a_n) := \varphi_1\left(\overrightarrow{\prod_{k\colon a_k\in A_1}} a_k\right) \prod_{k\colon a_k\in A_2} \varphi_2(a_k). \tag{2.1}$$

Since the right-hand side of (2.1) is a multilinear functional on $A_{i_1} \times A_{i_2} \times \cdots \times A_{i_n}$, it makes sense as a definition by the universality of tensor product of vector spaces. We denote the above construction as

$$(A, \varphi) = (A_1, \varphi_1) \triangleright (A_2, \varphi_2) = (A_1 \sqcup A_2, \varphi_1 \triangleright \varphi_2)$$

and call it the **monotone product** of (A_1, φ_1) and (A_2, φ_2).

The monotone product has certain associativity. For three nc-probability spaces (A_i, φ_i), $i = 1, 2, 3$, there is a natural isomorphism, called the associator,

$$\Psi \colon (A_1 \sqcup A_2) \sqcup A_3 \xrightarrow{\sim} A_1 \sqcup (A_2 \sqcup A_3).$$

This is defined by the natural rearrangement of the tensor components so that the resulting element belongs to the target space. For example, if $a_1 a_2 a_3 a_4 a_5 \in A_{2,3,2,1,3}$, then $a_1 \otimes a_2 \otimes (a_3 a_4) \otimes a_5$ is an element of $(A_1 \sqcup A_2) \sqcup A_3$, where $a_3 a_4$ stands for the multiplication in $A_1 \sqcup A_2$, while $\otimes$ is the multiplication in $(A_1 \sqcup A_2) \sqcup A_3$. Then

$$\Psi(a_1 \otimes a_2 \otimes (a_3 a_4) \otimes a_5) := (a_1 a_2 a_3) \otimes a_4 \otimes a_5 \in A_1 \sqcup (A_2 \sqcup A_3),$$

where $a_1 a_2 a_3$ in the right side is the multiplication in $A_2 \sqcup A_3$ and $\otimes$ is the multiplication in $A_1 \sqcup (A_2 \sqcup A_3)$. Omitting parentheses and Ψ, we simply write $a_1 a_2 \cdots a_n \in (A_1 \sqcup A_2) \sqcup A_3$ or $a_1 a_2 \cdots a_n \in A_1 \sqcup (A_2 \sqcup A_3)$, which does not cause any confusion because the appropriate arrangement of parentheses is uniquely determined.

Proposition 2.1. *For three nc-probability spaces (A_i, φ_i), $i = 1, 2, 3$, we have*

$$(\varphi_1 \rhd \varphi_2) \rhd \varphi_3 = (\varphi_1 \rhd (\varphi_2 \rhd \varphi_3)) \circ \Psi.$$

Proof. Let $i_1, i_2, \dots, i_n \in [3]$ with $i_1 \neq i_2$, $i_2 \neq i_3$, ..., $i_{n-1} \neq i_n$ and $a_1 a_2 \cdots a_n \in A_{i_1, i_2, \dots, i_n}$. For a subset $J \subseteq [3]$, let $S_J \subseteq [n]$ be defined by $S_J := \{p : i_p \in J\}$. Furthermore, we decompose $S_{2,3}$ into maximal intervals $T_1, T_2, \dots, T_r$ of $[n]$. For example, if $(i_1, i_2, i_3, i_4, i_5, i_6, i_7) = (1, 3, 2, 3, 1, 3, 2)$, then $S_1 = \{1, 5\}$ and $S_{2,3} = \{2, 3, 4, 6, 7\}$, and $S_{2,3}$ is decomposed into $T_1 = \{2, 3, 4\}$ and $T_2 = \{6, 7\}$. By the definition of $\rhd$ we have

$$\begin{aligned}
&(\varphi_1 \rhd (\varphi_2 \rhd \varphi_3))(a_1 a_2 \cdots a_n) \\
&= \varphi_1 \left(\overrightarrow{\prod_{p \in S_1}} a_p \right) (\varphi_2 \rhd \varphi_3) \left(\overrightarrow{\prod_{p \in T_1}} a_p \right) \cdots (\varphi_2 \rhd \varphi_3) \left(\overrightarrow{\prod_{p \in T_r}} a_p \right) \\
&= \varphi_1 \left(\overrightarrow{\prod_{p \in S_1}} a_p \right) \varphi_2 \left(\overrightarrow{\prod_{p \in T_1,\, i_p = 2}} a_p \right) \cdots \varphi_2 \left(\overrightarrow{\prod_{p \in T_r,\, i_p = 2}} a_p \right) \prod_{p \in S_3} \varphi_3(a_p),
\end{aligned}$$

where $\varphi_2 \left(\overrightarrow{\prod}_{p \in T_j,\, i_p = 2} a_p \right)$ is set to be 1 if the product range for p is empty. On the other hand, we have

$$((\varphi_1 \rhd \varphi_2) \rhd \varphi_3)(a_1 a_2 \cdots a_n) = (\varphi_1 \rhd \varphi_2) \left(\overrightarrow{\prod_{p \in S_{1,2}}} a_p \right) \prod_{p \in S_3} \varphi_3(a_p).$$

To compute the factor $(\varphi_1 \rhd \varphi_2) \left(\overrightarrow{\prod}_{p \in S_{1,2}} a_p \right)$, we decompose $S_{1,2}$ into S_1 and $S' := \{p \in S_{1,2} : i_p = 2\}$. Further, we decompose S' into maximal intervals of S' (not of $[n]$), which are exactly $T_1 \cap S_2, T_2 \cap S_2, \dots, T_r \cap S_2$, so we are done. □

With associativity in hand, we generalize the definition of the monotone product to an arbitrary totally ordered finite set I. We may assume that $I = [N]$. Let $(A_i, \varphi_i)_{i \in I}$ be a family of nc-probability spaces. We can identify

$$\bigsqcup_{i=1}^{N} A_i \simeq (\cdots (((A_1 \sqcup A_2) \sqcup A_3) \sqcup \cdots) \sqcup A_N, \tag{2.2}$$

where the isomorphism is defined similarly to Ψ; it is just a suitable rearrangement of letters of words. On the right-hand side of (2.2) we can define the linear function

$$(\cdots (((\varphi_1 \rhd \varphi_2) \rhd \varphi_3) \rhd \cdots) \rhd \varphi_N, \tag{2.3}$$

which induces a linear functional φ on $\bigsqcup_{i=1}^{N} A_i$ via the isomorphism. The associativity guarantees that the definition of φ does not change if we select another way of adding parentheses in (2.2). This definition of φ means that, when we compute $\varphi(a_1 a_2 \cdots a_n)$ for $a_1 a_2 \cdots a_n \in A_{i_1, i_2, \dots, i_n}$, $i_1, i_2, \dots, i_n \in [N]$ with $i_1 \neq i_2$, $i_2 \neq i_3$, ..., $i_{n-1} \neq i_n$, we first factor out

$$\prod_{p:\, i_p = N} \varphi_N(a_p).$$

Then we apply the same procedure to the rest $\varphi \left(\overrightarrow{\prod}_{p:\, i_p \in [N-1]} a_p \right)$, i.e., factor out the expectations of a_p's in A_{N-1}, regarding consecutive a_p's in A_{N-1} as single elements. We repeat this procedure to the remaining algebras $A_{N-2}, A_{N-3}, \dots, A_2$, and then the factor $\varphi_1 \left(\overrightarrow{\prod}_{p:\, i_p = 1} a_p \right)$ appears in the end.

Finally, we extend the definition of the monotone product to a possibly infinite totally ordered set I. For this purpose, it suffices to define φ on each direct summand $A_{i_1, i_2, \dots, i_n}$; then the definition can be extended by linearity to $A := \bigsqcup_{i \in I} A_i$. This can be done since $A_{i_1, i_2, \dots, i_n}$ can be regarded as a subspace of the free product $\bigsqcup_{j \in J} A_j$, where $J := \{i_1, i_2, \dots, i_n\} \subseteq I$ is a finite totally ordered set. We denote this construction as

$$(A, \varphi) = \mathop{\rhd}_{i \in I} (A_i, \varphi_i) = \left(\bigsqcup_{i \in I} A_i, \mathop{\rhd}_{i \in I} \varphi_i \right)$$

and call it the **monotone product** of $(A_i, \varphi_i)_{i \in I}$. We also call φ the monotone product of $(\varphi_i)_{i \in I}$.

The associativity of the monotone product can be stated in a more general way as follows. First, for a family of algebras $(A_i)_{i \in I}$ and a disjoint decomposition $I = J \cup K$, we denote the natural isomorphism as

$$\Phi_{J,K} \colon \bigsqcup_{i \in I} A_i \xrightarrow{\sim} \left(\bigsqcup_{j \in J} A_j \right) \sqcup \left(\bigsqcup_{k \in K} A_k \right).$$

The definition of $\Phi_{J,K}$ is similar to Ψ and is omitted.

Proposition 2.2. *Suppose that a totally ordered set I decomposes as $I = J \cup K$, where J, K are nonempty disjoint subsets of I such that $j < k$ for all $j \in J$ and $k \in K$. For any family of nc-probability spaces $(A_i, \varphi_i)_{i \in I}$ we have*

$$\mathop{\rhd}_{i \in I} \varphi_i = \left[\left(\mathop{\rhd}_{j \in J} \varphi_j \right) \rhd \left(\mathop{\rhd}_{k \in K} \varphi_k \right) \right] \circ \Phi_{J,K}.$$

Proof. This is a direct consequence of Proposition 2.1. More precisely, it suffices to consider the case of finite totally ordered set I because the definition of the infinite case is based on the finite case. Then the desired identity is just a rearrangement of parentheses, which can be justified by iterative use of Proposition 2.1. □

In our definition (2.3) of the monotone product, we first factored out $\varphi_{i_p}(a_p)$ for all p for which i_p has the largest value among $i_1, i_2, \dots, i_n$. Actually, we can factor out $\varphi(a_p)$'s when i_p is just a local maximum.

Proposition 2.3. *Let $(A_i, \varphi_i)_{i\in I}$ be a family of nc-probability spaces where I is a totally ordered set. Let (A, φ) be the monotone product of $(A_i, \varphi_i)_{i\in I}$. For any $i_1, i_2, \dots, i_n \in I$ and $(a_1, a_2, \dots, a_n) \in A_{i_1} \times A_{i_2} \times \cdots \times A_{i_n}$, we have*

$$\varphi(a_1a_2\cdots a_n) = \begin{cases} \varphi_{i_\ell}(a_\ell)\varphi(a_1\cdots a_{\ell-1}a_{\ell+1}\cdots a_n) & \text{if } 2 \le \ell \le n-1 \text{ and } i_{\ell-1} < i_\ell > i_{\ell+1}, \\ \varphi_{i_1}(a_1)\varphi(a_2a_3\cdots a_n) & \text{if } i_1 > i_2, \\ \varphi_{i_n}(a_n)\varphi(a_1a_2\cdots a_{n-1}) & \text{if } i_{n-1} < i_n. \end{cases}$$

Proof. We fix $1 \le \ell \le n$ such that $i_{\ell-1} < i_\ell > i_{\ell+1}$ (when $\ell = 1$ or n only one of the inequalities is considered) and set $m := i_\ell$. We decompose I into subsets J and K, where $J := \{i \in I : i < m\}$ and $K := \{i \in I : i \ge m\}$. There is a natural isomorphism

$$A \simeq \left(\bigsqcup_{j\in J} A_j\right) \sqcup \left(\bigsqcup_{k\in K} A_k\right).$$

The linear functional φ on A induces a linear functional on the right-hand side, which is exactly

$$\left(\mathop{\triangleright}_{j\in J} \varphi_j\right) \sqcup \left(\mathop{\triangleright}_{k\in K} \varphi_k\right)$$

by the associativity of the monotone product. This means that

$$\varphi(a_1a_2\cdots a_n) = \varphi\left(\overrightarrow{\prod_{p:\, i_p\in J}} a_p\right)\varphi\left(\overrightarrow{\prod_{p\in T_1}} a_p\right)\varphi\left(\overrightarrow{\prod_{p\in T_2}} a_p\right)\cdots\varphi\left(\overrightarrow{\prod_{p\in T_r}} a_p\right),$$

where T_j are the maximal intervals of $[n]$ such that $\min T_j > 1$ implies $i_{\min T_j - 1} \in J$ and $\max T_j < n$ implies $i_{\max T_j + 1} \in J$. By the assumption $i_{\ell-1} < i_\ell > i_{\ell+1}$, some T_j is the singleton $\{\ell\}$. Factoring out this $\varphi(a_\ell)$ does not affect the other factorizations, so we get the conclusion. ☐

Comparing Proposition 2.3 and Definition 1.21, together with the fact that φ and φ_i coincide on A_i, yields the following.

Corollary 2.4. *Let $(A_i, \varphi_i)_{i\in I}$ be a family of nc-probability spaces with I a totally ordered set. Let (A, φ) be the monotone product of $(A_i, \varphi_i)_{i\in I}$. Then the family of subalgebras $(A_i)_{i\in I}$ is monotonically independent in (A, φ).*

The associativity of monotone product will be later used in the following form.

Corollary 2.5. *Let I be a totally ordered set and J, K be its nonempty disjoint subsets such that $I = J \cup K$ and $j < k$ whenever $j \in J$ and $k \in K$. Let (A, φ) be a nc-probability space and $(A_i)_{i\in I}$ be a family of monotonically independent subalgebras of A. Then the two subsets $\bigcup_{j\in J} A_j$ and $\bigcup_{k\in K} A_k$ are monotonically independent.*

Proof. We want to show that $B_1 := \langle A_j : j \in J\rangle$ and $B_2 := \langle A_k : k \in K\rangle$ are monotonically independent. For this, we refer to a universal space. Let $(\hat{A}, \hat{\varphi})$ be the monotone product of $(A_i, \varphi\restriction_{A_i})_{i\in I}$. By Proposition 2.2, the subalgebras $\bigsqcup_{j\in J} A_J$ and $\bigsqcup_{k\in K} A_i$ are monotonically independent in $(\hat{A}, \hat{\varphi})$. With a slight abuse of notation, we have $\varphi\restriction_{A_i} = \hat{\varphi}\restriction_{A_i}$ for all $i \in I$. Since $(A_i)_{i\in I}$ is monotonically independent in both (A, φ) and $(\hat{A}, \hat{\varphi})$, by Proposition 1.25, we have $\varphi(b_1b_2\cdots b_n) = \hat{\varphi}(b_1b_2\cdots b_n)$ for $b_p \in B_1 \cup B_2$. Therefore,

$$\begin{aligned}\varphi(b_1b_2\cdots b_n) &= \hat{\varphi}(b_1b_2\cdots b_n) \\ &= \hat{\varphi}\left(\overrightarrow{\prod_{p:\, b_p\in B_1}} b_p\right)\prod_{p:\, b_p\in B_2}\hat{\varphi}(b_p) = \varphi\left(\overrightarrow{\prod_{p:\, b_p\in B_1}} b_p\right)\prod_{p:\, b_p\in B_2}\varphi(b_p),\end{aligned}$$

showing that B_1, B_2 are monotonically independent. ☐

2.3. **Monotone product of $*$-probability spaces.** If $(A_i)_{i\in I}$ is a family of $*$-algebras, the free product $\bigsqcup_{i\in I} A_i$ also becomes a $*$-algebra with involution defined by

$$(a_1a_2\cdots a_n)^* := a_n^* a_{n-1}^* \cdots a_1^*$$

and extended by antilinearity to the whole algebra. The following proposition shows that the monotone product preserves restricted states. Only in this section we say that a linear operator a on a pre-Hilbert space H is adjointable if there is a linear operator a^* on H such that $\langle a\xi, \eta\rangle = \langle \xi, a^*\eta\rangle$ for all $\xi, \eta \in H$. We introduce the notation

$$\mathbb{L}(H) := \{a \colon H \to H \mid \text{linear and adjointable}\},$$

which forms a unital $*$-algebra.

Proposition 2.6. *Let $(A_i, \varphi_i)_{i\in I}$ be a family of $*$-probability spaces where I is a totally ordered set. Then $\triangleright_{i\in I}(A_i, \varphi_i)$ is also a $*$-probability space.*

Proof. Let $(A, \varphi) := \triangleright_{i\in I}(A_i, \varphi_i)$ and $\tilde{\varphi}$ be the unital extension of $\triangleright_{i\in I}\varphi_i$ to $\tilde{A} := \mathbb{C} \oplus A$. What has to be shown is the positivity $\tilde{\varphi}(a^*a) \ge 0$ for each $a \in \tilde{A}$. As a is a (finite) linear combination of elements of $\mathbb{C}$ and $A_{i_1,i_2,\dots,i_n}$, only finitely many A_i's are involved. Therefore, we may and do assume below that I is the finite set $[N]$.

Let $\tilde{A}_i$ be the unitization of A_i and $\tilde{\varphi}_i \colon \tilde{A}_i \to \mathbb{C}$ be the unital extension of φ_i, which is positive. We take a triplet (π_i, H_i, ξ_i) consisting of a $*$-representation $\pi_i \colon A_i \to \mathbb{L}(H_i)$, a pre-Hilbert space H_i, and a unit vector $\xi_i \in H_i$ such that $\varphi_i(a) = \langle \xi_i, \pi_i(a)\xi_i\rangle$ for all $a \in A_i$. Note that such a triplet exists by restricting the (algebraic) GNS-construction for $\tilde{A}_i$ onto A_i, see [90, Theorem

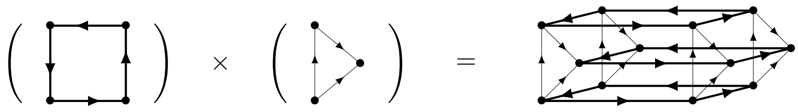

FIGURE 2. Cartesian product.

1.19] for further details. Let $p_i\colon H_i \to \mathbb{C}\xi_i$ be the rank-one projection, $H := H_1 \otimes H_2 \otimes \cdots \otimes H_N$, $\xi := \xi_1 \otimes \xi_2 \otimes \cdots \otimes \xi_N$ as in Example 1.23, and $\lambda_i\colon \mathbb{L}(H_i) \to \mathbb{L}(H)$ be defined by

$$\lambda_i(x_i) := \mathrm{id}_{H_1} \otimes \cdots \otimes \mathrm{id}_{H_{i-1}} \otimes x_i \otimes p_{i+1} \otimes \cdots \otimes p_N,$$

which is a $*$-homomorphism. Then we define a $*$-representation $\tilde{\pi}\colon \tilde{A} \to \mathbb{L}(H)$ by

$$\tilde{\pi}(\mathbf{1}_{\tilde{A}}) := \mathrm{id}_H,$$
$$\tilde{\pi}(a_1a_2\cdots a_n) := \lambda_{i_1}(\pi_{i_1}(a_1))\lambda_{i_2}(\pi_{i_2}(a_2))\cdots\lambda_{i_n}(\pi_{i_n}(a)), \qquad a_1a_2\cdots a_n \in A_{i_1,i_2,\dots,i_n}.$$

We show the formula

$$\tilde{\varphi}(a) = \langle \xi, \tilde{\pi}(a)\xi\rangle_H, \qquad a \in \tilde{A}. \tag{2.4}$$

This is easy for $a = \mathbf{1}_{\tilde{A}}$ and for $a \in A_i$. For $a = a_1a_2\cdots a_n$, when computing $\varphi(a_1a_2\cdots a_n)$ we can use the monotone independence of $(A_i)_{i\in I}$ with respect to φ. On the other hand, when computing $\langle \xi, \pi(a_1a_2\cdots a_n)\xi\rangle = \langle \xi, b_1b_2\cdots b_n\xi\rangle$ where $b_k := \lambda_{i_k}(\pi_{i_k}(a_k))$, we can also use monotone independence of $B_i := \lambda_i(\mathbb{L}(H_i)) \subseteq \mathbb{L}(H)$ shown in Example 1.23. This fact and Proposition 1.25 yield (2.4) on A. Finally, formula (2.4) implies the positivity of $\tilde{\varphi}$ because

$$\tilde{\varphi}(a^*a) = \langle \xi, \tilde{\pi}(a^*a)\xi\rangle = \langle \xi, \tilde{\pi}(a)^*\tilde{\pi}(a)\xi\rangle = \langle \tilde{\pi}(a)\xi, \tilde{\pi}(a)\xi\rangle \geq 0, \qquad a \in \tilde{A}.$$ □

2.4. **Graph products and independence.** A universal construction of tensor and monotone independence can be given from the perspective of graphs. A **(directed) graph** is a pair $\mathrm{G} = (V, E)$, where V is a set and $E \subseteq V \times V$. Each element $v \in V$ is called a **vertex** and $(u, v) \in E$ is called an **edge** (from u to v). In particular, an edge $(v, v) \in E$ is called a **loop** at $v \in V$. Two graphs $\mathrm{G}_1 = (V_1, E_1)$ and $\mathrm{G}_2 = (V_2, E_2)$ are called **isomorphic** if there is a bijection $\Phi\colon V_1 \to V_2$ such that for all $u, v \in V_1$, one has $(u, v) \in E_1$ if and only if $(\Phi(u), \Phi(v)) \in E_2$. The map Φ is called a **graph isomorphism**.

For simplicity, let us assume that $\mathrm{G} = (V, E)$ is a finite graph, i.e., V is a finite set. Let $\ell^2(V)$ be the finite-dimensional Hilbert space of the functions $f\colon V \to \mathbb{C}$ equipped with inner product $\langle f, g\rangle := \sum_{v\in V} \overline{f(v)}g(v)$. The **adjacency matrix** of G is a linear operator $\boldsymbol{A}_{\mathrm{G}}$ on $\ell^2(V)$ defined by

$$(\boldsymbol{A}_{\mathrm{G}}f)(v) := \sum_{u\in V,\, (u,v)\in E} f(u).$$

If we assume $V = \{1, 2, \dots, N\}$ and identify f with the row vector $(f(1), f(2), \dots, f(N)) \in \mathbb{C}^N$, then $\boldsymbol{A}_{\mathrm{G}}$ can be identified with the $N \times N$ matrix whose (i, j) entry is 1 if $(i, j) \in E$ and zero otherwise, acting on the row vector f from the right.* The adjacency matrix contains all information about G.

2.4.1. *Cartesian product and tensor independence.*

Definition 2.7. Let $\mathrm{G}_i = (V_i, E_i)$, $i = 1, 2$, be two graphs. The **Cartesian product** of G_1 and G_2, denoted as $\mathrm{G}_1 \times \mathrm{G}_2$, is the graph with vertex set $V := V_1 \times V_2$ and edge set $E \subseteq V \times V$ defined as follows: for $(u_1, u_2), (v_1, v_2) \in V$,

$$((u_1, u_2), (v_1, v_2)) \in E \iff \begin{cases} (u_1, v_1) \in E_1 \text{ and } u_2 = v_2, \text{ or} \\ u_1 = v_1 \text{ and } (u_2, v_2) \in E_2. \end{cases}$$

See Figure 2 for an example.

Proposition 2.8. *Let $\mathrm{G}_i = (V_i, E_i)$, $i = 1, 2$, be two finite graphs. Under the natural identification $\ell^2(V_1 \times V_2) = \ell^2(V_1) \otimes \ell^2(V_2)$, we have*

$$\boldsymbol{A}_{\mathrm{G}_1\times\mathrm{G}_2} = \boldsymbol{A}_{\mathrm{G}_1} \otimes \mathrm{id}_{\ell^2(V_2)} + \mathrm{id}_{\ell^2(V_1)} \otimes \boldsymbol{A}_{\mathrm{G}_2}.$$

Proof. The proof is analogous to Proposition 2.11 below and is omitted. □

From the above formula and Example 1.14, given any linear functionals φ_i on $\mathbb{B}(\ell^2(V_i))$, $i = 1, 2$, the adjacency matrix $\boldsymbol{A}_{\mathrm{G}_1\times\mathrm{G}_2}$ is the sum of tensor independent matrices with respect to the state $\varphi_1 \otimes \varphi_2$.

The Cartesian product is associative in the following sense.

Proposition 2.9. *Let $\mathrm{G}_i = (V_i, E_i)$, $i = 1, 2, 3$, be graphs. The natural bijection $V_1 \times (V_2 \times V_3) \simeq (V_1 \times V_2) \times V_3$ induces the isomorphism*

$$\mathrm{G}_1 \times (\mathrm{G}_2 \times \mathrm{G}_3) \simeq (\mathrm{G}_1 \times \mathrm{G}_2) \times \mathrm{G}_3.$$

Proof. The proof is straightforward and is omitted. □

*Multiplying a row vector by a matrix from the right is a common convention in some fields, e.g. in the theory of Markov chains.

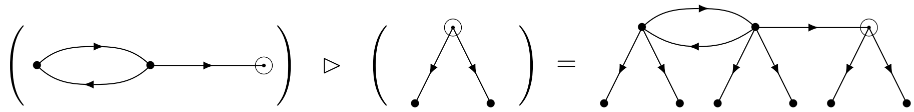

FIGURE 3. Comb product. The roots of the graphs are denoted as $\odot$.

2.4.2. *Comb product and monotone independence.* Monotone independence can be realized by another graph product for rooted graphs. A **rooted graph** is a pair $\mathrm{R} = (\mathrm{G}, o)$ of a graph G and a vertex o of G. The specified vertex o is called the **root** of R. We denote by $\boldsymbol{A}_{\mathrm{R}}$ the adjacency matrix $\boldsymbol{A}_{\mathrm{G}}$ of the underlying graph. Two rooted graphs $\mathrm{R}_1 = (\mathrm{G}_1, o_1)$ and $\mathrm{R}_2 = (\mathrm{G}_2, o_2)$ are called **isomorphic** if there is a graph isomorphism $\Phi\colon \mathrm{G}_1 \to \mathrm{G}_2$ such that $\Phi(o_1) = o_2$.

Definition 2.10. Let $\mathrm{R}_i = (V_i, E_i, o_i)$, $i = 1, 2$, be two rooted graphs. The **comb product**† of R_1 and R_2, denoted as $\mathrm{R}_1 \rhd \mathrm{R}_2$, is the rooted graph with vertex set $V := V_1 \times V_2$, root $o := (o_1, o_2)$, and edge set $E \subseteq V \times V$ defined as follows: for $(u_1, u_2), (v_1, v_2) \in V$,

$$((u_1, u_2), (v_1, v_2)) \in E \iff \begin{cases} (u_1, v_1) \in E_1 \text{ and } u_2 = v_2 = o_2, \text{ or} \\ u_1 = v_1 \text{ and } (u_2, v_2) \in E_2. \end{cases}$$

The comb product is the graph made by gluing copies of R_2 to each vertex of R_1 at o_2, see Figure 3.

For a rooted graph $\mathrm{R} = (V, E, o)$, let $\delta_o\colon V \to \mathbb{C}$ be the function defined by

$$\delta_o(v) = \begin{cases} 1, & v = o, \\ 0, & v \neq o. \end{cases}$$

We set the vector state $\varphi_o\colon \mathbb{B}(\ell^2(V)) \to \mathbb{C}$ by $\varphi_o(a) := \langle \delta_o, a\delta_o \rangle$. The nth moment $\varphi_o(\boldsymbol{A}_{\mathrm{R}}^n)$ is exactly the number of paths of length n on R started from o and terminated at o.

Proposition 2.11. *Let $\mathrm{R}_i = (V_i, E_i, o_i)$, $i = 1, 2$, be finite rooted graphs. Under the natural identification $\ell^2(V_1 \times V_2) = \ell^2(V_1) \otimes \ell^2(V_2)$, we have*

$$\boldsymbol{A}_{\mathrm{R}_1 \rhd \mathrm{R}_2} = \boldsymbol{A}_{\mathrm{R}_1} \otimes p_2 + \mathrm{id}_{\ell^2(V_1)} \otimes \boldsymbol{A}_{\mathrm{R}_2},$$

where p_2 is the orthogonal projection from $\ell^2(V_2)$ onto $\mathbb{C}\delta_{o_2}$.

Proof. It suffices to check the formula on simple tensors, i.e., functions of separated variables $f(v_1, v_2) = g(v_1)h(v_2)$:

$$\begin{aligned}
(\boldsymbol{A}_{\mathrm{R}_1 \rhd \mathrm{R}_2} f)(v_1, v_2) &= \sum_{\substack{(u_1,u_2) \in V_1 \times V_2 \\ ((u_1,u_2),(v_1,v_2)) \in E}} g(u_1)h(u_2) \\
&= \sum_{\substack{(u_1,u_2) \in V_1 \times V_2 \\ (u_1,v_1) \in E_1,\, u_2 = v_2 = o_2}} g(u_1)h(u_2) + \sum_{\substack{(u_1,u_2) \in V_1 \times V_2 \\ u_1 = v_1,\, (u_2,v_2) \in E_2}} g(u_1)h(u_2) \\
&= \left(\sum_{\substack{u_1 \in V_1 \\ (u_1,v_1) \in E_1}} g(u_1) \right) h(o_2)\delta_{o_2}(v_2) + g(v_1) \left(\sum_{\substack{u_2 \in V_2 \\ (u_2,v_2) \in E_2}} h(u_2) \right) \\
&= (\boldsymbol{A}_{\mathrm{R}_1} g)(v_1)(p_2 h)(v_2) + g(v_1)(\boldsymbol{A}_{\mathrm{R}_2} h)(v_2) \\
&= [(\boldsymbol{A}_{\mathrm{R}_1} \otimes p_2 + \mathrm{id}_{\ell^2(V_1)} \otimes \boldsymbol{A}_{\mathrm{R}_2}) f](v_1, v_2).
\end{aligned}$$

$\square$

Observe from Example 1.17 and Proposition 2.11 that $\boldsymbol{A}_{\mathrm{R}_1 \rhd \mathrm{R}_2}$ is the sum of monotonically independent random variables with respect to the state $\varphi_{(o_1,o_2)}(\cdot) = \langle \delta_{(o_1,o_2)}, \cdot\, \delta_{(o_1,o_2)} \rangle$.

The comb product satisfies the following associativity.

Proposition 2.12. *Let $\mathrm{R}_i = (V_i, E_i, o_i)$, $i = 1, 2, 3$, be rooted graphs. The natural bijection $V_1 \times (V_2 \times V_3) \simeq (V_1 \times V_2) \times V_3$ gives the isomorphism*

$$\mathrm{R}_1 \rhd (\mathrm{R}_2 \rhd \mathrm{R}_3) \simeq (\mathrm{R}_1 \rhd \mathrm{R}_2) \rhd \mathrm{R}_3.$$

Proof. The proof is straightforward from the definition of the comb product. Here we give an alternative proof for finite rooted graphs based on the adjacency matrices. Since the roots are clearly preserved by the bijection, it remains to show that the bijection induces a bijection of the edge sets. For this purpose, it suffices to show that the two adjacency matrices coincide on $\ell^2(V_1) \otimes \ell^2(V_2) \otimes \ell^2(V_3)$ under the natural identification of Hilbert spaces. On one hand we have

$$\begin{aligned}
\boldsymbol{A}_{\mathrm{R}_1 \rhd (\mathrm{R}_2 \rhd \mathrm{R}_3)} &= \boldsymbol{A}_{\mathrm{R}_1} \otimes p_{2,3} + \mathrm{id}_{\ell^2(V_1)} \otimes \boldsymbol{A}_{\mathrm{R}_2 \rhd \mathrm{R}_3} \\
&= \boldsymbol{A}_{\mathrm{R}_1} \otimes (p_2 \otimes p_3) + \mathrm{id}_{\ell^2(V_1)} \otimes (\boldsymbol{A}_{\mathrm{R}_2} \otimes p_3 + \mathrm{id}_{\ell^2(V_2)} \otimes \boldsymbol{A}_{\mathrm{R}_3}) \\
&\simeq \boldsymbol{A}_{\mathrm{R}_1} \otimes p_2 \otimes p_3 + \mathrm{id}_{\ell^2(V_1)} \otimes \boldsymbol{A}_{\mathrm{R}_2} \otimes p_3 + \mathrm{id}_{\ell^2(V_1)} \otimes \mathrm{id}_{\ell^2(V_2)} \otimes \boldsymbol{A}_{\mathrm{R}_3},
\end{aligned} \tag{2.5}$$

where $p_{2,3}$ is the orthogonal projection from $\ell^2(V_2 \times V_3)$ onto $\mathbb{C}\delta_{(o_2,o_3)}$, which is $p_2 \otimes p_3$ under the identification $\ell^2(V_2 \times V_3) = \ell^2(V_2) \otimes \ell^2(V_3)$. A similar calculation shows that $\boldsymbol{A}_{(\mathrm{R}_1 \rhd \mathrm{R}_2) \rhd \mathrm{R}_3}$ has the same expression (2.5):

$$\begin{aligned}
\boldsymbol{A}_{(\mathrm{R}_1 \rhd \mathrm{R}_2) \rhd \mathrm{R}_3} &= (\boldsymbol{A}_{\mathrm{R}_1} \otimes p_2 + \mathrm{id}_{\ell^2(V_1)} \otimes \boldsymbol{A}_{\mathrm{R}_2}) \otimes p_3 + (\mathrm{id}_{\ell^2(V_1)} \otimes \mathrm{id}_{\ell^2(V_2)}) \otimes \boldsymbol{A}_{\mathrm{R}_3} \\
&\simeq \boldsymbol{A}_{\mathrm{R}_1} \otimes p_2 \otimes p_3 + \mathrm{id}_{\ell^2(V_1)} \otimes \boldsymbol{A}_{\mathrm{R}_2} \otimes p_3 + \mathrm{id}_{\ell^2(V_1)} \otimes \mathrm{id}_{\ell^2(V_2)} \otimes \boldsymbol{A}_{\mathrm{R}_3}.
\end{aligned}$$

$\square$

†also called the rooted product

Let $\mathrm{R} = (V, E, o)$ be a finite rooted undirected graph, where "undirected" means that $(u, v) \in E$ implies $(v, u) \in E$, or equivalently, $\boldsymbol{A}_{\mathrm{R}}$ is a self-adjoint operator. Assume further that $(o, o) \notin E$ and

$$\deg(o) := \#\{v \in V : (v, o) \in E\} \geq 1.$$

Let $\mathrm{R}_N := \mathrm{R}^{\rhd N}$ be the N-fold comb product of R and o_N denote the root of R_N. Having no loop at o implies that $\boldsymbol{A}_{\mathrm{R}}$ has mean 0 with respect to φ_o. We can actually show

$$\lim_{N\to\infty} \varphi_{o_N}\left[\left(\frac{\boldsymbol{A}_{\mathrm{R}_N}}{\sqrt{\deg(o)N}}\right)^k\right] = \int_{-\sqrt{2}}^{\sqrt{2}} x^k \frac{\mathrm{d}x}{\pi\sqrt{2-x^2}} = \begin{cases} \frac{(k-1)!!}{(k/2)!}, & \text{if } k \text{ is even},\\ 0, & \text{otherwise}.\end{cases} \tag{2.6}$$

This is a direct consequence of the fact that the adjacency matrix of R_N is the sum of monotonically independent, identically distributed random variables of the form (1.8) with mean 0 and of the monotone CLT shown later in Theorem 3.18. Note that $\deg(o)$ is exactly the second moment of $\boldsymbol{A}_{\mathrm{R}}$ with respect to φ_o. The convergence (2.6) means that the number of paths of length k on R_N started from o_N and terminated at o_N is asymptotically

$$\left[\int_{-\sqrt{2}}^{\sqrt{2}} x^k \frac{\mathrm{d}x}{\pi\sqrt{2-x^2}} + o(1)\right](\deg(o)N)^{\frac{k}{2}}, \qquad N\to\infty.$$

2.5. **Notes.** The associativity of monotone independence is addressed in [66] in the setting of $*$-probability spaces, in which the proof was based on the operator model in Example 1.23. In order to handle the monotone product of nc-probability spaces, we adopted a more combinatorial proof of Proposition 2.1. The proof of positivity in Proposition 2.6 is similar to the case of free product of states, see e.g. [149, Definition 1.5.4].

Besides the monotone product, there are other universal ways to construct a nc-probability space $(A_1 \sqcup A_2, \varphi_1\varphi_2)$ from two nc-probability spaces (A_1, φ_1) and (A_2, φ_2). Of particular interest is the following four as all of them satisfy the associativity.

a) The antimonotone product

$$(\varphi_1 \lhd \varphi_2)(a_1a_2\cdots a_n) := \left[\prod_{k:\, a_k\in A_1} \varphi_1(a_k)\right]\varphi_2\left(\overrightarrow{\prod_{k:\, a_k\in A_2}} a_k\right),$$

which is just the flip of the monotone product and is essentially the same.

b) The tensor product

$$(\varphi_1 \otimes \varphi_2)(a_1a_2\cdots a_n) := \varphi_1\left(\overrightarrow{\prod_{k:\, a_k\in A_1}} a_k\right)\varphi_2\left(\overrightarrow{\prod_{k:\, a_k\in A_2}} a_k\right).$$

c) The Boolean product

$$(\varphi_1 \diamond \varphi_2)(a_1a_2\cdots a_n) := \prod_{k:\, a_k\in A_1} \varphi_1(a_k) \prod_{k:\, a_k\in A_2} \varphi_2(a_k).$$

d) The last one is called the free product and its definition is of different flavour. First we consider the unitization $\tilde{A}_i := \mathbb{C}\oplus A_i$ that naturally embed into $\tilde{A} := \mathbb{C}\oplus(A_1\sqcup A_2)$. We define φ on $\tilde{A}$ by requiring that $\varphi(a_1a_2\cdots a_n) = 0$ whenever $(a_1, a_2, \dots, a_n) \in \tilde{A}_{i_1}\times\tilde{A}_{i_2}\times\cdots\times\tilde{A}_{i_n}$, $i_1\neq i_2$, $i_2 \neq i_3$, $\dots$, $i_{n-1}\neq i_n$ and $\varphi(a_k) = 0$ for all $k\in[n]$. The free product $\varphi_1 * \varphi_2$ on $A_1\sqcup A_2$ is defined as the restriction $\varphi\restriction_{A_1\sqcup A_2}$.

The idea of identifying notions of independence as products of linear functionals on the free product algebra provides an appropriate framework for a classification program of independence. Speicher [146], and then Ben Ghorbal and Schürmann [28], showed that the associative products of linear functionals on the free product algebra with some conditions are only tensor, free and Boolean products. Muraki dropped one assumption of Ben Ghorbal and Schürmann, and as a result, the classification list contained two more products: monotone and antimonotone [123]. Muraki [124], Gerhold and Lachs [75] gave further results in this direction. As a closely related problem, Gerhold, Hasebe and Ulrich axiomatized and classified "universal operator models" that contain the ones in Examples 1.14 and 1.17 [74].

The relation between the comb product of graphs and monotone independence appeared in Accardi, Ben Ghorbal and Obata's work [1]. Arizmendi, Hasebe and Lehner studied the empirical eigenvalue distribution of the adjacency matrix of $\mathrm{R}^{\rhd N}$ in the large N limit [12]. The limit distribution is not universal and depends heavily on the original rooted graph R. There are other graph products that correspond to other independences in noncommutative probability. Among all, the adjacency matrix of the "free product of graphs" is the sum of freely independent operators. Accardi, Lenczewski and Sałapata decomposed the free product graph into the comb product of two subgraphs related to subordination of free convolution [2]. Lenczewski constructed a graph product that corresponds to conditionally monotone independence [106]. Garza-Vargas and Kulkarni constructed an amalgamated free product of graphs and applied it to the spectral analysis of Jacobi matrices of graphs [72]. Obata's monograph [127] contains other notable examples of graph products and references to earlier works.

3. Monotone cumulants

Cumulants are equivalents of moments and sometimes provide a clearcut description of random variables. In particular, (normalized) cumulants up to order four are called mean, variance, skewness and kurtosis, and are used in statistics. In probability theory, the characteristic function (or the Fourier transform) is often more powerful than cumulants because cumulants require that random variables have finite moments, while the characteristic function does not. However, in noncommutative probability theory, cumulants are more useful because the theory substantially builds upon moments. In free probability theory, Voiculescu introduced single-variate free cumulants [150] and then Speicher defined multivariate free cumulants [145] that have discovered a wide range of applications so far.

In classical probability theory, (single-variate) cumulants are numbers $\mathbb{K}_n(X)$, $n \geq 1$, defined for random variables X with finite moments, that satisfy the following axioms.

(C1) There are universal polynomials $P_n^{\otimes}(t_1, t_2, \ldots, t_{n-1})$, $n \geq 2$, with convention $P_1^{\otimes} = 0$, such that for all $X \in L^{\infty -}(\Omega, \mathcal{F}, \mathbb{P})$,

$$\mathbb{K}_n(X) = \mathbb{E}[X^n] + P_n^{\otimes}(\mathbb{E}[X], \mathbb{E}[X^2], \ldots, \mathbb{E}[X^{n-1}]), \qquad n \in \mathbb{N}. \qquad \text{(Polynomiality)}$$

(C2) For any $X \in L^{\infty -}(\Omega, \mathcal{F}, \mathbb{P})$,

$$\mathbb{K}_n(\lambda X) = \lambda^n \mathbb{K}_n(X), \qquad n \in \mathbb{N},\ \lambda \in \mathbb{C}. \qquad \text{(Homogeneity)}$$

(C3) If $X, Y \in L^{\infty -}(\Omega, \mathcal{F}, \mathbb{P})$ are independent then

$$\mathbb{K}_n(X + Y) = \mathbb{K}_n(X) + \mathbb{K}_n(Y), \qquad n \in \mathbb{N}. \qquad \text{(Additivity)}$$

Remark 3.1. In (C1), by "universal" we emphasize that $P_n^{\otimes}$ does not depend on X or the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Axiom (C1) is equivalent to the following reverted form of moments in terms of cumulants: there are universal polynomials $Q_n^{\otimes}(t_1, t_2, \ldots, t_{n-1})$, $n \geq 2$, with convention $Q_1^{\otimes} = 0$, such that for all $n \geq 1$ and $X \in L^{\infty -}(\Omega, \mathcal{F}, \mathbb{P})$

$$\mathbb{E}[X^n] = \mathbb{K}_n(X) + Q_n^{\otimes}(\mathbb{K}_1(X), \mathbb{K}_2(X), \ldots, \mathbb{K}_{n-1}(X)).$$

We can give a construction of $\mathbb{K}_n$ as the coefficients of the logarithm of exponential moment generating function:

$$\log \mathbb{E}[e^{zX}] = \log\left(\sum_{n=0}^{\infty} \frac{\mathbb{E}[X^n]}{n!} z^n\right) = \sum_{n=1}^{\infty} \frac{\mathbb{K}_n(X)}{n!} z^n. \tag{3.1}$$

Note that under the assumption $X \in L^{\infty -}(\Omega, \mathcal{F}, \mathbb{P})$, the above series might have convergence radius zero; then the above equalities are to be interpreted as formal power series.

Example 3.2. From the defining formula (3.1), the cumulants $\mathbb{K}_1, \mathbb{K}_2, \mathbb{K}_3$ are computed as

$$\begin{aligned} \mathbb{K}_1(X) &= \mathbb{E}[X], \\ \mathbb{K}_2(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2, \\ \mathbb{K}_3(X) &= \mathbb{E}[X^3] - 3\,\mathbb{E}[X^2]\mathbb{E}[X] + 2\,\mathbb{E}[X]^3. \end{aligned}$$

Our objective is to discover a monotone counterpart of cumulants, which we denote by κ_n. The most natural definition would be to replace the pair $(L^{\infty -}(\Omega, \mathcal{F}, \mathbb{P}), \mathbb{E})$ with a nc-probability space (A, φ), and replace the independence assumption in (C3) with monotone independence. However, the third condition would contradict the asymmetry of monotone independence. More precisely, suppose that x, y are monotonically independent real random variables in a unital C^*-probability space. Then $\kappa_n(x + y) = \kappa_n(x) + \kappa_n(y)$ does not depend on whether we assume x, y are monotonically independent or y, x are. However, monotone independence of x, y implies $M_{x+y}(z) = M_x(M_y(z))$ and monotone independence of y, x implies $M_{x+y}(z) = M_{y+x}(z) = M_y(M_x(z))$, and therefore, the distribution of $x + y$ is typically different if we switch the independence assumption.

Here we propose a weaker version of additivity, which we call extensivity because of its resemblance to the corresponding notion in thermodynamics.

Definition 3.3. A rule that associates with each nc-probability space (A, φ) and random variable $x \in A$ a sequence of complex numbers $(\kappa_n(x))_{n\geq 1}$ is called **monotone cumulants** if the following conditions hold.

(M1) There are universal polynomials $P_n^{\triangleright}(t_1, t_2, \ldots, t_{n-1})$, $n \geq 2$, with convention $P_1^{\triangleright} = 0$, such that for any nc-probability space (A, φ) and $x \in A$, we have

$$\kappa_n(x) = \varphi(x^n) + P_n^{\triangleright}(\varphi(x), \varphi(x^2), \ldots, \varphi(x^{n-1})), \qquad n \in \mathbb{N}. \qquad \text{(Polynomiality)}$$

(M2) For any nc-probability space (A, φ) and any $x \in A$, we have

$$\kappa_n(\lambda x) = \lambda^n \kappa_n(x), \qquad n \in \mathbb{N},\ \lambda \in \mathbb{C}. \qquad \text{(Homogeneity)}$$

(M3) If $N \in \mathbb{N}$ and $x_1, x_2, \ldots, x_N$ are monotonically independent and identically distributed (**monotonically iid** for short) random variables in a nc-probability space (A, φ), then

$$\kappa_n(x_1 + x_2 + \cdots + x_N) = N\kappa_n(x_1), \qquad n \in \mathbb{N}. \qquad \text{(Extensivity)}$$

Remark 3.4. Analogously to Remark 3.1, a recursive argument shows that condition (M1) is equivalent to that there are universal polynomials $Q_n^{\triangleright}(t_1, t_2, \ldots, t_{n-1})$, $n \geq 2$, with convention $Q_1^{\triangleright} := 0$, such that

$$\varphi(x^n) = \kappa_n(x) + Q_n^{\triangleright}(\kappa_1(x), \kappa_2(x), \ldots, \kappa_{n-1}(x)) \tag{3.2}$$

for all $n \geq 1$, $x \in A$ and (A, φ).

3.1. **Cumulants from moments of random walk.** We begin with showing the uniqueness of monotone cumulants, which also indicates how to show the existence. Note that the same reasoning below also applies to showing that classical cumulants $(\mathbb{K}_n)_{n\geq 1}$ are unique. For that purpose an elementary lemma on polynomials is needed.

Lemma 3.5. *Let $P(N) = a_0 + a_1 N + a_2 N^2 + \cdots + a_k N^k$ and $Q(N) = b_0 + b_1 N + b_2 N^2 + \cdots + b_k N^k$ be two polynomial functions on $\mathbb{N}$ with complex coefficients a_i, b_i. If $P(N) = Q(N)$ for all $N \in \mathbb{N}$ then $a_i = b_i$ for all $0 \leq i \leq k$.*

Remark 3.6. This lemma allows us to naturally extend a polynomial $P(N)$ defined for $N \in \mathbb{N}$ to a polynomial $P(t)$ defined for $t \in \mathbb{R}$.

The result easily extends to polynomials in several variables. For the case of two variables, if $P(N,M) = \sum_{i,j=1}^k a_{i,j} N^i M^j$ and $Q(N,M) = \sum_{i,j=1}^k b_{i,j} N^i M^j$ are polynomials with complex coefficients $a_{i,j}, b_{i,j}$ and $P(N,M) = Q(N,M)$ for all $N, M \in \mathbb{N}$ then $a_{i,j} = b_{i,j}$ for all $i, j \in [k]$. The proof is just to fix one variable, say M, and apply the lemma to $P(\cdot, M)$ and $Q(\cdot, M)$, which yields $\sum_{j=1}^k a_{i,j} M^j = \sum_{j=1}^k b_{i,j} M^j$ for all $M \in \mathbb{N}$ and i. Then again applying the lemma gives the conclusion $a_{i,j} = b_{i,j}$.

Proof. In fact, a weaker assumption is enough; suppose $P(N) = Q(N)$ holds at distinct positive integers $N_1 < N_2 < \cdots < N_{k+1}$. Then, by setting $c_i := a_i - b_i$, we have

$$\begin{pmatrix} 1 & N_1 & N_1^2 & \cdots & N_1^k \\ 1 & N_2 & N_2^2 & \cdots & N_2^k \\ \vdots & & \vdots & & \vdots \\ 1 & N_{k+1} & N_{k+1}^2 & \cdots & N_{k+1}^k \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Since the coefficient matrix has nonzero determinant (called the Vandermonde determinant), the numbers c_i must be zero. □

Proposition 3.7. *Monotone cumulants are unique.*

Proof. Suppose that $(\kappa_n)_{n\ge1}$ are monotone cumulants determined by universal polynomials $(P_n^{\rhd})_{n\ge1}$. We can see that $P_n^{\rhd}$ for $n \ge 2$ contains no linear terms or a constant term. Indeed, if we write

$$P_n^{\rhd}(t_1, t_2, \dots, t_{n-1}) = \sum_{k_1, k_2, \dots, k_{n-1} \ge 0} c_{k_1,k_2,\dots,k_{n-1}} t_1^{k_1} t_2^{k_2} \cdots t_{n-1}^{k_{n-1}},$$

where $c_{k_1,k_2,\dots,k_{n-1}}$ are complex constants independent of (A, φ) and the tuple $(k_1, k_2, \dots, k_{n-1})$ runs over a finite subset of $\mathbb{N}_0^{n-1}$, then the homogeneity condition reads

$$\lambda^n \kappa_n(x) = \lambda^n \varphi(x^n) + \sum_{k_1,k_2,\dots,k_{n-1}\ge 0} c_{k_1,k_2,\dots,k_{n-1}} \lambda^{k_1+2k_2+\cdots+(n-1)k_{n-1}} \varphi(x)^{k_1} \varphi(x^2)^{k_2} \cdots \varphi(x^{n-1})^{k_{n-1}}.$$

Since this holds for all $\lambda \in \mathbb{C}$, comparing the coefficients of λ^p $(p \ne n)$ yields

$$\sum_{\substack{k_1,k_2,\dots,k_{n-1}\ge 0 \\ k_1+2k_2+\cdots+(n-1)k_{n-1}=p}} c_{k_1,k_2,\dots,k_{n-1}} \varphi(x)^{k_1} \varphi(x^2)^{k_2} \cdots \varphi(x^{n-1})^{k_{n-1}} = 0, \qquad p \ne n.$$

Because of the arbitrariness of (A, φ) and $x \in A$, the tuple $(\varphi(x), \varphi(x^2), \dots, \varphi(x^{n-1}))$ can take arbitrary vectors in $\mathbb{C}^{n-1}$. Therefore, we conclude $c_{k_1,k_2,\dots,k_{n-1}} = 0$ unless $k_1 + 2k_2 + \cdots + (n-1)k_{n-1} = n$. In particular, the constant term and linear terms of $P_n^{\rhd}$ are all zero. This also implies that the constant and linear terms of $Q_n^{\rhd}$ in (3.2) are all zero.

Let us take monotonically iid random variables $x_1, x_2, \dots, x_N$ in some nc-probability space. Then the extensivity condition yields, with notation $x = x_1$,

$$\varphi((x_1 + x_2 + \cdots + x_N)^n) = N\kappa_n(x) + Q_n^{\rhd}(N\kappa_1(x), N\kappa_2(x), \dots, N\kappa_{n-1}(x)).$$

The right-hand side is a polynomial in positive integers N and hence, by Lemma 3.5, their coefficients are uniquely determined. In particular, since the $Q_n^{\rhd}$ part has no linear term, $\kappa_n(x)$ is uniquely determined as the coefficient of N of $\varphi((x_1 + x_2 + \cdots + x_N)^n)$. □

The above proof also indicates how we can find monotone cumulants: $\kappa_n(x)$ should be the coefficient of N of *the nth moment of monotone random walk* $\varphi((x_1 + x_2 + \cdots + x_N)^n)$. In order for this definition to make sense, we need to show $\varphi((x_1 + x_2 + \cdots + x_N)^n)$ is a polynomial in N.

Proposition 3.8. *There are universal polynomials $U_n^{\rhd}(s, t_1, t_2, \dots, t_{n-1})$, $n \in \mathbb{N}$, such that $U_1^{\rhd}(s) = 0$, $U_n^{\rhd}(0, t_1, t_2, \dots, t_{n-1}) = 0$ $(n \ge 2)$ and*

$$\varphi((x_1 + x_2 + \cdots + x_N)^n) = N\varphi(x^n) + U_n^{\rhd}(N, \varphi(x), \varphi(x^2), \dots, \varphi(x^{n-1})), \qquad n \ge 1$$

for any monotonically iid random variables $x_1, x_2, \dots, x_N$ in any nc-probability space (A, φ) with notation $x = x_1$.

Proof. The proof is based on induction on n. For $n = 1$ we have

$$\varphi(x_1 + x_2 + \cdots + x_N) = N\varphi(x).$$

Suppose that $n \ge 2$ and the statement is the case up to $n-1$. We set $s_k := x_1 + x_2 + \cdots + x_k$ and $s_0 := 0$. By Corollary 2.5, s_{N-1} and x_N are monotonically independent. According to the moment calculation (1.10), there exists a universal polynomial $T_n^{\rhd}$ of $2n-2$ variables such that

$$\begin{aligned} \varphi(s_N^n) &= \varphi((s_{N-1} + x_N)^n) \\ &= \varphi(s_{N-1}^n) + \varphi(x_N^n) + T_n^{\rhd}(\varphi(s_{N-1}), \varphi(s_{N-1}^2), \dots, \varphi(s_{N-1}^{n-1}), \varphi(x_N), \varphi(x_N^2), \dots, \varphi(x_N^{n-1})) \\ &= \varphi(s_{N-1}^n) + \varphi(x^n) + \underbrace{T_n^{\rhd}(\varphi(s_{N-1}), \varphi(s_{N-1}^2), \dots, \varphi(s_{N-1}^{n-1}), \varphi(x), \varphi(x^2), \dots, \varphi(x^{n-1}))}_{=: S_n}. \end{aligned}$$

By the induction hypothesis, for each $1 \le k \le n-1$, $\varphi(s_{N-1}^k)$ is a polynomial in $N-1$ and $\varphi(x), \varphi(x^2), \dots, \varphi(x^k)$, so that S_n is a polynomial in $N, \varphi(x), \varphi(x^2), \dots, \varphi(x^{n-1})$. Consequently, there are polynomials $V_{n,k}^{\rhd}(t_1, t_2, \dots, t_{n-1})$, $0 \le k \le d_n$, such that

$$\varphi(s_N^n) - \varphi(s_{N-1}^n) = \varphi(x^n) + \sum_{k=0}^{d_n} V_{n,k}^{\rhd}(\varphi(x), \varphi(x^2), \dots, \varphi(x^{n-1}))N^k.$$

Taking the sum over N we obtain

$$\varphi(s_N^n) = N\varphi(x^n) + \sum_{k=0}^{d_n} V_{n,k}^{\rhd}(\varphi(x), \varphi(x^2), \dots, \varphi(x^{n-1})) \sum_{M=1}^{N} M^k.$$

By Faulharbor's formula, $F_k(N) := \sum_{M=1}^{N} M^k$ is a polynomial in N of degree $k+1$ without a constant term. □

Remark 3.9. A more straightforward proof will be provided in Example 3.12.

We are now ready to prove the existence part of monotone cumulants.

Theorem 3.10. *Let (A, φ) be a nc-probability space and $x \in A$. We take a sequence $(x_i)_{i=1}^{\infty}$ of monotonically iid random variables with the same distribution as x in a possibly different nc-probability space $(\hat{A}, \hat{\varphi})$. Let $\kappa_n(x)$ be the coefficient of N of the polynomial $\hat{\varphi}((x_1 + x_2 + \dots + x_N)^n)$ in N. Then κ_n, $n \in \mathbb{N}$, are monotone cumulants, i.e., conditions* (M1)–(M3) *hold. Moreover, the polynomials $P_n^{\rhd}$ $(n \ge 2)$ have no constant or linear terms.*

Remark 3.11. The above definition of $\kappa_n(x)$ does not depend on how to select $(\hat{A}, \hat{\varphi})$ and (x_i) as the nth moment $\hat{\varphi}((x_1 + x_2 + \dots + x_N)^n)$ is determined only by the distribution of x and $N \in \mathbb{N}$. In general, we cannot take (x_i) in the same space (A, φ). A canonical construction of $(\hat{A}, \hat{\varphi})$ is as follows: we set $A_i := A$, $\varphi_i := \varphi$ and

$$(\hat{A}, \hat{\varphi}) := \mathop{\rhd}_{i \in \mathbb{N}} (A_i, \varphi_i).$$

Then we consider the natural embeddings of A into $\hat{A}$ as $\iota_i \colon A \to A_i \subseteq \hat{A}$, $i \in \mathbb{N}$. For each $x \in A$, the random variables $x_i := \iota_i(x)$, $i \in \mathbb{N}$, are by construction monotonically independent and have the same distribution as x, see Section 2.

Proof of Theorem 3.10. For notational simplicity we denote $\hat{\varphi}$ as φ. Condition (M1) is clear from Proposition 3.8. Condition (M2) holds because $\varphi((\lambda x_1 + \lambda x_2 + \dots + \lambda x_N)^n) = \lambda^n \varphi((x_1 + x_2 + \dots + x_N)^n)$. In order to show condition (M3), we fix $N \in \mathbb{N}$ and set

$$y_i := x_{N(i-1)+1} + x_{N(i-1)+2} + \dots + x_{Ni}, \qquad i \in \mathbb{N}.$$

By Corollary 2.5, the sequence $(y_i)_{i=1}^{\infty}$ is monotonically independent. Also, $(y_i)_{i=1}^{\infty}$ is identically distributed. For each $n \in \mathbb{N}$, the coefficient of M of

$$\varphi((y_1 + y_2 + \dots + y_M)^n)$$

equals $\kappa_n(y_1)$, which is $\kappa_n(x_1 + x_2 + \dots + x_N)$. On the other hand,

$$\varphi((y_1 + y_2 + \dots + y_M)^n) = \varphi((x_1 + x_2 + \dots + x_{MN})^n)$$

is a polynomial in MN whose coefficient of MN is $\kappa_n(x_1)$, and therefore the coefficient of M is $N\kappa_n(x_1)$. Combining the above arguments we conclude (M3). The last assertion on $P_n^{\rhd}$ is already proved in Proposition 3.7. □

Example 3.12. From condition (M1), we have $\kappa_1(x) = \varphi(x)$. We compute monotone cumulants κ_2, κ_3 by finding the polynomials $U_2^{\rhd}, U_3^{\rhd}$ in Proposition 3.8. The method here is more straightforward than the proof of Proposition 3.8.

Formula for $U_2^{\rhd}$. It should be kept in mind that $\varphi(x_i^n) = \varphi(x^n)$ does not depend on i because $x = x_1, x_2, x_3, \dots$ have an identical distribution. We first compute

$$\begin{aligned}\varphi(s_N^2) &= \sum_{i,j=1}^{N} \varphi(x_i x_j) = \sum_{i \ne j} \varphi(x_i x_j) + \sum_i \varphi(x_i^2) \\ &= \sum_{i \ne j} \varphi(x_i)\varphi(x_j) + \sum_i \varphi(x_i^2) = \sum_{i \ne j} \varphi(x)\varphi(x) + \sum_i \varphi(x^2) \\ &= N(N-1)\varphi(x)\varphi(x) + N\varphi(x^2),\end{aligned}$$

so that $U_2^{\rhd}(s,t) = s(s-1)t^2$.

Formula for $U_3^{\rhd}$. We begin with

$$\begin{aligned}\varphi(s_N^3) &= \sum_{i,j,k=1}^{N} \varphi(x_i x_j x_k) \\ &= \sum_{i,j,k \text{ distinct}} \varphi(x_i x_j x_k) + \sum_{i=j \ne k} \varphi(x_i^2 x_k) + \sum_{i=k \ne j} \varphi(x_i x_j x_i) + \sum_{i \ne k=j} \varphi(x_i x_j^2) + \sum_{i=j=k} \varphi(x_i^3) \\ &= \sum_{i,j,k \text{ distinct}} \varphi(x_i x_j x_k) + \sum_{i=j \ne k} \varphi(x_i^2)\varphi(x_k) + \sum_{i=k \ne j} \varphi(x_i x_j x_i) + \sum_{i \ne k=j} \varphi(x_i)\varphi(x_j^2) + \sum_{i=j=k} \varphi(x_i^3).\end{aligned}$$

In the above, $\varphi(x_i x_j x_k)$ for distinct integers i, j, k always factorizes into $\varphi(x_i)\varphi(x_j)\varphi(x_k)$; for example, if $i < j > k$ and $i \ne k$ then $\varphi(x_i x_j x_k) = \varphi(x_i x_k)\varphi(x_j) = \varphi(x_i)\varphi(x_j)\varphi(x_k)$ and if $i < j < k$ then $\varphi(x_i x_j x_k) = \varphi(x_i x_j)\varphi(x_k) = \varphi(x_i)\varphi(x_j)\varphi(x_k)$. On the other hand, the sum over $i = k \ne j$ is more delicate. In order to use monotone independence, we need to further specify the inequality between i and j. If $i<j$ then $\varphi(x_ix_jx_i)=\varphi(x_j)\varphi(x_i^2)$. If $i>j$ then $\varphi(x_ix_jx_i)=\varphi(x_i)\varphi(x_jx_i)=\varphi(x_i)\varphi(x_j)\varphi(x_i)$. Therefore,

$$\sum_{i=k\neq j}\varphi(x_ix_jx_i)=\sum_{i<j}\varphi(x_ix_jx_i)+\sum_{i>j}\varphi(x_ix_jx_i)=\sum_{i<j}\varphi(x_i^2)\varphi(x_j)+\sum_{i>j}\varphi(x_i)^2\varphi(x_j).$$

Overall, we arrive at

$$\begin{aligned}\varphi(s_N^3)&=N(N-1)(N-2)\varphi(x)^3+N(N-1)\varphi(x^2)\varphi(x)+\frac{N(N-1)}{2}\varphi(x^2)\varphi(x)\\&\quad+\frac{N(N-1)}{2}\varphi(x)^3+N(N-1)\varphi(x^2)\varphi(x)+N\varphi(x^3)\\&=N(N-1)\left(N-\frac{3}{2}\right)\varphi(x)^3+\frac{5N(N-1)}{2}\varphi(x^2)\varphi(x)+N\varphi(x^3).\end{aligned}$$

This means

$$U_3^{\rhd}(s,t_1,t_2)=s(s-1)\left(s-\frac{3}{2}\right)t_1^3+\frac{5s(s-1)}{2}t_1t_2.$$

The above method can be generalized to any $U_n^{\rhd}$, which provides an alternative proof of Proposition 3.8.

Formulas for κ_2,κ_3. The monotone cumulants $\kappa_2(x)$ and $\kappa_3(x)$ are identified with the coefficients of N of $\varphi(s_N^2)$ and of $\varphi(s_N^3)$ respectively:

$$\kappa_2(x)=\varphi(x^2)-\varphi(x)^2; \tag{3.3}$$

$$\kappa_3(x)=\varphi(x^3)-\frac{5}{2}\varphi(x^2)\varphi(x)+\frac{3}{2}\varphi(x)^3. \tag{3.4}$$

A recursive formula for computing $\kappa_n(x)$ will be provided in Proposition 3.15. A combinatorial formula for $\varphi(x^n)$ in terms of $\kappa_\ell(x)$, $1\leq\ell\leq n$, will be given in Theorem 3.26.

Remark 3.13. The nth monotone cumulant $\kappa_n(x)$ is determined by the moments of the random variable x up to order n. Therefore, for any probability measure μ having finite moments up to order n, we can define $\kappa_n(\mu):=\kappa_n(x)$ by taking a random variable x in a nc-probability space (A,φ) such that $\varphi(x^p)=\int_{\mathbb{R}}t^p\,\mu(\mathrm{d}t)$, $1\leq p\leq n$. We call $\kappa_n(\mu)$ the **nth monotone cumulant of** μ.

Example 3.14. Generally, monotone cumulants $\kappa_n(\mu)$ are hard to calculate unless the probability measure μ embeds into a *monotone convolution semigroup*, see Section 5.2. Although the semicircle distribution $\mathrm{S}(0,1)=(2\pi)^{-1}\sqrt{4-t^2}\,\chi_{[-2,2]}(t)\,\mathrm{d}t$ does not admit such an embedding (see Example 5.25), a remarkable recurrence formula is given in [56]. Let $d_n(N)$ denote the expectation $\varphi((x_1+x_2+\cdots+x_N)^{2n})$ where x_i has the analytic distribution $\mathrm{S}(0,1)$; note that odd moments are all zero. Then the following recurrence relation holds:

$$d_n(N)=\sum_{k=1}^{n}\sum_{j=1}^{N}d_{n-k}(N)d_{k-1}(j).$$

This yields e.g. $d_1(N)=N$, $d_2(N)=\frac{3}{2}N^2+\frac{1}{2}N$ and $d_3(N)=\frac{5}{2}N^3+2N^2+\frac{1}{2}N$, so that the monotone cumulants up to order six are given by $(0,1,0,1/2,0,1/2)$.

3.2. **Differential recursion for monotone cumulants.** To compute monotone cumulants, differential recursion is helpful.

Proposition 3.15. *Let x be a random variable in a nc-probability space (A,φ). Let $(x_i)_{i=1}^{\infty}$ be a sequence of monotonically iid random variables with the same distribution as x in a possibly different nc-probability space $(\hat{A},\hat{\varphi})$, see Remark 3.11. As we have seen above, for each $n\in\mathbb{N}$ the evaluation $\hat{\varphi}((x_1+x_2+\cdots+x_N)^n)$ is a polynomial in N, so it can be extended to a polynomial in real variable $t\in\mathbb{R}$, which we denote by $\mathrm{m}_n^{\rhd}(t)\equiv\mathrm{m}_n^{\rhd}(t;x)$. Then we have $\mathrm{m}_0^{\rhd}(t)\equiv1$ and for $n\in\mathbb{N}$*

$$\begin{cases}\dfrac{\mathrm{d}}{\mathrm{d}t}\mathrm{m}_n^{\rhd}(t)=\displaystyle\sum_{\ell=0}^{n-1}(\ell+1)\kappa_{n-\ell}(x)\mathrm{m}_\ell^{\rhd}(t),\\ \mathrm{m}_n^{\rhd}(0)=0.\end{cases} \tag{3.5}$$

Proof. For simplicity we denote $\hat{\varphi}$ as φ. Let $s_N:=x_1+x_2+\cdots+x_N$, $s_0:=0$. Note first that $\mathrm{m}_n^{\rhd}(0)=0$ for $n\geq1$ comes from Proposition 3.8 in which it is shown that $\varphi(s_N^n)$ has no constant term in N. Let $s'_M:=x_{N+1}+x_{N+2}+\cdots+x_{N+M}$. By Corollary 2.5, s_N and s'_M are monotonically independent. Note that s'_M also depends on N as an element of $\hat{A}$; however its distribution only depends on the number M of the summands x_i's, so that we omit explicitly mentioning the dependence on N. In the obvious formula

$$\mathrm{m}_n^{\rhd}(N+M)=\varphi((x_1+x_2+\cdots+x_{N+M})^n)=\varphi((s_N+s'_M)^n),$$

the right-hand side is exactly the nth moment of additive monotone convolution, so we can use the calculation in (1.10):

$$\varphi((s_N+s'_M)^n)=\sum_{\ell=0}^{n}\sum_{\substack{k_0,k_1,\dots,k_\ell\geq0,\\k_0+k_1+\cdots+k_\ell=n-\ell}}\varphi(s_N^\ell)\varphi(s_M^{k_0})\varphi(s_M^{k_1})\cdots\varphi(s_M^{k_\ell}). \tag{3.6}$$

Since each $\varphi(s_M^k)$ is a polynomial in M without a constant term, the contributions to the monomial M^1 in the sum (3.6) only come from the tuples $(k_0, k_1, \dots, k_\ell)$, $0 \le \ell \le n-1$, such that exactly one of k_i's is nonzero:

$$\varphi((s_N + s'_M)^n) = \sum_{\ell=0}^{n-1} (\ell+1)\varphi(s_N^\ell)\varphi(s_M^{n-\ell}) + R_1(M)$$
$$= M \sum_{\ell=0}^{n-1} (\ell+1)\varphi(s_N^\ell)\kappa_{n-\ell}(x) + R_2(M),$$

where $R_1(M)$ and $R_2(M)$ are polynomials in M without a constant or linear term. Since all the terms are polynomials, we can extend the variables N and M to real numbers t and s, so that

$$\mathrm{m}_n^{\triangleright}(t+s) = s\sum_{\ell=0}^{n-1}(\ell+1)\kappa_{n-\ell}(x)\mathrm{m}_\ell^{\triangleright}(t) + R_2(s).$$

The desired formula follows by taking the derivative $\frac{\mathrm{d}}{\mathrm{d}s}\big|_{s=0}$. □

The differential recursion gives an efficient method for computing $\kappa_n(x)$.

Example 3.16. We set $n = 1$ in (3.5) to obtain $(\mathrm{m}_1^{\triangleright})'(t) = \kappa_1(x)$, which integrates to

$$\mathrm{m}_1^{\triangleright}(t) = \kappa_1(x)t.$$

Setting $t = 1$, we obtain

$$\kappa_1(x) = \mathrm{m}_1^{\triangleright}(1) = \varphi(x),$$

which is already known. Formula (3.5) for $n = 2$ reads

$$(\mathrm{m}_2^{\triangleright})'(t) = \kappa_2(x)\mathrm{m}_0^{\triangleright}(t) + 2\kappa_1(x)\mathrm{m}_1^{\triangleright}(t) = \kappa_2(x) + 2\kappa_1(x)^2 t,$$

and hence

$$\mathrm{m}_2^{\triangleright}(t) = \kappa_2(x)t + \kappa_1(x)^2t^2.$$

Setting $t = 1$ and using $\kappa_1(x) = \varphi(x)$ we obtain formula (3.3) for κ_2. In a similar manner we obtain

$$\mathrm{m}_3^{\triangleright}(t) = \kappa_3(x)t + \frac{5}{2}\kappa_2(x)\kappa_1(x)t^2 + \kappa_1(x)^3t^3$$

and so $\kappa_3(x)$ is given by formula (3.4). If n becomes larger and larger, this method of differential recursion for computing κ_n seems more efficient than the one in Example 3.12.

Corollary 3.17. *As a polynomial in t, we have $\deg(\mathrm{m}_n^{\triangleright}(t)) \le n$ for all $n \in \mathbb{N}_0$.*

Proof. Integrating the differential recursion in Proposition 3.15 yields

$$\mathrm{m}_n^{\triangleright}(t) = \sum_{\ell=0}^{n-1}(\ell+1)\kappa_{n-\ell}(x)\int_0^t \mathrm{m}_\ell^{\triangleright}(s)\,\mathrm{d}s, \qquad n \ge 1.$$

Starting from $\mathrm{m}_0^{\triangleright}(t) \equiv 1$, we can show the bound $\deg(\mathrm{m}_n^{\triangleright}(t)) \le n$ by induction. □

3.3. **Monotone central limit theorem.** In probability theory, a basic form of the central limit theorem says if $(X_i)_{i\ge1}$ is real-valued iid random variables such that $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^2] = \sigma^2 > 0$, then the distribution of

$$\frac{X_1 + X_2 + \cdots + X_N}{\sqrt{N}}$$

converges weakly to $\mathrm{N}(0, \sigma^2)$ as $N \to \infty$. We consider a similar problem for monotonically independent random variables.

Theorem 3.18. *Let (A, φ) be a unital C^*-probability space and $(x_i)_{i\ge1}$ be a sequence of monotonically iid real random variables in A. Suppose that $\varphi(x_1) = 0$ and $\varphi(x_1^2) = \sigma^2 > 0$. Then, for any $n \in \mathbb{N}$*

$$\lim_{N\to\infty}\varphi\left(\left(\frac{x_1 + x_2 + \cdots + x_N}{\sqrt{N}}\right)^n\right) = \int_{-\sqrt{2}\sigma}^{\sqrt{2}\sigma}\frac{t^n}{\pi\sqrt{2\sigma^2 - t^2}}\,\mathrm{d}t.$$

In particular, the analytic distribution of $(x_1 + x_2 + \cdots + x_N)/\sqrt{N}$ converges weakly to the arcsine law $\mathrm{A}(0, \sigma^2)$.

Proof. We first prove the convergence of monotone cumulants of $a_N := (x_1 + x_2 + \cdots + x_N)/\sqrt{N\sigma^2}$. By using conditions (M1) and (M2) we have

$$\kappa_n(a_N) = (N\sigma^2)^{-\frac{n}{2}}\kappa_n(x_1 + x_2 + \cdots + x_N) = (N\sigma^2)^{-\frac{n}{2}}N\kappa_n(x_1).$$

Recall here that $\kappa_1(x_1) = \varphi(x_1) = 0$ and $\kappa_2(x_1) = \varphi(x_1^2) - \varphi(x_1)^2 = \sigma^2$. Passing to the limit yields

$$\kappa_n := \lim_{N\to\infty}\kappa_n(a_N) = \begin{cases} 1, & \text{if } n = 2, \\ 0, & \text{otherwise.}\end{cases}$$

By the polynomiality $\varphi(a_N^n) = \kappa_n(a_N) + Q_n^{\triangleright}(\kappa_1(a_N), \kappa_2(a_N), \dots, \kappa_{n-1}(a_N))$ and taking the limit, we obtain the convergence

$$\lim_{N\to\infty}\varphi(a_N^n) = \kappa_n + Q_n^{\triangleright}(\kappa_1, \kappa_2, \dots, \kappa_{n-1}), \qquad n \ge 1.$$

To compute this limit, we use the differential recursion. Let $\mathrm{m}_n^{\triangleright}(t; a_N)$ be the polynomial constructed for $x := a_N$ as in Proposition 3.15. From Example 3.16, we have $\mathrm{m}_1^{\triangleright}(t; a_N) = \varphi(a_N)t = 0$. Since $\mathrm{m}_n^{\triangleright}(t; a_N)$ is a polynomial in t and $\varphi(a_N^k)$, $1 \le k \le n$, the limit

$$\mathrm{m}_n(t) := \lim_{N\to\infty}\mathrm{m}_n^{\triangleright}(t; a_N)$$

exists. Since $\deg(\mathrm{m}_n^{\rhd}(t;a_N)) \le n$ and the coefficient of each monomial t^k $(1 \le k \le n)$ of $\mathrm{m}_n^{\rhd}(t;a_N)$ converges, the limit function $\mathrm{m}_n(t)$ is also a polynomial with degree $\le n$ and the convergence is uniform on each finite interval of $\mathbb{R}$. By taking the limit in the integrated form of Proposition 3.15, we obtain

$$\mathrm{m}_n(t) = \lim_{N\to\infty} \mathrm{m}_n^{\rhd}(t;a_N) = \lim_{N\to\infty} \sum_{\ell=0}^{n-1} (\ell+1)\kappa_{n-\ell}(a_N) \int_0^t \mathrm{m}_\ell^{\rhd}(s;a_N)\,\mathrm{d}s = \int_0^t (n-1)\mathrm{m}_{n-2}(s)\,\mathrm{d}s.$$

Since $\mathrm{m}_0(t) = 1$ and $\mathrm{m}_1(t) = 0$ for all $t \in \mathbb{R}$, this can be easily solved by iterated integrals as

$$\mathrm{m}_{2k}(t) = \frac{(2k-1)!!}{k!} t^k, \qquad \mathrm{m}_{2k-1}(t) = 0, \qquad k \ge 1.$$

As $\mathrm{m}_n^{\rhd}(1;a_N) = \varphi(a_N^n)$, we have thus obtained

$$\lim_{N\to\infty} \varphi(a_N^n) = \mathrm{m}_n(1) = \begin{cases} \dfrac{(n-1)!!}{(n/2)!}, & \text{if } n \text{ is even,} \\ 0, & \text{if } n \text{ is odd.} \end{cases} \tag{3.7}$$

This limit is exactly the moment sequence of the standard arcsine law $\mathrm{A}(0,1)$, which has compact support and hence its moment sequence is determinate (Proposition A.4). The weak convergence is a consequence of Proposition A.7. ☐

3.4. **Poisson's law of small numbers.** The second limit theorem to be discussed is an analogue of Poisson's law of small numbers. In probability theory, the simplest formulation is as follows. Let $\lambda > 0$ be a fixed parameter. For each $N \in \mathbb{N}$ with $N > \lambda$, suppose that $X_{N,1}, X_{N,2}, \dots, X_{N,N}$ are independent random variables that have the identical Bernoulli distribution

$$\mathbb{P}(X_{N,i} = 0) = 1 - \frac{\lambda}{N}, \qquad \mathbb{P}(X_{N,i} = 1) = \frac{\lambda}{N}, \qquad 1 \le i \le N. \tag{3.8}$$

Then the distribution of $X_{N,1} + X_{N,2} + \cdots + X_{N,N}$ converges weakly to the Poisson distribution with rate λ

$$\mathrm{Poi}^{\otimes}(\lambda) = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \delta_n.$$

Replacing the notion of independence with monotone independence, we formulate and obtain the following limit theorem.

Theorem 3.19. *Let $\lambda > 0$. For each $N \in \mathbb{N}$ with $N > \lambda$, let $(x_{N,i})_{i=1}^N$ be monotonically iid real random variables in a unital C^*-probability space (A, φ) such that*

$$\mu_{x_{N,i}} = \left(1 - \frac{\lambda}{N}\right)\delta_0 + \frac{\lambda}{N}\delta_1, \qquad 1 \le i \le N.$$

Then there exists a probability measure $\mathrm{Poi}^{\rhd}(\lambda)$ whose monotone cumulants are all equal to λ such that

$$\lim_{N\to\infty} \varphi((x_{N,1} + x_{N,2} + \cdots + x_{N,N})^n) = \int_{\mathbb{R}} t^n \,\mathrm{Poi}^{\rhd}(\lambda)(\mathrm{d}t), \qquad n \in \mathbb{N}.$$

Remark 3.20. We will study $\mathrm{Poi}^{\rhd}(\lambda)$ further in Example 5.21, where $\mathrm{Poi}^{\rhd}(\lambda)$ turns out to have compact support. Therefore, the analytic distribution of $x_{N,1} + x_{N,2} + \cdots + x_{N,N}$ converges weakly to $\mathrm{Poi}^{\rhd}(\lambda)$. The measure $\mathrm{Poi}^{\rhd}(\lambda)$ will be called the **monotone Poisson distribution** with parameter $\lambda > 0$.

Proof of Theorem 3.19. Let $a_N := x_{N,1} + x_{N,2} + \cdots + x_{N,N}$. Since $(x_{N,i})_{i=1}^N$ is monotonically iid, we have

$$\kappa_n(a_N) = N\kappa_n(x_{N,1}).$$

Observe here that $\varphi(x_{N,1}^n) = \lambda/N$. From condition (M1), since $P_n^{\rhd}$ has no constant or linear terms,

$$N\kappa_n(x_{N,1}) = N\varphi(x_{N,1}^n) + NP_n^{\rhd}(\varphi(x_{N,1}), \varphi(x_{N,1}^2), \dots, \varphi(x_{N,1}^{n-1})) = \lambda + o(1).$$

Therefore we conclude that

$$\lim_{N\to\infty} \kappa_n(a_N) = \lambda.$$

This in turn implies

$$\lim_{N\to\infty} \varphi(a_N^n) = \lambda + Q_n^{\rhd}(\lambda, \lambda, \dots, \lambda). \tag{3.9}$$

Because for each N the sequence $\varphi(a_N^n)$, $n = 0, 1, 2, \dots$ is positive semi-definite, the limit sequence is also positive semi-definite. Theorem A.1 ensures that the limit (3.9) is the moment sequence of some probability measure on $\mathbb{R}$. ☐

3.5. **Monotone set partitions and monotone cumulants.** Cumulants are known to be intimately connected to set partitions. Let S be a finite set. A decomposition of S into nonempty disjoint subsets is called a **set partition** of S. A set partition is denoted as $\rho = \{B_1, B_2, \dots, B_k\}$, where B_i's are the nonempty disjoint subsets of S such that $S = B_1 \cup \cdots \cup B_k$. Each B_i is called a **block** of the set partition ρ. The number k of the blocks of ρ is denoted by $|\rho|$. Let $\mathbf{P}(S)$ stand for the set of all set partitions of S. For the special case $S = [n]$, we denote $\mathbf{P}(n) := \mathbf{P}([n])$. For example $\mathbf{P}(2)$ has two elements $\{\{1,2\}\}$ and $\{\{1\},\{2\}\}$ and $\mathbf{P}(3)$ has five elements $\{\{1,2,3\}\}, \{\{1,2\},\{3\}\}, \{\{1,3\},\{2\}\}, \{\{1\},\{2,3\}\}, \{\{1\},\{2\},\{3\}\}$.

Let $(\alpha_n)_{n\ge1}$ be a sequence of complex numbers. For each set partition $\rho = \{B_1, B_2, \dots, B_k\}$ of S, we define

$$\alpha_\rho := \alpha_{|B_1|}\alpha_{|B_2|}\cdots\alpha_{|B_k|}.$$

With this notation, the following formula holds.

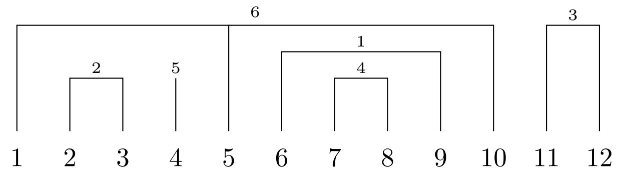

FIGURE 4. The diagram of $\pi = (B_1, B_2, \ldots, B_6) \in \mathbf{OP}(12)$ where $B_1 = \{6, 9\}$, $B_2 = \{2, 3\}$, $B_3 = \{11, 12\}$, $B_4 = \{7, 8\}$, $B_5 = \{4\}$, and $B_6 = \{1, 5, 10\}$. This is not a monotone set partition because e.g. B_6 covers B_1. For example, the permuted one $\tilde{\pi} = (B_6, B_2, B_5, B_1, B_4, B_3)$ is a monotone set partition.

Proposition 3.21. *For a random variable $X \in L^{\infty-}(\Omega, \mathcal{F}, \mathbb{P})$, we have*

$$\mathbb{E}[X^n] = \sum_{\rho \in \mathbf{P}(n)} \mathbb{K}_\rho(X), \qquad n \in \mathbb{N}. \tag{3.10}$$

Proof. Although we do not need this formula later, the proof is sketched for the reader's convenience. We can see that, given a sequence of positive integers $(i_1, i_2, \ldots, i_k)$ with $i_1 + 2i_2 + \cdots + ki_k = n$, the number of $\rho \in \mathbf{P}(n)$ that has i_1 blocks of cardinality one, i_2 blocks of cardinality two, $\ldots$, i_k blocks of cardinality k, equals

$$\begin{aligned} &\binom{n}{i_1}\binom{n - i_1}{2i_2}\binom{n - i_1 - 2i_2}{3i_3} \cdots \binom{n - i_1 - 2i_2 - \cdots - (k-1)i_{k-1}}{ki_k} \prod_{p=1}^{k} \frac{(pi_p)!}{(p!)^{i_p}(i_p!)} \\ &= \frac{n!}{i_1! i_2! \cdots i_k! (1!)^{i_1}(2!)^{i_2} \cdots (k!)^{i_k}}, \end{aligned} \tag{3.11}$$

so that the coefficient of $\mathbb{K}_1(X)^{i_1}\mathbb{K}_2(X)^{i_2} \cdots \mathbb{K}_k(X)^{i_k}$ in (3.10) is exactly (3.11). On the other hand, recall that the definition of cumulants is given by $\mathbb{E}[e^{zX}] = \exp\left(\sum_{n \ge 1} \frac{\mathbb{K}_n(X)}{n!} z^n\right)$, so that

$$\sum_{n \ge 0} \frac{\mathbb{E}[X^n]}{n!} z^n = e^{\mathbb{K}_1(X) z} e^{\frac{\mathbb{K}_2(X)}{2!} z^2} e^{\frac{\mathbb{K}_3(X)}{3!} z^3} \cdots \tag{3.12}$$

in the sense of formal power series. The coefficient of $\mathbb{K}_1(X)^{i_1}\mathbb{K}_2(X)^{i_2} \cdots \mathbb{K}_k(X)^{i_k} z^n$ in (3.12) is easily seen to be the number (3.11) divided by $n!$, as desired. ☐

Example 3.22. For $n = 1, 2, 3$, formula (3.10) reads

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{K}_1(X), \\ \mathbb{E}[X^2] &= \mathbb{K}_2(X) + \mathbb{K}_1(X)^2, \\ \mathbb{E}[X^3] &= \mathbb{K}_3(X) + 3\,\mathbb{K}_2(X)\mathbb{K}_1(X) + \mathbb{K}_1(X)^3, \end{aligned}$$

which are equivalent to the formulas in Example 3.2.

Our goal is to discover a similar formula for monotone cumulants. It turns out that an ordered set partition provides a suitable framework. An **ordered set partition** of a finite set S is a sequence $\pi = (B_1, B_2, \ldots, B_k)$, where $\{B_1, B_2, \ldots, B_k\}$ is a set partition, i.e., $B_1, B_2, \ldots, B_k$ are nonempty disjoint subsets of S such that S is the union of them. We can also consider that π is the set partition $\{B_1, B_2, \ldots, B_k\}$ equipped with the linear order on its blocks $B_1 \le B_2 \le \cdots \le B_k$. In this way, an equivalent definition is that an ordered set partition is a pair $\pi = (\rho, \le)$ of a set partition $\rho \in \mathbf{P}(S)$ and a linear (or total) order on ρ. We set the notation $|\pi| := |\rho|$ that is the length of the sequence π, and $\overline{\pi} := \rho$. Let $\mathbf{OP}(S)$ be the set of the ordered set partitions of S, and in particular, let $\mathbf{OP}(n) := \mathbf{OP}([n])$. With a slight abuse of notation, we will write $B_i \in \pi$, which more precisely means $B_i \in \overline{\pi}$.

Let T be a totally ordered finite set. A set partition $\rho \in \mathbf{P}(T)$ is said to have a crossing if there are two blocks $B_1, B_2 \in \rho$ and elements $a, b \in B_1$ and $b, c \in B_2$ such that $a < c < b < d$. A set partition that has no crossings is called a **noncrossing set partition**.

We consider a partial order on each $\rho \in \mathbf{P}(T)$ defined by a covering relation. For nonempty subsets $B_1, B_2 \subseteq T$, we say B_1 covers B_2, denoted as $B_1 \preceq B_2$, if $\min B_1 \le i \le \max B_1$ for all $i \in B_2$. On a set partition of T, the relation $\preceq$ becomes a partial order.

Definition 3.23. Let T be a totally ordered finite set. An ordered set partition $\pi = (\rho, \le)$ of T is called a **monotone set partition** if

- ρ is a noncrossing set partition,
- if $B, B' \in \rho$ satisfies $B \preceq B'$ then $B \le B'$.

The set of monotone set partitions of T is denoted by $\mathbf{M}(T)$. For notational simplicity, we set $\mathbf{M}(n) := \mathbf{M}([n])$

The monotone set partitions $\mathbf{M}(T)$ can be generated from the following recursion.

Proposition 3.24. *Let T be a totally ordered finite set. There exists a canonical bijection*

$$\beta \colon \mathbf{M}(T) \to \bigcup_{I} \mathbf{M}(T \setminus I),$$

where I runs over the set of nonempty intervals of T, the complement $T \setminus I$ is endowed with the linear order induced by T, and $\mathbf{M}(\emptyset) := \{\emptyset\}$. The bijection is given by $\beta\colon (\rho, \le) \mapsto \rho \setminus \{B_{\max}\} \in \mathbf{M}(T \setminus B_{\max})$, where $B_{\max}$ is the largest block of ρ with respect to $\le$.

Proof. Observe first that $B_{\max}$ is a nonempty interval of T since $B_{\max}$ does not cover any other element and ρ is noncrossing. It is straightforward that $\rho \setminus \{B_{\max}\}$ is a monotone set partition of $T \setminus B_{\max}$. Conversely, the map $\bigcup_I \mathbf{M}(T \setminus I) \ni (I, \pi') = (I, (\rho', \le')) \mapsto (\rho' \cup \{I\}, \le)$ where $\le$ is the extension of $\le'$ such that I is larger than any block of ρ', also gives a monotone set partition of T, and this is the inverse map. ☐

Proposition 3.25. *Let T be a totally ordered finite set. The cardinality of $\mathbf{M}(T)$ is $\frac{(|T|+1)!}{2}$.*

Proof. We may assume that $T = [n]$. Let $t_n := |\mathbf{M}(n)|$, $n \in \mathbb{N}$, and $t_0 := 1$. The previous bijection yields $t_n = \sum_I t_{n-|I|}$. For each $1 \le k \le n$, there are $n - k + 1$ intervals I such that $|I| = k$. This yields

$$t_n = \sum_{k=1}^{n} (n-k+1) t_{n-k} = \sum_{p=0}^{n-1} (p+1) t_p, \qquad n \ge 1; \qquad t_0 = 1.$$

Computing $t_n - t_{n-1}$ yields $t_n = (n+1)t_{n-1}$, so an induction argument shows the desired formula $t_n = (n+1)!/2$ for $n \in \mathbb{N}$. ☐

Given a sequence $(\alpha_n)_{n\ge1}$ of complex numbers and $\pi = (B_1, B_2, \dots, B_k) \in \mathbf{OP}(S)$, we define

$$\alpha_\pi := \alpha_{|B_1|}\alpha_{|B_2|}\cdots\alpha_{|B_k|}.$$

Equivalently, we set $\alpha_\pi := \alpha_\rho$ for $\pi = (\rho, \le)$.

Theorem 3.26. *On any nc-probability space (A, φ) and for any $x \in A$, we have*

$$\varphi(x^n) = \sum_{\pi \in \mathbf{M}(n)} \frac{1}{|\pi|!} \kappa_\pi(x), \qquad n \in \mathbb{N}. \tag{3.13}$$

Remark 3.27. For the classical cumulants $(\mathbb{K}_n)_{n\ge1}$, the moment-cumulant formula (3.10) can be written in the equivalent form

$$\mathbb{E}[X^n] = \sum_{\pi \in \mathbf{OP}(n)} \frac{1}{|\pi|!} \mathbb{K}_\pi(X), \qquad n \in \mathbb{N},\ X \in L^{\infty-}, \tag{3.14}$$

since for each $\rho \in \mathbf{P}(n)$ there are $|\rho|!$ number of $\pi \in \mathbf{OP}(n)$ such that $\overline{\pi} = \rho$, corresponding to the permutations of the blocks. Comparing with formula (3.14) somehow justifies the naturality of the factor $1/|\pi|!$ in (3.13).

Proof of Theorem 3.26. Let $\mathrm{m}_n^{\triangleright}(t) = \mathrm{m}_n^{\triangleright}(t; x)$ be the polynomial in $t \in \mathbb{R}$ defined in Proposition 3.15. We prove a generalized formula

$$\mathrm{m}_n^{\triangleright}(t) = \sum_{\pi \in \mathbf{M}(n)} \frac{t^{|\pi|}}{|\pi|!} \kappa_\pi(x), \qquad n \in \mathbb{N}, \tag{3.15}$$

which coincides with (3.13) for $t = 1$. We extend formula (3.15) to $n = 0$ by interpreting $\mathbf{M}(0) := \{\emptyset\}$, $|\emptyset| := 0$, $0! := 1$ and $\kappa_\emptyset(x) := 1$, so that the following calculations make sense. Then formula (3.15) is obviously the case for $n = 0$. Suppose that the formula holds up to $n - 1$. Then, using the differential recursion in Proposition 3.15, we proceed as

$$\begin{aligned}
\mathrm{m}_n^{\triangleright}(t) &= \sum_{\ell=0}^{n-1} (\ell+1)\kappa_{n-\ell}(x) \int_0^t \mathrm{m}_\ell^{\triangleright}(s)\, \mathrm{d}s \\
&= \sum_{\ell=0}^{n-1} (\ell+1)\kappa_{n-\ell}(x) \sum_{\pi \in \mathbf{M}(\ell)} \frac{1}{|\pi|!} \kappa_\pi(x) \int_0^t s^{|\pi|}\, \mathrm{d}s \\
&= \sum_{\ell=0}^{n-1} (\ell+1) \sum_{\pi \in \mathbf{M}(\ell)} \frac{t^{|\pi|+1}}{(|\pi|+1)!} \kappa_{n-\ell}(x)\kappa_\pi(x) \\
&= \sum_{p=1}^{n} (n-p+1) \sum_{\pi \in \mathbf{M}(n-p)} \frac{t^{|\pi|+1}}{(|\pi|+1)!} \kappa_p(x)\kappa_\pi(x).
\end{aligned} \tag{3.16}$$

Since there are $n - p + 1$ intervals $I \subseteq [n]$ of size p, (3.16) can be written in the form

$$\sum_{\substack{\emptyset \ne I \subseteq [n] \\ \text{interval}}} \sum_{\pi \in \mathbf{M}([n]\setminus I)} \frac{t^{|\pi|+1}}{(|\pi|+1)!} \kappa_{|I|}(x)\kappa_\pi(x). \tag{3.17}$$

The last formula can be well described by the bijection β in Proposition 3.24: the ordered set partition $\sigma := \beta(I, \pi)$ runs over all elements of $\mathbf{M}(n)$ exactly once as (I, π) runs over the summation range of (3.17), and it holds that $|\pi| + 1 = |\sigma|$ and $\kappa_{|I|}(x)\kappa_\pi(x) = \kappa_\sigma(x)$. Therefore, the last expression (3.17) is exactly the desired (3.15). ☐

The monotone CLT says that the monotone cumulant sequence $(0, 1, 0, 0, 0, \dots)$ corresponds to the moment sequence (3.7) of the arcsine law. This fact and Theorem 3.26 yield the cardinality of the set of monotone pair partitions

$$\mathbf{M}_2(2n) := \{\pi \in \mathbf{M}(2n) : \text{every block of } \pi \text{ has cardinality two}\}.$$

Corollary 3.28. *The cardinality of $\mathbf{M}_2(2n)$ is $(2n-1)!!$.*

The above proof of the moment-cumulant formula in Theorem 3.26 does not clarify well why the monotone set partitions appear. In fact, monotone set partitions have a more intrinsic meaning: they naturally appear when characterizing a "tensor-like" factorization of mixed moments.

Definition 3.29. Let I be a totally ordered set and $i_1, i_2, \dots, i_n \in I$. Let $\ker(i_1, i_2, \dots, i_n)$ be the ordered set partition of $[n]$, called the **kernel**, defined as follows: let $A_j := \{p \in [n] : i_p = j\}$, $j \in I$, and we collect all the nonempty sets $A_{j_1}, A_{j_2}, \dots, A_{j_r}$, $j_1 < j_2 < \dots < j_r$, and define $\ker(i_1, i_2, \dots, i_n) := (A_{j_1}, A_{j_2}, \dots, A_{j_r})$.

Example 3.30. Let $I = \mathbb{N}$. Then $\ker(3, 5, 2, 1, 5, 3, 5)$ is given by $(\{4\}, \{3\}, \{1, 6\}, \{2, 5, 7\})$.

Proposition 3.31. *Let I be a totally ordered set and $i_1, i_2, \dots, i_n \in I$. Then $\ker(i_1, i_2, \dots, i_n) \in \mathbf{M}(n)$ if and only if the factorization*

$$\varphi(a_1 a_2 \cdots a_n) = \prod_{B \in \ker(i_1, i_2, \dots, i_n)} \varphi\left(\overrightarrow{\prod_{p \in B}} a_p\right) \tag{3.18}$$

holds for any random variables $a_1 \in A_{i_1}$, $a_2 \in A_{i_2}$, $\dots$, $a_n \in A_{i_n}$ and any monotonically independent subalgebras $(A_i)_{i \in I}$ in any nc-probability space (A, φ).

Remark 3.32. The factorization (3.18) is exactly the formula that always holds irrespective of $\ker(i_1, i_2, \dots, i_n)$ provided $(A_i)_{i \in I}$ were tensor independent. This proposition therefore characterizes the arrangements of random variables such that the factorization coincides with the case of tensor independence.

Proof. Let us check the statement through examples. In the following, $(A_i)_{i=1}^{\infty}$ are monotonically independent subalgebras in a nc-probability space (A, φ), $\boldsymbol{i} = (i_1, i_2, \dots, i_n) \in \mathbb{N}^n$ and $a_1 \in A_{i_1}$, $a_2 \in A_{i_2}$, $\dots$, $a_n \in A_{i_n}$.

Case 1: $\boldsymbol{i} = (i_1, i_2, \dots, i_8) = (2, 4, 4, 4, 3, 2, 1, 1)$. Then the kernel

$$\pi_1 := \ker(\boldsymbol{i}) = (\{7, 8\}, \{1, 6\}, \{5\}, \{2, 3, 4\})$$

is a monotone set partition. We first focus on the largest block $\{2, 3, 4\}$ and obtain

$$\varphi(a_1 a_2 \cdots a_8) = \varphi(a_1 (a_2 a_3 a_4) a_5 a_6 a_7 a_8) = \varphi(a_2 a_3 a_4) \varphi(a_1 a_5 a_6 a_7 a_8)$$

because $i_1 < i_2 = i_3 = i_4 > i_5$ and $a_2 a_3 a_4 \in A_{i_2}$. The remaining sequence $(i_1, i_5, i_6, i_7, i_8)$ associates the kernel ordered set partition $(\{7, 8\}, \{1, 6\}, \{5\})$, which is also a monotone set partition. Since now $i_1 < i_5 > i_6$ we have

$$\varphi(a_1 a_5 a_6 a_7 a_8) = \varphi(a_5) \varphi(a_1 a_6 a_7 a_8),$$

and finally we arrive at

$$\varphi(a_1 a_2 \cdots a_8) = \varphi(a_2 a_3 a_4) \varphi(a_5) \varphi(a_1 a_6) \varphi(a_7 a_8) = \prod_{B \in \pi_1} \varphi\left(\overrightarrow{\prod_{p \in B}} a_p\right).$$

In general, when the kernel is a monotone set partition, we can first factor out the expectation of elements corresponding to the largest block of the kernel, and then by Proposition 3.24 the remaining blocks still form a monotone set partition. Then we can repeat the same procedure to get the tensor-like factorization.

Case 2: $\boldsymbol{i} = (2, 1, 2, 1)$. The associated kernel $\pi_2 := \ker(\boldsymbol{i}) = (\{2, 4\}, \{1, 3\})$ is not a monotone set partition because $\overline{\pi_2}$ has a crossing. By the definition of monotone independence we get

$$\varphi(a_1 a_2 a_3 a_4) = \varphi(a_1) \varphi(a_2 a_3 a_4),$$

so that the block $\{1, 3\}$ "splits" into the singletons $\{1\}$ and $\{3\}$. This shows the tensor-like factorization does not hold.

Case 3: $\boldsymbol{i} = (2, 1, 1, 2)$. Then $\pi_3 := \ker(\boldsymbol{i}) = (\{2, 3\}, \{1, 4\})$. In this case $\overline{\pi_3}$ is noncrossing but the total order on $\overline{\pi_3}$ is not compatible with the covering relation, so that π_3 is not a monotone set partition. It holds that

$$\varphi(a_1 a_2 a_3 a_4) = \varphi(a_1) \varphi(a_2 a_3 a_4),$$

so that the block $\{1, 4\}$ again splits.

In general, as soon as $\ker(\boldsymbol{i})$ is not a monotone set partition, there always exists a block that splits, so that the tensor-like factorization fails. □

3.6. **Notes.** The monotone cumulants were defined by Hasebe and Saigo [87], and we basically followed this original paper, providing more detailed arguments. More general multivariate monotone cumulants are introduced in [86]. The proof of Theorem 3.26 followed [86]. The original definition of monotone cumulants was inspired by "umbral calculus" in combinatorics, in which "umbrae" correspond to iid copies of random variables, and a "dot operation" corresponds to the sum of iid random variables. The definition of monotone cumulants builds upon a Lie-theoretic approach, which was already exploited by Voiculescu in the definition of free cumulants [150]. The Lie-theoretic aspect of cumulants has been further pursued in the literature by the use of Hopf algebras, see e.g. [9, 60, 85, 111].

The proof of Theorem 3.18 (the monotone CLT) and Theorem 3.19 (monotone Poisson's law of small numbers) more or less followed the lines of [87], being different from the proof of the original article [121]. Hora, Obata [90, Theorem 8.23] and Saigo [135] proved the monotone CLT allowing some non-identically distributed random variables. Wang analytically proved the monotone CLT only by assuming the existence of finite second moment [153]. Arizmendi, Salazar and Wang provided a Berry-Esseen type result [15]. Other limit theorems are studied in the literature: Wang and Wendler showed a law of large numbers [154] using a martingale technique; Wang obtained a limit theorem of stable type [152]; Anshelevich and Williams established a rather general limit theorem converging to monotonically infinitely divisible distributions [7]; Franz, Hasebe and Schleißinger studied monotone convolutions of infinitesimal triangular arrays that allow non-identical probability measures [70]. These results can be seen as nontrivial limit theorems for iterated compositions of holomorphic self-maps and some of them have connections to ergodic theory.

The monotone set partitions first appeared in [122] in the form of Proposition 3.31. Lie-theoretic approaches can make it clearer how Proposition 3.31 leads to the appearance of monotone set partitions in the moment-cumulant formula in Theorem 3.26, see [85, 111].

4. Cauchy transform

In this section we collect results on the Cauchy transform of probability measures and its relatives. Using these results we extend monotone convolutions to probability measures with unbounded support and analyze them in later sections.

4.1. **Measures.** As different sources use different terminology in measure theory, we start by specifying our conventions. Let X be a topological space and $\mathcal{B}(X)$ be the set of Borel subsets of X, i.e., $\mathcal{B}(X) \subseteq 2^X$ is the smallest σ-field that contains all open subsets of X. A **Borel measure** is a function $\mu\colon \mathcal{B}(X) \to [0, +\infty]$ such that

- $\mu(\emptyset) = 0$,
- $\mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n)$ whenever $B_1, B_2, B_3, \dots$ are disjoint Borel subsets of X.

A Borel measure μ on X is called

- a **probability measure** if the total mass $\mu(X)$ is one;
- **finite** if the total mass is finite;
- **locally finite** if every point of X has an open neighborhood with finite mass; and
- **atomless** if $\mu(\{x\}) = 0$ for all $x \in X$.

Related to the last condition above, an element $x \in X$ is called an **atom** of μ if $\mu(\{x\}) \in (0, +\infty]$.

We say that a Borel measure μ is **supported on** B if $\mu(X \setminus B) = 0$. The **support** of μ is the smallest closed subset B of X such that $\mu(X \setminus B) = 0$, and the support is denoted as $\mathrm{supp}(\mu)$ if it exists. The support of a Borel measure always exists if X is a separable metric space [59, Problem 3, Section 7.1].

A **complex Borel measure** is a function of the form $\mu = \mu_1 - \mu_2 + \mathrm{i}(\mu_3 - \mu_4)$, where μ_k are Borel measures. The domain of μ is $\{B \in \mathcal{B}(X) : \mu_k(B) < +\infty,\ k = 1, 2, 3, 4\}$.

4.1.1. *Handling σ-fields.* The definition of $\mathcal{B}(X)$ is implicit in that no necessary and sufficient condition is given for subsets of X to be Borel. In measure theory and probability theory, one often sees such sets, e.g. Lebesgue measurable sets or sets in product σ-fields. To handle such sets, a standard technique is to consider classes of sets rather than individual sets. The following classes are widely used.

Definition 4.1. Let Ω be a set and $\mathcal{F} \subseteq 2^\Omega$.

(i) $\mathcal{F}$ is called a **π-system** if $A \cap B \in \mathcal{F}$ for all $A, B \in \mathcal{F}$.

(ii) $\mathcal{F}$ is called a **λ-system** if the following conditions hold:
  - $\Omega \in \mathcal{F}$;
  - if $A, B \in \mathcal{F}$ and $A \subseteq B$ then $B \setminus A \in \mathcal{F}$;
  - if $A_n \in \mathcal{F}$ $(n \in \mathbb{N})$ and $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$ then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

(iii) $\mathcal{F}$ is called an **algebra** if the following conditions hold:
  - $\Omega \in \mathcal{F}$;
  - if $A, B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$;
  - if $A \in \mathcal{F}$ then $\Omega \setminus A \in \mathcal{F}$.

(iv) $\mathcal{F}$ is called a **monotone class** if the following conditions hold:
  - if $A_n \in \mathcal{F}$ $(n \in \mathbb{N})$ and $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$ then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$;
  - if $A_n \in \mathcal{F}$ $(n \in \mathbb{N})$ and $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$ then $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

Note that any algebra is a π-system and any λ-system is a monotone class.

The following π-λ theorem and monotone class theorem are standard methods to extend a statement about a class $\mathcal{C}$ of sets to a statement about the σ-field $\sigma(\mathcal{C})$ generated by the class $\mathcal{C}$; the reader is referred to [97, Theorem 1.19] and [59, Theorem 4.4.2] for the proofs, respectively.

Theorem 4.2 (π-λ theorem)**.** *Let Ω be a set. Let $\mathcal{P}$ be a π-system and $\mathcal{L}$ be a λ-system of subsets of Ω such that $\mathcal{P} \subseteq \mathcal{L}$. Then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.*

Theorem 4.3 (Monotone class theorem)**.** *Let Ω be a set. Let $\mathcal{A}$ be an algebra and $\mathcal{M}$ be a monotone class of subsets of Ω such that $\mathcal{A} \subseteq \mathcal{M}$. Then $\sigma(\mathcal{A}) \subseteq \mathcal{M}$.*

A typical application of these theorems is the following.

Proposition 4.4. *Let τ_1, τ_2 be locally finite Borel measures on $\mathbb{R}$ such that $\tau_1(I) = \tau_2(I)$ for all open intervals I of finite length. Then $\tau_1 = \tau_2$.*

Proof. The goal is to show $\tau_1(B) = \tau_2(B)$ for all $B \in \mathcal{B}(\mathbb{R})$. Since

$$\tau_i(B) = \lim_{N\to\infty} \tau_i(B \cap (-N, N)), \qquad i = 1, 2,$$

it suffices to show $\tau_1(B) = \tau_2(B)$ for bounded Borel subsets B. We therefore fix $N \in \mathbb{N}$. Let us consider the set $\mathcal{P} \subseteq 2^{(-N,N)}$ consisting of the empty set and the open subintervals of $(-N, N)$, and

$$\mathcal{L} := \{B \in \mathcal{B}((-N, N)) : \tau_1(B) = \tau_2(B)\}.$$

We can see that $\mathcal{P}$ is a π-system, $\mathcal{L}$ is a λ-system and, by assumption, $\mathcal{P}$ is contained in $\mathcal{L}$. By the π-λ theorem, $\sigma(\mathcal{P})$ is contained in $\mathcal{L}$, the former of which is known to be equal to $\mathcal{B}((-N, N))$. □

Remark 4.5. One could also use the monotone class theorem. Observe first that the above $\mathcal{L}$ is also a monotone class. Instead of $\mathcal{P}$ one could consider the algebra $\mathcal{A} \subseteq 2^{(-N,N)}$ consisting of the empty set and finite disjoint unions of the intervals of the forms $(a_i, b_i]$ $(-N \le a_i < b_i \le N)$; note that $(a_i, N]$ is to be interpreted as (a_i, N). By taking limits one would see that each $(a_i, b_i]$ belongs to $\mathcal{L}$, and therefore $\mathcal{A} \subseteq \mathcal{L}$. By the monotone class theorem, $\mathcal{L}$ would contain $\sigma(\mathcal{A}) = \mathcal{B}((-N, N))$.

Sometimes, finding an appropriate algebra is harder than finding a π-system, and therefore π-λ theorem is more useful. Later in Theorem 6.11 we also see the opposite situation where the second condition of the λ-system is hard to check, and thus monotone class theorem is more useful.

4.1.2. *Weak convergence of measures.* There are several notions of convergence for measures. Here we consider weak convergence and its properties. In Appendix A, a relation of weak convergence to convergence of moments is discussed.

Definition 4.6. Let X be a topological space. The set of the finite Borel measures on X is denoted by $\mathscr{M}_{\mathrm{fin}}(X)$. A sequence $(\tau_n)_{n\ge1}$ in $\mathscr{M}_{\mathrm{fin}}(X)$ is said to **converge weakly** to $\tau \in \mathscr{M}_{\mathrm{fin}}(X)$ if for any bounded continuous function $f : X \to \mathbb{R}$ one has

$$\lim_{n\to\infty} \int_X f(x)\tau_n(\mathrm{d}x) = \int_X f(x)\tau(\mathrm{d}x). \tag{4.1}$$

Sometimes, it is convenient to take a smaller set of test functions. For the proof of the following result, the reader is referred to [97, Theorem 13.16].

Proposition 4.7. *Let X be a locally compact, complete separable metric space and $\tau_n, \tau \in \mathscr{M}_{\mathrm{fin}}(X)$, $n \in \mathbb{N}$. Then τ_n converges weakly to τ if and only if $\lim_{n\to\infty} \tau_n(X) = \tau(X)$ and (4.1) holds for all continuous functions f with compact support.*

The following is a somewhat trivial reformulation of weak convergence. This is, however, useful when combined with relative compactness, see Remark 4.13.

Lemma 4.8. *Let X be a topological space. Let $\tau, \tau_n \in \mathscr{M}_{\mathrm{fin}}(X)$ $(n \in \mathbb{N})$. Then the weak convergence $\tau = \lim_{n\to\infty} \tau_n$ holds if and only if any subsequence of $(\tau_n)_{n\ge1}$ has a further subsequence that converges weakly to τ.*

Proof. The "only if" part is obvious. For the "if" part, suppose to the contrary that τ_n does not converge to τ; then there exist a bounded continuous function $f : X \to \mathbb{R}$, $\varepsilon > 0$ and a subsequence $(\tau_{n(j)})_{j\ge1}$ such that $|\int f d\tau_{n(j)} - \int f d\tau| \ge \varepsilon$ for all j. This contradicts the assumption that $(\tau_{n(j)})$ has a further subsequence that converges to τ. □

Definition 4.9. Let X be a topological space and $\mathscr{M} \subseteq \mathscr{M}_{\mathrm{fin}}(X)$.

(i) $\mathscr{M}$ is said to be **tight** if for any $\varepsilon > 0$ there exists a compact subset $K \subseteq X$ such that $\mu(X \setminus K) < \varepsilon$ for all $\mu \in \mathscr{M}$.

(ii) $\mathscr{M}$ is said to be **relatively compact** (with respect to the weak convergence) if any sequence in $\mathscr{M}$ has a subsequence that is weakly convergent in $\mathscr{M}_{\mathrm{fin}}(X)$.

Remark 4.10. The above definition of relative compactness coincides with the standard definition in topology when $\mathscr{M}_{\mathrm{fin}}(X)$ is metrizable, which is the case if X is separable; see e.g. [36, Theorem 5, Appendix III].

Theorem 4.11 (Prokhorov)**.** *Let X be a complete separable metric space and $\mathscr{M}$ be a subset of $\mathscr{M}_{\mathrm{fin}}(X)$. Then $\mathscr{M}$ is relatively compact if and only if $\mathscr{M}$ is tight and $\{\tau(X) : \tau \in \mathscr{M}\} \subseteq [0, +\infty)$ is bounded.*

Proof. For the case of probability measures, the reader is referred to [36, Theorems 6.1 and 6.2]. For finite Borel measures, dividing the measures by their total masses reduces the problem to probability measures. □

Remark 4.12. Actually we will use Theorem 4.11 only for $X = \mathbb{R}$ or circles in the Riemann sphere $\mathbb{C} \cup \{\infty\}$. Then this theorem is a direct consequence of Helly's selection theorem, see e.g. [97, Theorem 13.33]. If X is a compact space, the whole set $\mathscr{M}_{\mathrm{fin}}(X)$ is automatically tight and so relative compactness is equivalent to the uniform boundedness of total masses.

Remark 4.13 (Compactness argument)*.* To show the weak convergence of a sequence of finite Borel measures $(\tau_n)_{n\ge1}$, a useful method is a **compactness argument**. One first shows that $(\tau_n)_{n\ge1}$ is tight and uniformly bounded (i.e., relatively compact); this is of course easier when the underlying space X is compact. Then any subsequence $(\tau_{n(j)})_{j\ge1}$ of $(\tau_n)_{n\ge1}$ has a further subsequence $(\tau_{n(j(k))})_{k\ge1}$ weakly converging to a finite Borel measure τ. If one can somehow show that the limit τ does not depend on the choice of $(\tau_{n(j)})_{j\ge1}$ or $(\tau_{n(j(k))})_{k\ge1}$, Lemma 4.8 implies that the whole sequence $(\tau_n)_{n\ge1}$ converges to τ.

4.2. **Holomorphic functions.** In this section, we collect technical but preparatory results in complex analysis. Let $\mathbb{D}$ denote the unit disk $\{z \in \mathbb{C} : |z| < 1\}$ and $\mathbb{C}^+$ the complex upper half-plane $\{z \in \mathbb{C} : \Im(z) > 0\}$. We consider $\widehat{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ as a compact subset of the Riemann sphere $\mathbb{C} \cup \{\infty\}$.

4.2.1. *Interchange of integration and differentiation.* Here we note a useful criterion that allows us to differentiate a holomorphic function under the integral sign. This fact will be used below without being mentioned.

Proposition 4.14. *Let $(T, \mathcal{F}, \mu)$ be a measure space. Let D be an open subset of $\mathbb{C}$. Let $f\colon D\times T \to \mathbb{C}$ be a function such that*

- *for a.e. $t \in T$ the function $f(\cdot,t)$ is holomorphic on D,*
- *for each $z \in D$ the function $f(z,\cdot)$ is μ-integrable,*
- *there is a μ-integrable function $g\colon T \to [0,+\infty)$ such that $|f(z,t)| \le g(t)$ for all $z \in D$ and $t \in T$.*

Then the function $F(z) := \int_T f(z,t)\,\mu(\mathrm{d}t)$ is holomorphic on D and $F^{(n)}(z) = \int_T \partial_z^n f(z,t)\,\mu(\mathrm{d}t)$ for all $n \in \mathbb{N}$. Note that the assumptions above imply that $\partial_z^n f(z,\cdot)$ is μ-integrable for any $z \in D$ and $n \in \mathbb{N}_0$.

Proof. Let C be the circle centered at $z_0 \in D$ with radius $r > 0$, sufficiently small so that C and its interior are contained in D. By Cauchy's integral formula, for μ-a.e. t, we have

$$\partial_z f(z,t) = \frac{1}{2\pi\mathrm{i}}\int_C \frac{f(w,t)}{(w-z)^2}\,\mathrm{d}w.$$

This yields the estimate

$$|\partial_z f(z,t)| \le \frac{1}{2\pi}\int_C \frac{g(t)}{|w-z|^2}\,|\mathrm{d}w| \le \frac{4g(t)}{r}, \qquad |z-z_0| < \frac{r}{2}. \tag{4.2}$$

Writing $f(z,t) = u(x,y,t) + \mathrm{i}v(x,y,t)$ and $F(z) = U(x,y) + \mathrm{i}V(x,y)$ with notation $z = x+\mathrm{i}y$, the above estimate implies that the four functions $|\partial_x u|$, $|\partial_y u|$, $|\partial_x v|$, $|\partial_y v|$ are all bounded by $4g(t)/r$. Therefore, the usual criterion for the interchange of differentiation and integration yields that $U(x,y), V(x,y)$ are differentiable under the integral sign with respect to both x and y. Also we can check that U,V are C^1 functions by the dominated convergence theorem. This argument implies that the C^1 functions U,V satisfy the Cauchy–Riemann equations, and hence F is holomorphic and the desired formula $F'(z) = \int_T \partial_z f(z,t)\,\mu(\mathrm{d}t)$ holds.

As we have established (4.2), the function $\partial_z f$ also satisfies the assumptions of the proposition in a neighborhood of z_0, so that we can obtain the result for the second derivative. Repeating the above arguments, we obtain the desired formula for higher-order derivatives. □

4.2.2. *Boundary behavior.* The study of boundary behavior of holomorphic functions is an important theme below. A basic notion in this respect is the following.

Definition 4.15. For a function $f\colon \mathbb{C}^+ \to \mathbb{C}$ we say that f has a **nontangential limit** $\zeta \in \mathbb{C}\cup\{\infty\}$ at ∞ if for any $\gamma > 0$ we have

$$\lim_{\substack{z\in\nabla_\gamma\\ |z|\to\infty}} f(z) = \zeta, \tag{4.3}$$

where ∇_γ is the sector domain

$$\nabla_\gamma := \{z \in \mathbb{C}^+ : \gamma|\Re(z)| < \Im(z)\}.$$

The nontangential limit of f at ∞ is written as $\sphericalangle\lim_{z\to\infty} f(z)$ if it exists.

There is a conformal bijection of $\mathbb{C}^+$ that maps ∞ to $a \in \mathbb{R}$. This allows us to define a nontangential limit at a: f has a nontangential limit $\zeta \in \mathbb{C}\cup\{\infty\}$ at $a \in \mathbb{R}$ if for any $\gamma > 0$

$$\lim_{\substack{z\in a+\nabla_\gamma\\ z\to a}} f(z) = \zeta,$$

and we write $\zeta = \sphericalangle\lim_{z\to a} f(z)$.

In general, even if (4.3) exists for some $\gamma > 0$, the limit might not exist for smaller γ's. A remarkable fact is that such never happens for a large class of holomorphic functions: if (4.3) exists for some $\gamma > 0$, the limit exists for any $\gamma > 0$. This is a consequence of Lindelöf's theorem. Although this theorem is not essential below, we quote it here as it helps to better understand some results. For the proof we refer the reader to [40, Theorem 1.5.7] or [45, Theorem 2.3]. A stronger version can be found in [45, Theorem 2.20].

Theorem 4.16 (Lindelöf). *Let $N\colon \mathbb{C}^+ \to \mathbb{C}^+\cup\mathbb{R}$ be a holomorphic function. If there exists a continuous map $\gamma\colon [0,1) \to \mathbb{C}^+$ such that $\lim_{t\to1}\gamma(t) = \infty$ and $\zeta := \lim_{t\to1} N(\gamma(t)) \in \mathbb{C}^+\cup\widehat{\mathbb{R}}$ exists, then the nontangential limit of N at ∞ exists and equals ζ.*

4.2.3. *Convergence of holomorphic functions.* We will consider the locally uniform convergence for holomorphic functions. There are two equivalent formulations that could be different for general continuous functions in topological spaces.

Definition 4.17. Let X be a topological space and $f, f_n \in C(X)$, $n \in \mathbb{N}$.

(i) $(f_n)_{n\ge1}$ is said to **converge to f locally uniformly** if each point of X has a neighborhood on which f_n converges to f uniformly.

(ii) $(f_n)_{n\ge1}$ is said to **converge to f uniformly on compacta** if f_n converges to f uniformly on each compact subset of X.

If every point of X has a compact neighborhood then these two notions coincide.

The idea of Lemma 4.8 also works for the uniform convergence on compacta.

Lemma 4.18. *Let X be a topological space and let $f, f_n \in C(X)$, $n \in \mathbb{N}$. Then f_n converges to f uniformly on compacta if and only if any subsequence of $(f_n)_{n\ge1}$ has a further subsequence that converges to f uniformly on compacta.*

The relative compactness for functions is defined similarly to measures.

Definition 4.19. Let X be a topological space. A family $\mathscr{F} \subseteq C(X)$ is called **relatively compact** (with respect to uniform convergence on compacta), or normal, if any sequence in $\mathscr{F}$ has a subsequence that converges uniformly on compacta to a continuous function on X. Note that the limit function is not required to belong to $\mathscr{F}$.

Remark 4.20 (cf. Remark 4.10). The definition of relative compactness above coincides with the notion in topology if convergence on compacta is induced by a metric on the space $C(X)$. This is the case e.g. if X is of the form $X = \bigcup_{n\in\mathbb{N}} K_n$, where K_n are compact subsets such that any compact subset of X is contained in some K_n. Indeed, in this case

$$\rho(f,g) := \sum_{n=1}^{\infty} 2^{-n} \left(1 \wedge \sup_{x\in K_n} |f(x) - g(x)|\right)$$

provides a distance on $C(X)$ such that $\rho(f_n, f) \to 0$ if and only if $f_n \to f$ uniformly on compacta. For example, if X is an open subset of $\mathbb{C}$ then $K_n := \{z \in X : d(z, \mathbb{C} \setminus X) \geq 1/n,\ |z| \leq n\}$ meets this condition, where d is the Euclidean distance.

In the compactness argument, a characterization of relative compactness is a key ingredient. For holomorphic functions, this is given by Montel's theorem. The reader is referred to [133] for the proof.

Theorem 4.21 (Montel)**.** *Let $D \subseteq \mathbb{C}$ be a domain. Let $\mathscr{F}$ be a family of holomorphic functions from D to $\mathbb{C}$. Then $\mathscr{F}$ is relatively compact if and only if $\sup_{z\in K,\, f\in\mathscr{F}} |f(z)| < +\infty$ for each compact subset $K \subseteq D$. Note that a locally uniform limit of holomorphic functions is always holomorphic by Weierstrass' theorem.*

On the other hand, the identity theorem provides a sufficient condition for the uniqueness of limit points. Combining Montel's theorem, the identity theorem and Lemma 4.18 yields the following useful criterion for the locally uniform convergence.

Theorem 4.22 (Vitali)**.** *Let $D \subseteq \mathbb{C}$ be a domain. Let $f_n\colon D \to \mathbb{C}$ $(n = 1, 2, 3, \dots)$ be holomorphic functions that are uniformly bounded on each compact subset of D. Suppose that there is a sequence of distinct points $(z_k)_{k\geq1} \subseteq D$ such that $\lim_{k\to\infty} z_k \in D$ and $\lim_{n\to\infty} f_n(z_k)$ exists in $\mathbb{C}$ for all $k \in \mathbb{N}$. Then f_n converges to a holomorphic function locally uniformly on D.*

Proof. By Montel's theorem, (f_n) has a subsequence that converges locally uniformly to a holomorphic function f on D. This is a candidate for the limit.

Let us apply Lemma 4.18. Take any subsequence $(f_{n(k)})_{k\geq1}$. Again by Montel's theorem, it has a further subsequence $(f_{n(k(j))})_{j\geq1}$ that converges to a holomorphic function g locally uniformly on D. By the assumption that $\lim_{n\to\infty} f_n(z_k) \in \mathbb{C}$ exists, we conclude that $f(z_k) = g(z_k)$ for all $k \in \mathbb{N}$. The identity theorem forces f and g to coincide. Therefore, $(f_{n(k(j))})_{j\geq1}$ converges to f locally uniformly. Lemma 4.18 yields that the whole sequence $(f_n)_{n\geq1}$ converges to f on D. □

4.3. **Nevanlinna functions.** A holomorphic function

$$N\colon \mathbb{C}^+ \to \mathbb{C}^+ \cup \mathbb{R}$$

is called a **Nevanlinna function**. Nevanlinna functions play central roles below because many transforms of probability measures are holomorphic functions taking values in half-planes. We collect various properties of Nevanlinna functions. We first demonstrate an integral formula for Nevanlinna functions.

Upon fixing a conformal bijection $L\colon \mathbb{C}^+ \to \mathbb{D}$, the function $-\mathrm{i}N \circ L^{-1}$ is a holomorphic function on $\mathbb{D}$ taking values with nonnegative real part, called a **Herglotz function**. Sometimes Herglotz functions make arguments clearer so we work with them instead.

Lemma 4.23 (Poisson integral formula)**.** *Let $f : \mathbb{D} \to \mathbb{C}$ be a holomorphic function. For every $R \in (0, 1)$ it holds that*

$$f(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Re\left(\frac{Re^{\mathrm{i}\phi} + z}{Re^{\mathrm{i}\phi} - z}\right) f(Re^{\mathrm{i}\phi})\, \mathrm{d}\phi, \qquad |z| < R. \tag{4.4}$$

Proof. Let $z = re^{\mathrm{i}\theta}$ with $0 \leq r < R$. By Cauchy's integral formula

$$f(z) = \frac{1}{2\pi\mathrm{i}} \int_{|w|=R} \frac{f(w)}{w - z}\, \mathrm{d}w. \tag{4.5}$$

Let $z^* := \frac{R^2}{r} e^{\mathrm{i}\theta}$ called the reflection of z with respect to the circle $\{w : |w| = R\}$. Since $|z^*| > R$, Cauchy's integral theorem yields

$$0 = \frac{1}{2\pi\mathrm{i}} \int_{|w|=R} \frac{f(w)}{w - z^*}\, \mathrm{d}w. \tag{4.6}$$

Combining (4.5) and (4.6) gives

$$f(z) = \frac{1}{2\pi\mathrm{i}} \int_{|w|=R} f(w) \left(\frac{1}{w - z} - \frac{1}{w - z^*}\right) \mathrm{d}w.$$

With notation $w = Re^{\mathrm{i}\phi}$ we obtain

$$\frac{1}{w - z} - \frac{1}{w - z^*} = \frac{1}{Re^{\mathrm{i}\phi}} \Re\left(\frac{Re^{\mathrm{i}\phi} + z}{Re^{\mathrm{i}\phi} - z}\right).$$

The conclusion follows by observing $\mathrm{d}w = \mathrm{i}Re^{\mathrm{i}\phi}\, \mathrm{d}\phi$. □

Proposition 4.24. *For a Herglotz function f, there exist $b \in \mathbb{R}$ and a finite Borel measure σ on $\partial\mathbb{D}$ such that*

$$f(z) = \mathrm{i}b + \int_{\partial\mathbb{D}} \frac{\zeta + z}{\zeta - z}\, \sigma(\mathrm{d}\zeta), \qquad z \in \mathbb{D}. \tag{4.7}$$

Proof. Let $0 < R < 1$ and g be a holomorphic function defined by

$$g(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{Re^{\mathrm{i}\phi}+z}{Re^{\mathrm{i}\phi}-z}\,\Re[f(Re^{\mathrm{i}\phi})]\,\mathrm{d}\phi, \qquad |z| < R.$$

Since f has a representation in Lemma 4.23 we have $\Re[g(z)] = \Re[f(z)]$ for $|z| < R$. It is a well known consequence of the Cauchy–Riemann relations that a holomorphic function on a domain with a constant real part must be constant, which implies in our situation that $f(z) = g(z) + \mathrm{i}b$ for some constant $b \in \mathbb{R}$. Therefore,

$$f(Rz) = \mathrm{i}b + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{\mathrm{i}\phi}+z}{e^{\mathrm{i}\phi}-z}\,\Re[f(Re^{\mathrm{i}\phi})]\,\mathrm{d}\phi = \mathrm{i}b + \int_{\partial\mathbb{D}} \frac{\zeta+z}{\zeta-z}\,\sigma_R(\mathrm{d}\zeta), \qquad z \in \mathbb{D}, \tag{4.8}$$

where σ_R is the finite Borel measure on the unit circle $\partial\mathbb{D}$ defined by $\sigma_R(\mathrm{d}\zeta) := \frac{1}{2\pi}\Re[f(Re^{\mathrm{i}\phi})]\,\mathrm{d}\phi$, $\zeta = e^{\mathrm{i}\phi}$. Selecting $z = 0$ in (4.8) we obtain $\sigma_R(\partial\mathbb{D}) = \Re[f(0)]$, which implies that the family $\{\sigma_R(\partial\mathbb{D}) \mid 0 < R < 1\} \subseteq [0, +\infty)$ is bounded. Since $\partial\mathbb{D}$ is compact, the family $\{\sigma_R \mid 0 < R < 1\}$ is tight. By Theorem 4.11, there exists a sequence $R_n \uparrow 1$ and a finite Borel measure σ such that $\sigma = \lim_{n\to\infty} \sigma_{R_n}$ weakly. Setting $R = R_n$ in (4.8) and letting $n \to \infty$ amounts to the desired (4.7). ◻

Remark 4.25. The number $b \in \mathbb{R}$ and the finite Borel measure σ are unique. We will prove this in the next theorem for the equivalent setting of Nevanlinna functions.

Theorem 4.26 (Nevanlinna formula)**.** *For a Nevanlinna function N, there exist $a \geq 0$, $b \in \mathbb{R}$ and a finite Borel measure τ on $\mathbb{R}$ such that*

$$N(z) = az - b + \int_{\mathbb{R}} \frac{1+tz}{t-z}\,\tau(\mathrm{d}t) = -b + \int_{\widehat{\mathbb{R}}} \frac{1+tz}{t-z}\,\widehat{\tau}(\mathrm{d}t), \tag{4.9}$$

where $\widehat{\tau}$ is the finite Borel measure on $\widehat{\mathbb{R}}$ defined by $\widehat{\tau}\restriction_{\mathbb{R}} = \tau$ and $\widehat{\tau}(\{\infty\}) = a$, and $(1+\infty z)/(\infty - z) := z$. The triplet (a, b, τ) is uniquely determined as follows.

(i) $a = \sphericalangle \lim_{z\to\infty} \dfrac{N(z)}{z}$.

(ii) $b = -\Re[N(\mathrm{i})]$.

(iii) *The Borel measure $\rho(\mathrm{d}t) = (1+t^2)\tau(\mathrm{d}t)$ satisfies for each $-\infty < \alpha < \beta < \infty$*

$$\rho((\alpha,\beta)) + \frac{1}{2}(\rho(\{\alpha\}) + \rho(\{\beta\})) = \frac{1}{\pi}\lim_{y\to 0^+}\int_{\alpha}^{\beta} \Im[N(x+\mathrm{i}y)]\,\mathrm{d}x, \tag{4.10}$$

$$\rho(\{\alpha\}) = \lim_{y\to 0^+} (-\mathrm{i}y)N(\alpha+\mathrm{i}y), \tag{4.11}$$

*which will be referred to as the **Stieltjes inversion formula**.*

Remark 4.27. A further study of the behavior of $\Im[N(x+\mathrm{i}y)]$ as $y \to 0^+$ would provide more information about ρ, e.g., supporting sets for the absolutely continuous part and singular part of ρ, see [139, Appendix F].

Proof. The conformal bijection $L\colon \mathbb{C}^+ \to \mathbb{D}$ defined by

$$L(z) = \frac{\mathrm{i}z+1}{z+\mathrm{i}}$$

has the compositional inverse

$$L^{-1}(z) = \frac{-\mathrm{i}z+1}{z-\mathrm{i}}$$

and extends to a homeomorphism from $\mathbb{C}^+ \cup \widehat{\mathbb{R}}$ onto $\overline{\mathbb{D}}$. By Proposition 4.24 the function $-\mathrm{i}N \circ L^{-1}$ has a representation

$$-\mathrm{i}N(L^{-1}(w)) = \mathrm{i}b + \int_{-\pi}^{\pi} \frac{e^{\mathrm{i}\phi}+w}{e^{\mathrm{i}\phi}-w}\,\sigma(\mathrm{d}\phi) = \mathrm{i}b + \int_{\widehat{\mathbb{R}}} \frac{L(t)+w}{L(t)-w}\,\widehat{\tau}(\mathrm{d}t), \qquad w \in \mathbb{D}, \tag{4.12}$$

where $\widehat{\tau} = \sigma \circ (L\restriction_{\widehat{\mathbb{R}}})$ is the measure pushforwarded by the map $L^{-1}\restriction_{\partial\mathbb{D}}$. With the notation $w = L(z)$, $z \in \mathbb{C}^+$, it holds that

$$\frac{L(t)+L(z)}{L(t)-L(z)} = -\frac{\mathrm{i}(1+tz)}{t-z},$$

which transforms (4.12) into the desired (4.9).

(ii) is immediate.

(i) It suffices to show $\sphericalangle \lim_{z\to\infty} R(z)/z = 0$, where $R(z) := \int_{\mathbb{R}} \frac{1+tz}{t-z}\,\tau(\mathrm{d}t)$. Let $t \in \mathbb{R}$, $\gamma > 0$, $z = x+\mathrm{i}y \in \nabla_\gamma$ with $y \geq 1$. Then

$$\left|\frac{1+tz}{z(t-z)}\right| = \left|\frac{1}{z(t-z)} + \frac{z}{t-z} + 1\right| \leq 1 + \left|\frac{z}{t-z}\right|\left(1+\frac{1}{|z|^2}\right) \leq 1 + 2\left|\frac{z}{t-z}\right|.$$

Since

$$\left|\frac{z}{z-t}\right| = \sqrt{\frac{x^2+y^2}{(x-t)^2+y^2}} \leq \sqrt{\frac{\gamma^{-2}y^2+y^2}{y^2}} = \sqrt{1+\gamma^{-2}}, \tag{4.13}$$

the function $|(1+tz)/[z(t-z)]|$ can be uniformly bounded by a constant independent of (t, z). By the dominated convergence theorem, $\lim_{z\to\infty,\, z\in\nabla_\gamma} R(z)/z = 0$.

(iii) First observe that

$$\frac{1+tz}{t-z} = \left(\frac{1}{t-z} - \frac{t}{1+t^2}\right)(1+t^2), \tag{4.14}$$

which is sometimes useful. Now this formula immediately implies

$$\begin{aligned}\Im[N(x+\mathrm{i}y)] &= ay + \int_{\mathbb{R}} \frac{y}{(x-t)^2+y^2}\,\rho(\mathrm{d}t) \\ &= ay + \underbrace{\int_{(\alpha-1,\beta+1)} \frac{y}{(x-t)^2+y^2}\,\rho(\mathrm{d}t)}_{=:\,I_1(x,y)} + \underbrace{\int_{(-\infty,\alpha-1]\cup[\beta+1,\infty)} \frac{y}{(x-t)^2+y^2}\,\rho(\mathrm{d}t)}_{=:\,I_2(x,y)}.\end{aligned}$$

Since I_2 is continuous on $[\alpha,\beta]\times[0,+\infty)$ and vanishes on $y=0$, we have $\lim_{y\to 0^+}\int_{(\alpha,\beta)} I_2(x,y)\,\mathrm{d}x = 0$. Concerning I_1, we use Tonelli's theorem to interchange the integrals

$$\begin{aligned}\lim_{y\to0^+}\int_{(\alpha,\beta)} I_1(x,y)\,\mathrm{d}x &= \lim_{y\to0^+}\int_{(\alpha-1,\beta+1)}\rho(\mathrm{d}t)\int_{(\alpha,\beta)}\frac{y}{(x-t)^2+y^2}\,\mathrm{d}x \\ &= \lim_{y\to0^+}\int_{(\alpha-1,\beta+1)}\left(\arctan\frac{\beta-t}{y} - \arctan\frac{\alpha-t}{y}\right)\rho(\mathrm{d}t).\end{aligned}$$

Since

$$\lim_{y\to0^+}\left(\arctan\frac{\beta-t}{y} - \arctan\frac{\alpha-t}{y}\right) = \begin{cases} 0, & t\in(-\infty,\alpha)\cup(\beta,\infty), \\ \pi, & t\in(\alpha,\beta), \\ \frac{\pi}{2}, & t=\alpha,\beta, \end{cases}$$

the desired (4.10) follows by the dominated convergence theorem.

As for (4.11), due to the calculations

$$\begin{aligned}\mathrm{i}yN(\alpha+\mathrm{i}y) &= \mathrm{i}ya(\alpha+\mathrm{i}y) - \mathrm{i}yb + \mathrm{i}y\int_{\mathbb{R}}\left[\frac{1}{t-(\alpha+\mathrm{i}y)} - \frac{t}{1+t^2}\right]\rho(\mathrm{d}t) \\ &= o(1) - \rho(\{\alpha\}) + \int_{\mathbb{R}\setminus\{\alpha\}} \underbrace{\mathrm{i}y\left[\frac{(t-\alpha)+\mathrm{i}y}{(t-\alpha)^2+y^2} - \frac{t}{1+t^2}\right]}_{=:\,k(t,y)}\rho(\mathrm{d}t),\end{aligned}$$

it remains to show that the integral converges to zero as $y\to0^+$. We split the integral region $\mathbb{R}\setminus\{\alpha\}$ into $J_1 := \{t\in\mathbb{R} : 0 < |t-\alpha| < 1\}$ and $J_2 := \{t\in\mathbb{R} : |t-\alpha|\ge 1\}$. The integral over J_1 tends to zero by the dominated convergence theorem because $\rho(J_1) < +\infty$ and for all $y\in(0,1)$ and $t\in J_1$ we have

$$|k(t,y)| \le \frac{y|t-\alpha|+y^2}{(t-\alpha)^2+y^2} + \frac{|t|y}{1+t^2} \le \frac{[y^2+(t-\alpha)^2]/2+y^2}{(t-\alpha)^2+y^2} + \frac{t^2+y^2}{2(1+t^2)} \le 2.$$

The integral over J_2 also converges to zero by the dominated convergence because $\tau(J_2) < +\infty$ and

$$\sup_{\substack{t\in J_2\\ y\in(0,1)}} (1+t^2)|k(t,y)| = \sup_{\substack{t\in J_2\\ y\in(0,1)}} y\left|\frac{[t-\alpha+\mathrm{i}y](1+t^2) - t[(t-\alpha)^2+y^2]}{(t-\alpha)^2+y^2}\right| < +\infty.$$

Finally, we verify the uniqueness of τ. Suppose that (a,b,τ') is another triplet. By the Stieltjes inversion formula, the measure $\rho'(\mathrm{d}t) := (1+t^2)\tau'(\mathrm{d}t)$ satisfies $\rho'(I) = \rho(I)$ for all open intervals I of finite length. Proposition 4.4 yields $\rho' = \rho$, and therefore, $\tau' = \tau$. □

In many examples the Stieltjes inversion formula is used in the following form.

Corollary 4.28. *Let N be a Nevanlinna function with triplet (a,b,τ). If N extends to a continuous function $\tilde N\colon \mathbb{C}^+\cup[\alpha,\beta]\to\mathbb{C}^+\cup\mathbb{R}$ for some $-\infty<\alpha<\beta<\infty$, then the measure $\tau\restriction_{[\alpha,\beta]}$ is Lebesgue absolutely continuous and its density is given by*

$$\frac{1}{\pi(1+t^2)}\Im[\tilde N(t)], \qquad t\in[\alpha,\beta].$$

Here we characterize the convergence of Nevanlinna functions.

Proposition 4.29. *Let N_n, $n=1,2,3,\dots$, be Nevanlinna functions with representations*

$$N_n(z) = -b_n + \int_{\widehat{\mathbb{R}}} \frac{1+tz}{t-z}\,\widehat{\tau}_n(\mathrm{d}t).$$

The following statements are equivalent.

(1) *N_n converges to a function N locally uniformly on $\mathbb{C}^+$.*

(2) *There is a sequence of distinct points $(z_k)_{k\ge1}$ that converges to a point $z_\infty\in\mathbb{C}^+$, and $\lim_{n\to\infty} N_n(z_k)$ exists in $\mathbb{C}$ for all $k\in\mathbb{N}$.*

(3) *b_n converges to some $b\in\mathbb{R}$ and $(\widehat{\tau}_n)_{n\ge1}$ converges weakly to some finite Borel measure $\widehat{\tau}$ on $\widehat{\mathbb{R}}$.*

Moreover, if the above equivalent conditions hold, then the limit function N is the Nevanlinna function given by

$$N(z) = -b + \int_{\widehat{\mathbb{R}}} \frac{1+tz}{t-z}\,\widehat{\tau}(\mathrm{d}t).$$

Proof. (1) $\Longrightarrow$ (2) is obvious.

(2) $\Longrightarrow$ (3). By performing an affine transformation $z \mapsto pz + q$ ($p > 0$, $q \in \mathbb{R}$), we may assume that $z_1 = \mathrm{i}$. Since $N_n(\mathrm{i}) = -b_n + \mathrm{i}\widehat{\tau}_n(\widehat{\mathbb{R}})$, the sequence $(b_n)_{n\geq 1}$ converges to $b \in \mathbb{R}$ and the total mass $(\widehat{\tau}_n(\widehat{\mathbb{R}}))_{n\geq 1}$ converges to a finite nonnegative number. By the compactness of $\widehat{\mathbb{R}}$ and by Theorem 4.11, the sequence $(\widehat{\tau}_n)_{n\geq 1}$ has a weakly convergent subsequence $(\widehat{\tau}_{n'})$, whose limit is denoted by $\widehat{\tau}$. Let N be the Nevanlinna function determined by the pair $(b, \widehat{\tau})$. The definition of weak convergence implies that for each $z \in \mathbb{C}^+$

$$N_{n'}(z) = -b_{n'} + \int_{\widehat{\mathbb{R}}} \frac{1+tz}{t-z}\, \widehat{\tau}_{n'}(\mathrm{d}t) \to -b + \int_{\widehat{\mathbb{R}}} \frac{1+tz}{t-z}\, \widehat{\tau}(\mathrm{d}t) = N(z) \quad \text{as} \quad n' \to \infty. \tag{4.15}$$

We now take any subsequence $(\widehat{\tau}_{n(j)})_{j\geq 1}$ of $(\widehat{\tau}_n)_{n\in\mathbb{N}}$. By the same reasoning as above, it has a further subsequence $(\widehat{\tau}_{n(j(k))})_{k\geq 1}$ that converges to a finite Borel measure $\tilde{\tau}$. Denoting by $\tilde{N}$ the Nevanlinna function corresponding to $(b, \tilde{\tau})$, we obtain $N_{n(j(k))}(z) \to \tilde{N}(z)$ in the same way as (4.15). Therefore, for any $k \in \mathbb{N}$, we have $N(z_k) = \lim_{n\to\infty} N_n(z_k) = \tilde{N}(z_k)$. By the identity theorem we have $N = \tilde{N}$ on $\mathbb{C}^+$ and hence, by the uniqueness of the Nevanlinna formula, $\tilde{\tau} = \widehat{\tau}$. Lemma 4.8 implies the convergence of the whose sequence $\widehat{\tau}_n \to \widehat{\tau}$ as $n \to \infty$.

(3) $\Longrightarrow$ (1). The pointwise convergence $N_n(z) \to N(z)$ follows by the definition of weak convergence as $\widehat{\mathbb{R}} \ni t \mapsto \frac{1+tz}{t-z} \in \mathbb{C}^+$ is bounded and continuous. Moreover, for each compact subset $K \subseteq \mathbb{C}^+$, the function $(t, z) \mapsto (1 + tz)/(t - z)$ is bounded on $\widehat{\mathbb{R}} \times K$, and so N_n is uniformly bounded on K. By Vitali's theorem (Theorem 4.22), the convergence $N_n \to N$ holds locally uniformly. ☐

Remark 4.30. Be aware that even if a sequence of Nevanlinna functions converges, the triplets in (4.9) might fail to converge to that of the limit function. Take for example the triplet $(0, 0, \delta_n)$, $n \in \mathbb{N}$. Then

$$N_n(z) := \frac{1+zn}{n-z} \to N(z) := z$$

while the limit function has triplet $(1, 0, 0)$. By contrast, the pair $(b, \widehat{\tau})$ works perfectly with respect to the convergence. In the above example, as finite Borel measures on $\widehat{\mathbb{R}}$, the convergence $\delta_n \to \delta_\infty$ holds.

Here is a technical lemma on Nevanlinna functions to be used in later sections.

Lemma 4.31. *Let T be a metric space and $N \colon T \times \mathbb{C}^+ \to \mathbb{C}^+ \cup \mathbb{R}$ be a function such that*

- *for each $z \in \mathbb{C}^+$, the map $t \mapsto N(t, z)$ is continuous,*
- *for each $t \in T$, the map $z \mapsto N(t, z)$ is a Nevanlinna function.*

Then $\frac{\partial^k N}{\partial z^k}$ is continuous on $T \times \mathbb{C}^+$ for every $k \in \mathbb{N}_0$.

Proof. For each $t \in T$ we have the formula

$$N(t, z) = a_t z - b_t + \int_{\mathbb{R}} \frac{1+xz}{x-z}\, \tau_t(\mathrm{d}x), \qquad z \in \mathbb{C}^+.$$

We fix a point $(t_*, z_*) \in T \times \mathbb{C}^+$ and a sequence (t_n, z_n), $n \in \mathbb{N}$, converging to (t_*, z_*). Since $-b_t + \mathrm{i}[a_t + \tau_t(\mathbb{R})] = \Re[N(t, \mathrm{i})]$ is a continuous function of $t \in T$, the sequences a_{t_n}, b_{t_n}, $\tau_{t_n}(\mathbb{R})$ $(n \in \mathbb{N})$ are all bounded. This implies that $N(t_n, z)$, $n \in \mathbb{N}$, $z \in C$, is uniformly bounded for any compact $C \subseteq \mathbb{C}^+$. In pariticular, let C be a smooth simple closed curve surrounding the points z_*, z_n $(n \in \mathbb{N})$. By Cauchy's integral formula, we have

$$\frac{\partial^k N}{\partial z^k}(t_n, z_n) = \frac{k!}{2\pi \mathrm{i}} \int_C \frac{N(t_n, w)}{(w - z_n)^{k+1}}\, \mathrm{d}w.$$

As $N(t_n, w)$ is uniformly bounded, the dominated convergence theorem allows us to conclude $\partial_z^k N(t_n, z_n) \to \partial_z^k N(t_*, z_*)$. ☐

4.4. **Cauchy transform and its relatives.** Let x be a real random variable in a unital C^*-probability space (A, φ). In Theorem 1.27 we encountered the shifted moment generating function $M_x(z)$ that can be written as

$$M_x(z) = \sum_{n=0}^{\infty} \varphi(x^n) z^{n+1} = \int_{\mathbb{R}} \frac{z}{1-zt}\, \mu_x(\mathrm{d}t).$$

Replacing the variable z with $1/z$ gives a function called the Cauchy transform, which is widely used in noncommutative probability. In this section we also introduce some other related functions.

Definition 4.32. Let μ be a finite Borel measure on $\mathbb{R}$. The function

$$G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z-t}\, \mu(\mathrm{d}t), \qquad z \in \mathbb{C}^+$$

is called the **Cauchy transform** of μ; sometimes it is called the Stieltjes transform or Borel transform. The function $F_\mu(z) := 1/G_\mu(z)$ is called the **reciprocal Cauchy transform** of μ.

Proposition 4.33. *For a finite Borel measure μ on $\mathbb{R}$ and $-\infty < \alpha < \beta < \infty$, one has*

$$\mu((\alpha, \beta)) + \frac{1}{2}(\mu(\{\alpha\}) + \mu(\{\beta\})) = -\frac{1}{\pi} \lim_{y\to 0^+} \int_\alpha^\beta \Im[G_\mu(x + \mathrm{i}y)]\, \mathrm{d}x,$$
$$\mu(\{\alpha\}) = \lim_{y\to 0^+} \mathrm{i}y G_\mu(\alpha + \mathrm{i}y).$$

In particular, the map $\mu \mapsto G_\mu$ is injective.

Proof. The formula

$$-G_\mu(z) = -b + \int_{\mathbb{R}} \frac{1+tz}{t-z} \cdot \frac{\mu(\mathrm{d}t)}{1+t^2}, \qquad b := -\int_{\mathbb{R}} \frac{t}{1+t^2}\,\mu(\mathrm{d}t),$$

gives a Nevanlinna formula for $-G_\mu$ and allows us to apply Theorem 4.26 (iii). The injectivity follows from the uniqueness part of the measure τ in Theorem 4.26. □

The following is a consequence of Corollary 4.28.

Corollary 4.34. *For a finite Borel measure μ on $\mathbb{R}$ and $-\infty < \alpha < \beta < \infty$, suppose that G_μ extends to a continuous function $\tilde{G}_\mu \colon \mathbb{C}^+ \cup [\alpha,\beta] \to (-\mathbb{C}^+) \cup \mathbb{R}$. Then $\mu\restriction_{[\alpha,\beta]}$ is Lebesgue absolutely continuous and its density is given by $-\frac{1}{\pi}\Im[\tilde{G}_\mu(t)]$, $t \in [\alpha,\beta]$.*

The Cauchy transform is characterized as follows.

Proposition 4.35. *Let $G\colon \mathbb{C}^+ \to (-\mathbb{C}^+) \cup \mathbb{R}$ be a holomorphic function. Then the following statements are equivalent.*

(1) *$G = G_\mu$ for some probability measure μ.*

(2) $\lim_{y\to+\infty} \mathrm{i}yG(\mathrm{i}y) = 1$.

(3) $\sphericalangle \lim_{z\to\infty} zG(z) = 1$.

Proof. (1) $\Longrightarrow$ (3). By (4.13) and the dominated convergence theorem

$$zG_\mu(z) = \int_{\mathbb{R}} \frac{z}{z-t}\,\mu(\mathrm{d}t) \to 1 \qquad \text{as } z\to\infty,\ z \in \nabla_\gamma.$$

(3) $\Longrightarrow$ (2) is obvious.

(2) $\Longrightarrow$ (1). From the Nevanlinna formula for $-G$ we have

$$G(z) = -az + b + \int_{\mathbb{R}} \frac{1+tz}{z-t}\,\tau(\mathrm{d}t),$$

where $a \ge 0$, $b\in\mathbb{R}$ and τ is a finite Borel measure on $\mathbb{R}$. Moreover,

$$\Re[\mathrm{i}yG(\mathrm{i}y)] = ay^2 + \int_{\mathbb{R}} \frac{y^2}{y^2+t^2}(1+t^2)\,\tau(\mathrm{d}t).$$

From assumption (2) it must hold that $a=0$, and from $\frac{y^2}{y^2+t^2} \uparrow 1$ $(y\uparrow+\infty)$ the monotone convergence theorem yields

$$1 = \lim_{y\to+\infty} \Re[\mathrm{i}yG(\mathrm{i}y)] = \int_{\mathbb{R}} (1+t^2)\,\tau(\mathrm{d}t).$$

Therefore, the measure $\mu(\mathrm{d}t) = (1+t^2)\tau(\mathrm{d}t)$ is a probability measure on $\mathbb{R}$. The function G can be expressed in the form

$$G(z) = b + \int_{\mathbb{R}} \left(\frac{1}{z-t} + \frac{t}{1+t^2}\right)\mu(\mathrm{d}t) = G_\mu(z) + b + \int_{\mathbb{R}} \frac{t}{1+t^2}\,\mu(\mathrm{d}t),$$

which, together with assumption (2), implies $b + \int_{\mathbb{R}} \frac{t}{1+t^2}\,\mu(\mathrm{d}t) = 0$ and hence $G = G_\mu$. □

Proposition 4.36. *Let μ, μ_n $(n\in\mathbb{N})$ be probability measures on $\mathbb{R}$. The following are equivalent.*

(1) *$\mu_n \to \mu$ weakly.*

(2) *G_{μ_n} converges to G_μ locally uniformly on $\mathbb{C}^+$.*

(3) *There exists a sequence of distinct points $(z_k)_{k=1}^\infty \subset \mathbb{C}^+$ such that $\lim_{k\to\infty} z_k$ exists in $\mathbb{C}^+$, and $\lim_{n\to\infty} G_{\mu_n}(z_k) = G_\mu(z_k)$ for every $k\in\mathbb{N}$.*

Proof. (2) $\Longleftrightarrow$ (3) follows from Proposition 4.29 (1) and (2).

(1) $\Longrightarrow$ (3) is obvious from the definition of weak convergence.

(2) $\Longrightarrow$ (1). We extend μ_n to a probability measure $\widehat{\mu}_n$ on $\widehat{\mathbb{R}}$ by setting $\widehat{\mu}_n(\{\infty\}) := 0$. Since $\{\widehat{\mu}_n(\widehat{\mathbb{R}}) = 1\}_{n\ge1}$ is bounded, by Theorem 4.11, there is a subsequence $(\widehat{\mu}_{n(j)})_{j\ge1}$ that converges weakly to a probability measure $\widehat{\mu}$ on $\widehat{\mathbb{R}}$. Since $\mathbb{R} \ni t \mapsto 1/(z-t)$ can be regarded as a bounded continuous function on $\widehat{\mathbb{R}}$ vanishing at infinity, we have

$$G_{\mu_n}(z) = \int_{\widehat{\mathbb{R}}} \frac{1}{z-t}\,\widehat{\mu}_n(\mathrm{d}t) \to \int_{\widehat{\mathbb{R}}} \frac{1}{z-t}\,\widehat{\mu}(\mathrm{d}t) = \int_{\mathbb{R}} \frac{1}{z-t}\,\widehat{\mu}(\mathrm{d}t) = G_{\widehat{\mu}\restriction_{\mathbb{R}}}(z), \qquad z\in\mathbb{C}^+.$$

By assumption (2), the limit function $G_{\widehat{\mu}\restriction_{\mathbb{R}}}$ must be $G_\mu(z)$. Since the Cauchy transform determines the underlying finite Borel measure on $\mathbb{R}$ uniquely (see Proposition 4.33), we have $\widehat{\mu}\restriction_{\mathbb{R}} = \mu$. Recalling that μ is a probability measure on $\mathbb{R}$ and $\widehat{\mu}$ is a probability measure on $\widehat{\mathbb{R}}$, we must have $\widehat{\mu}(\{\infty\}) = 0$. The weak convergence $\widehat{\mu}_{n(j)} \to \widehat{\mu}$ implies

$$\int_{\mathbb{R}} f(x)\mu_{n(j)}(\mathrm{d}x) \to \int_{\mathbb{R}} f(x)\,\mu(\mathrm{d}x) \tag{4.16}$$

for any continuous function f on $\mathbb{R}$ with compact support since such an f can be regarded as a continuous function on $\widehat{\mathbb{R}}$. By Proposition 4.7, the convergence (4.16) implies that $\mu_{n(j)}$ converges weakly to μ on $\mathbb{R}$. By Lemma 4.8, the weak convergence of the whole sequence $\mu_n \to \mu$ holds as one can apply the above arguments to any subsequence of (μ_n) instead of (μ_n) itself. □

Proposition 4.37. *Let N be a Nevanlinna function. The following are equivalent.*

(1) $N = F_\mu$ *for some probability measure* μ *on* $\mathbb{R}$.

(2) $\displaystyle\lim_{y\to+\infty} \frac{N(\mathrm{i}y)}{\mathrm{i}y} = 1.$

(3) $\displaystyle\sphericalangle \lim_{z\to\infty} \frac{N(z)}{z} = 1.$

(4) *There are* $b \in \mathbb{R}$ *and a finite Borel measure* τ *on* $\mathbb{R}$ *such that*

$$N(z) = z - b + \int_{\mathbb{R}} \frac{1+tz}{t-z}\tau(\mathrm{d}t). \tag{4.17}$$

The probability measure μ *is unique.*

Proof. The equivalence (1) $\iff$ (2) $\iff$ (3) follows from Proposition 4.35 applied to $G := 1/N$. Theorem 4.26 implies the equivalence (3) $\iff$ (4). The uniqueness is a consequence of the injectivity of $\mu \mapsto G_\mu$ addressed in Proposition 4.33. □

Remark 4.38. For a probability measure μ the following inequality holds:

$$\Im[F_\mu(z)] \geq \Im(z), \qquad z \in \mathbb{C}^+, \tag{4.18}$$

and the equality holds at some $z \in \mathbb{C}^+$ if and only if μ is a delta measure. The latter statement holds because the equality holds if and only if $\tau = 0$.

A uniform version of Proposition 4.37 (2) or (3) for a family of probability measures characterizes the tightness.

Proposition 4.39. *Let* $\mathscr{P}$ *be a family of probability measures on* $\mathbb{R}$. *Then the following are equivalent.*

(1) $\mathscr{P}$ *is tight.*

(2) $\displaystyle\sphericalangle \lim_{z\to\infty} \sup_{\mu\in\mathscr{P}} \left|\frac{F_\mu(z)}{z} - 1\right| = 0.$

(3) $\displaystyle\lim_{y\to+\infty} \sup_{\mu\in\mathscr{P}} \left|\frac{F_\mu(\mathrm{i}y)}{\mathrm{i}y} - 1\right| = 0.$

Proof. Observe first that conditions (2) and (3) are respectively equivalent to

(2′) $\displaystyle\sphericalangle \lim_{z\to\infty} \sup_{\mu\in\mathscr{P}} |zG_\mu(z) - 1| = 0,$

(3′) $\displaystyle\lim_{y\to+\infty} \sup_{\mu\in\mathscr{P}} |\mathrm{i}yG_\mu(\mathrm{i}y) - 1| = 0.$

(2′) $\Longrightarrow$ (3′) is obvious.

(3′) $\Longrightarrow$ (1) follows from the estimates for $y > 0$:

$$-\Re[\mathrm{i}yG_\mu(\mathrm{i}y) - 1] = \int_{\mathbb{R}} \frac{t^2}{t^2+y^2}\,\mu(\mathrm{d}t) \geq \int_{|t|>y} \frac{t^2}{t^2+y^2}\,\mu(\mathrm{d}t) \geq \frac{1}{2}\mu(\mathbb{R}\setminus[-y,y]).$$

(1) $\Longrightarrow$ (2′). We fix $\gamma > 0$ for the nontangential domain. By the tightness assumption, for every $\varepsilon > 0$ there is $T > 0$ such that

$$\mu(\mathbb{R}\setminus[-T,T]) \leq \frac{\varepsilon}{1+\sqrt{1+\gamma^{-2}}}, \qquad \mu \in \mathscr{P}.$$

Combining the above inequality and (4.13), for all $\mu \in \mathscr{P}$ and $z \in \nabla_\gamma$ we have

$$\begin{aligned}|zG_\mu(z) - 1| &= \left|\int_{\mathbb{R}} \left(\frac{z}{z-t} - 1\right)\mu(\mathrm{d}t)\right| \\ &\leq \int_{[-T,T]} \left|\frac{t}{z-t}\right| \mu(\mathrm{d}t) + (\sqrt{1+\gamma^{-2}}+1)\mu(\mathbb{R}\setminus[-T,T]) \\ &\leq \int_{[-T,T]} \frac{|t|}{\Im(z)}\mu(\mathrm{d}t) + \varepsilon \leq \frac{T}{\Im(z)} + \varepsilon.\end{aligned}$$

Therefore, we obtain $\displaystyle\sup_{\substack{\mu\in\mathscr{P} \\ z\in\nabla_\gamma,\, \Im(z)>T/\varepsilon}} |zG_\mu(z) - 1| < 2\varepsilon.$ □

Finally we introduce and characterize transforms useful to study multiplicative monotone convolution.

Definition 4.40. Let μ be a probability measure on $\mathbb{R}$. The function

$$\psi_\mu(z) := \int_{\mathbb{R}} \frac{zt}{1-zt}\,\mu(\mathrm{d}t), \qquad z \in \mathbb{C}^+$$

is called the ψ**-transform** of μ (or the moment generating function of μ) and

$$\eta_\mu(z) := \frac{\psi_\mu(z)}{1+\psi_\mu(z)}, \qquad z \in \mathbb{C}^+$$

is called the η**-transform** of μ.

We can check by straightforward calculations that $\psi_\mu(z) = zG_\mu(1/z) - 1$, where G_μ is defined on the lower half-plane $-\mathbb{C}^+$ by the same formula in Definition 4.32, i.e., $G_\mu(z) := \overline{G_\mu(\overline{z})}$, and also

$$\eta_\mu(z) = 1 - zF_\mu\left(\frac{1}{z}\right), \tag{4.19}$$

where $F_\mu := 1/G_\mu$ on $-\mathbb{C}^+$.

Proposition 4.41. *For any probability measure μ on $\mathbb{R}$ such that $\mu \neq \delta_0$, the following hold.*

(i) *η_μ is a holomorphic map from $\mathbb{C}^+$ into $\mathbb{C} \setminus [0, +\infty)$.*

(ii) *$\arg z \le \arg \eta_\mu(z) \le \arg z + \pi$, i.e., $\eta_\mu(z)/z \in \mathbb{C}^+ \cup \mathbb{R}$, for all $z \in \mathbb{C}^+$.*

(iii) *$\sphericalangle \lim_{z\to 0} \eta_\mu(z) = 0$.*

Conversely, if a holomorphic map $\eta\colon \mathbb{C}^+ \to \mathbb{C} \setminus [0, +\infty)$ satisfies $\arg z \le \arg \eta(z) \le \arg z + \pi$ and $\sphericalangle \lim_{z\to 0} \eta(z) = 0$, then there is a unique probability measure μ on $\mathbb{R}$ such that $\mu \neq \delta_0$ and $\eta = \eta_\mu$.

Proof. The holomorphicity is immediate from (4.19). Inequality (4.18) implies $1/z - F_\mu(1/z) \in (\mathbb{C}^+ \cup \mathbb{R}) \setminus \{0\}$, which in turn yields $\eta_\mu(\mathbb{C}^+) \subseteq \mathbb{C} \setminus [0, +\infty)$ and condition (ii). Condition (iii) is equivalent to Proposition 4.37 (3) for $N = F_\mu$.

For the last converse statement, we can see that the function $N(z) := z[1 - \overline{\eta(1/\overline{z})}]$ is a Nevanlinna function and satisfies $\lim_{y\to+\infty} N(\mathrm{i}y)/(\mathrm{i}y) = 1$. Therefore, by Proposition 4.37 there is a unique probability measure μ on $\mathbb{R}$ such that $N(z) = F_\mu(z)$. Since η is assumed not to take 0, the value $N(z)$ never equals z and so $\mu \neq \delta_0$. We thus obtain

$$\eta_\mu(z) = 1 - z\overline{N\left(\frac{1}{\overline{z}}\right)} = \eta(z). \qquad \square$$

4.5. **Support and moments of probability measures.** From the Cauchy transform one can extract information about support and moments of the underlying probability measures. As a general symbol, for a finite Borel measure τ on $\mathbb{R}$ and $n \in \mathbb{N}_0$, if $\int_\mathbb{R} |t|^n\, \tau(\mathrm{d}t) < +\infty$ we set

$$\mathrm{m}_n(\tau) := \int_\mathbb{R} t^n\, \tau(\mathrm{d}t)$$

and call it the n**th moment of** τ. Note that Hölder's inequality implies

$$\int_\mathbb{R} |x|^k \tau(\mathrm{d}x) \le \tau(\mathbb{R})^{1-\frac{k}{\ell}} \left[\int_\mathbb{R} |x|^\ell \tau(\mathrm{d}x)\right]^{\frac{k}{\ell}}, \qquad 0 < k < \ell < +\infty, \tag{4.20}$$

so if the absolute ℓth moment is finite, then the lower order moments all exist.

Proposition 4.42. *Let μ be a probability measure on $\mathbb{R}$, C a closed subset of $\mathbb{R}$, and $-\infty < \alpha < \beta < \infty$.*

(i) *$\mu(C) = 1$ holds if and only if G_μ has an analytic continuation $\tilde{G}_\mu$ on $\mathbb{C} \setminus C$ such that $\tilde{G}_\mu(\overline{z}) = \overline{\tilde{G}_\mu(z)}$.*

(ii) *If $\mu([\alpha, \beta]) = 1$ then $\tilde{G}_\mu$ on $\mathbb{C} \setminus [\alpha, \beta]$ considered above has the Laurent series expansion*

$$\tilde{G}_\mu(z) = \sum_{n\ge 0} \frac{\mathrm{m}_n(\mu)}{z^{n+1}}, \qquad z \in \mathbb{C},\ |z| > \max\{|\alpha|, |\beta|\}.$$

(iii) *If $\mu([\alpha, \beta]) = 1$ then F_μ has an analytic continuation $\tilde{F}_\mu$ on $\mathbb{C} \setminus [\alpha, \beta]$ such that $\tilde{F}_\mu(\overline{z}) = \overline{\tilde{F}_\mu(z)}$ and there exist real numbers $b_1, b_2, \dots$ such that*

$$\tilde{F}_\mu(z) = z - \sum_{n\ge 1} \frac{b_n}{z^{n-1}}, \qquad z \in \mathbb{C},\ |z| > \max\{|\alpha|, |\beta|\}.$$

(iv) *Suppose that F_μ has an analytic continuation $\tilde{F}_\mu$ on $\mathbb{C} \setminus [\alpha, \beta]$ such that $\tilde{F}_\mu(\overline{z}) = \overline{\tilde{F}_\mu(z)}$. Then $\operatorname{supp}(\mu) \cap (\beta, +\infty)$ contains at most one point. An atom exists in $(\beta, +\infty)$ if and only if $\tilde{F}_\nu$ has a zero on $(\beta, +\infty)$, in which case μ has an atom at the zero of $\tilde{F}$. Similarly, $\operatorname{supp}(\mu) \cap (-\infty, \alpha)$ contains at most one point and an atom exists in $(-\infty, \alpha)$ if and only if $\tilde{F}_\nu$ has a zero on $(-\infty, \alpha)$.*

Proof. (i) Suppose that $\mu(C) = 1$. Then the integral formula

$$\tilde{G}_\mu(z) = \int_C \frac{1}{z - t}\, \mu(\mathrm{d}t) \tag{4.21}$$

gives an analytic continuation of G_μ to $\mathbb{C} \setminus C$ with $\tilde{G}_\mu(\overline{z}) = \overline{\tilde{G}_\mu(z)}$. Conversely, if G_μ has an analytic continuation $\tilde{G}_\mu$ on $\mathbb{C} \setminus C$ with $\tilde{G}_\mu(\overline{z}) = \overline{\tilde{G}_\mu(z)}$, then the latter condition implies $\tilde{G}_\mu(x)$ takes real values for $x \in \mathbb{R} \setminus C$. For each interval $[a, b] \subseteq \mathbb{R} \setminus C$, by the Stieltjes inversion formula in Corollary 4.34 we have $\mu([a, b]) = 0$. This implies $\mu(\mathbb{R} \setminus C) = 0$.

(ii) The Laurent series expansion can be obtained from formula (4.21) and the fact that the series

$$\frac{1}{z - t} = \sum_{n\ge 0} \frac{t^n}{z^{n+1}}$$

converges uniformly over $t \in [\alpha, \beta]$ and $z \in \{w \in \mathbb{C} : |w| \ge \varepsilon + \max\{|\alpha|, |\beta|\}\}$ for each $\varepsilon > 0$.

(iii) Since $\tilde{G}_\mu$ in (4.21) has no zeros in $\mathbb{C}\setminus[\alpha,\beta]$, its reciprocal $\tilde{F}_\mu := \tilde{G}_\mu$ is the desired analytic continuation. For the series expansion, we take a Nevanlinna formula for F_μ

$$F_\mu(z) = z - b + \int_{\mathbb{R}} \frac{1+tz}{t-z}\,\tau(\mathrm{d}t), \qquad z\in\mathbb{C}^+. \tag{4.22}$$

Since the extension $\tilde{F}_\mu$ takes real values on $\mathbb{R}\setminus[\alpha,\beta]$, the Stieltjes inversion (see Corollary 4.28) implies τ is supported on $[\alpha,\beta]$. Hence, the analytic continuation is given by

$$\tilde{F}_\mu(z) = z - b + \int_{[\alpha,\beta]} \frac{1+tz}{t-z}\,\tau(\mathrm{d}t), \qquad z\in\mathbb{C}\setminus[\alpha,\beta]. \tag{4.23}$$

The remaining argument is similar to (ii) thanks to formula (4.14); the coefficients b_n are given by $b_1 = b + \int_{[\alpha,\beta]} t\,\tau(\mathrm{d}t)$ and $b_n = \int_{[\alpha,\beta]} t^{n-2}(1+t^2)\,\tau(\mathrm{d}t)$, $n\geq 2$.

(iv) As discussed in (iii), the assumption implies that F_μ has an analytic extension $\tilde{F}_\mu$ of the form (4.23). One can see that

$$\tilde{F}_\mu'(x) = 1 + \int_{[\alpha,\beta]} \frac{1+t^2}{(t-x)^2}\,\tau(\mathrm{d}t) \geq 1, \qquad x\in\mathbb{R}\setminus[\alpha,\beta], \tag{4.24}$$

so that $\tilde{F}_\mu$ is strictly increasing on $\mathbb{R}\setminus[\alpha,\beta]$ and $\lim_{x\to\pm\infty}\tilde{F}_\nu(x) = \pm\infty$. In particular, $\tilde{F}_\nu$ has at most one zero in each interval $(-\infty,\alpha)$ and $(\beta,+\infty)$. If $\tilde{F}_\nu$ has a zero $\gamma\in(\beta,+\infty)$ then $\tilde{G}_\mu := 1/\tilde{F}_\mu$ is analytic in $(\beta,+\infty)\setminus\{\gamma\}$ taking real values there, and

$$q := \lim_{y\to 0} \mathrm{i}yG_\mu(\gamma+\mathrm{i}y) = \lim_{y\to 0}\frac{\mathrm{i}y}{\tilde{F}_\mu(\gamma+\mathrm{i}y)-\tilde{F}_\mu(\gamma)} = \frac{1}{\tilde{F}_\mu'(\gamma)} \in (0,1].$$

Therefore, μ assigns the mass q to the point γ and $\mu((\beta,+\infty)\setminus\{\gamma\}) = 0$. By contrast, if $\tilde{F}_\mu$ has no zeros on $(\beta,+\infty)$ then $\tilde{G}_\mu = 1/\tilde{F}_\mu$ is analytic in $\mathbb{C}\setminus(-\infty,\beta]$ taking real values, so that μ is supported on $(-\infty,\beta]$. A similar analysis is valid on the interval $(-\infty,\alpha)$. □

A particular class is the set of probability measures on the nonnegative real line. This class admits characterizations by means of the reciprocal Cauchy transform and η-transform.

Proposition 4.43. *Let μ be a probability measure on $\mathbb{R}$. Let (b,τ) be the pair appearing for $N = F_\mu$ in (4.17). The following are equivalent.*

1. *μ is supported on $[0,+\infty)$.*
2. *F_μ has an analytic continuation $\tilde{F}_\mu$ defined on $\mathbb{C}\setminus[0,+\infty)$ such that $\tilde{F}_\mu(\overline{z}) = \overline{\tilde{F}_\mu(z)}$ and $\lim_{x\uparrow 0}\tilde{F}_\mu(x)\in(-\infty,0]$.*
3. *τ is supported on $(0,+\infty)$ and $\int_0^\infty t^{-1}\,\tau(\mathrm{d}t)\leq b$.*

Proof. (1) $\Longrightarrow$ (2). The existence of analytic continuation $\tilde{F}_\mu$ can be proved in a similar way to the proof of Proposition 4.42 (iii); it is given by $\tilde{F}_\mu = 1/\tilde{G}_\mu$, where

$$\tilde{G}_\mu(z) = \int_{[0,+\infty)} \frac{1}{z-t}\,\mu(\mathrm{d}t), \qquad z\in\mathbb{C}\setminus[0,+\infty).$$

Obviously, $\tilde{G}_\mu(x) < 0$ for all $x<0$, so that $\tilde{F}_\mu(x) < 0$. Since $\tilde{F}_\mu$ is (strictly) increasing on $(-\infty,0)$ we obtain $\tilde{F}_\mu(0-)\in(-\infty,0]$.

(2) $\Longrightarrow$ (3). Let (b,τ) be the pair in (4.17) for $N = F_\mu$. By the Stieltjes inversion formula, τ is supported on $[0,+\infty)$. Then the Nevanlinna formula (4.17) naturally gives the expression for $\tilde{F}_\mu$. We write

$$\tilde{F}_\mu(z) = z - b + \underbrace{\int_{[0,1)} \frac{1+zt}{t-z}\,\tau(\mathrm{d}t)}_{=:I_1(z)} + \underbrace{\int_{[1,+\infty)} \frac{1+zt}{t-z}\,\tau(\mathrm{d}t)}_{=:I_2(z)}, \qquad z\in\mathbb{C}\setminus[0,+\infty). \tag{4.25}$$

Since the function $(-1,0)\ni x\mapsto (1+xt)/(t-x)$ is increasing and positive for all $t\in[0,1)$, we can use the monotone convergence theorem to conclude that $I_1(x)\uparrow\int_{[0,1)} t^{-1}\,\tau(\mathrm{d}t)$ as $x\to 0^-$. On the other hand, for I_2 we can use the dominated convergence theorem to deduce $I_2(x)\to\int_{[1,+\infty)} t^{-1}\,\tau(\mathrm{d}t)$ as $x\to 0^-$ because $|(1+tx)/(t-x)|\leq(1+t)/t$ for all $t\geq 1$ and $-1<x<0$. By the assumption $\tilde{F}_\mu(0-)\leq 0$, we must have $\int_{[0,1)} t^{-1}\,\tau(\mathrm{d}t) < +\infty$; in particular, $\tau(\{0\}) = 0$. The condition $\tilde{F}_\mu(0-)\leq 0$ now reads $\int_0^\infty t^{-1}\,\tau(\mathrm{d}t)\leq b$.

(3) $\Longrightarrow$ (1). Repeating the above arguments we have that $\tilde{F}_\mu$ in (4.25) satisfies $\tilde{F}_\mu(0-) = -b + \int_0^\infty t^{-1}\,\tau(\mathrm{d}t)\leq 0$. Since $\tilde{F}_\mu$ is increasing on $(-\infty,0)$, it is negative and so the holomorphic function $\tilde{G}_\mu := 1/\tilde{F}_\mu$ on $\mathbb{C}\setminus[0,+\infty)$ satisfies the condition of Proposition 4.42 (i). □

The η-transform of a probability measure on $[0,+\infty)$ analytically extends to $\mathbb{C}\setminus[0,+\infty)$ due to (4.19). The following proposition gives a useful characterization of η-transforms.

Proposition 4.44. *Let $\eta\colon\mathbb{C}\setminus[0,+\infty)\to\mathbb{C}$ be a holomorphic function. There is a probability measure $\mu\neq\delta_0$ on $[0,+\infty)$ such that $\eta = \eta_\mu$ on $\mathbb{C}^+$ if and only if the following conditions hold.*

- (i) *$\eta(\overline{z}) = \overline{\eta(z)}$ on $\mathbb{C}\setminus[0,+\infty)$.*
- (ii) *η is a self-map of $\mathbb{C}\setminus[0,+\infty)$ and $\arg z\leq\arg\eta_\mu(z)<\pi$ on $\mathbb{C}^+$.*

(iii) *For any $\theta \in (0,\pi)$ we have*

$$\lim_{\substack{z\to 0\\ \theta<\arg z<2\pi-\theta}} \eta(z) = 0.$$

Moreover, if the above conditions hold then η has the following formula

$$\eta(z) = b' z + \int_{(0,+\infty)} \frac{z}{1-tz}\cdot\frac{1+t^2}{t}\,\tau(\mathrm{d}t), \qquad z\in\mathbb{C}\setminus[0,+\infty), \tag{4.26}$$

where $b'\ge 0$ and τ is a finite Borel measure on $(0,+\infty)$ such that $\int_0^\infty t^{-1}\,\tau(\mathrm{d}t) < +\infty$.

Proof. Suppose that $\eta\restriction_{\mathbb{C}^+} = \eta_\mu$ for some probability measure $\mu\neq\delta_0$ on $[0,+\infty)$. From Proposition 4.43 (2) and (4.19), η is given by

$$\eta(z) = 1 - z\tilde{F}_\mu\left(\frac{1}{z}\right).$$

This implies condition (i). Before proving condition (ii) we first establish (4.26): combining formula (4.25) and Proposition 4.43 (3) we have

$$\eta(z) = \left(b - \int_0^\infty \frac{1}{t}\,\tau(\mathrm{d}t)\right) z + \int_0^\infty \left(\frac{z(z+t)}{1-zt} + \frac{z}{t}\right)\tau(\mathrm{d}t),$$

which is formula (4.26) with $b' := b - \int_0^\infty t^{-1}\,\tau(\mathrm{d}t)$. As one of b' and τ is nonzero, formula (4.26) obviously yields that $\eta(x)<0$ for all $x<0$. In addition, as $z/(1-tz)\in\mathbb{C}^+$ for all $z\in\mathbb{C}^+$ and $t>0$, we have $\eta(z)\in\mathbb{C}^+$. Combining the fact $\eta(z)\in\mathbb{C}^+$ and the inequality $\arg z\le\eta(z)\le\arg z+\pi$ known in Proposition 4.41 we conclude condition (ii). Condition (iii) is equivalent to

$$\lim_{\substack{|z|\to\infty\\ \theta<\arg z<2\pi-\theta}} \frac{\tilde{F}_\mu(z)}{z} = 1, \tag{4.27}$$

which can be shown as in Theorem 4.26 (i). More precisely, the bound (4.13) actually holds for all $z\in\mathbb{C}$ with $\theta<\arg z<2\pi-\theta$ and $t\ge 0$ with $\theta := \arctan\gamma$, because $|z/(z-t)|\le 1$ for $\Re(z)<0$. The remaining proof is the same as Theorem 4.26 (i).

For the converse, suppose that η satisfies conditions (i)–(iii). Let $F(z) := z[1-\eta(1/z)]$, $z\in\mathbb{C}\setminus[0,+\infty)$, which is expected to be the reciprocal Cauchy transform of the desired μ. Condition (i) implies F is holomorphic with $F(\overline{z}) = \overline{F(z)}$. Condition (ii) implies $F\restriction_{\mathbb{C}^+}$ is a Nevanlinna function. Condition (iii) implies $F(z)/z\to 1$ as $z\to\infty$ within ∇_γ for any $\gamma>0$. Therefore, by Proposition 4.37 there is a probability measure μ on $\mathbb{R}$ such that $F = F_\mu$ on $\mathbb{C}^+$. Moreover, condition (ii) implies η takes negative values on $(-\infty,0)$, so that F takes also negative values there. By Proposition 4.43, μ is supported on $[0,+\infty)$. ☐

Remark 4.45. The limit of a function $f\colon\mathbb{C}\setminus[0,+\infty)\to\mathbb{C}$ as $z\to 0$ satisfying $\arg z\in(\theta,2\pi-\theta)$ could be called the nontangential limit of f at 0. This is because the domain $\mathbb{C}\setminus[0,+\infty)$ is conformally equivalent to $\mathbb{C}^+$ by the map $z\mapsto\sqrt{z}$, and then the domain $\{z : \arg z\in(\theta,2\pi-\theta)\}$ is mapped exactly onto the sector $\nabla_{\tan(\theta/2)}$. Accordingly, one could give an alternative proof of (4.27) based on Lindelöf's theorem: observing that the function $\tilde{F}_\mu(z)/z$ maps $\mathbb{C}\setminus[0,+\infty)$ into $\mathbb{C}\setminus(-\infty,0]$ that is also conformally equivalent to $\mathbb{C}^+$ (see Proposition 4.43 (2)), one could use Lindelöf's theorem to extend the known nontangential limit from the upper half-plane in Theorem 4.26 (i) to (4.27).

We turn our attention to the existence of finite moments.

Proposition 4.46. *Let μ be a probability measure on $\mathbb{R}$ and let $n\in\mathbb{N}$. Let τ be the finite Borel measure in (4.17) for $N = F_\mu$. Then the following conditions are equivalent.*

(1) $\displaystyle\int_{\mathbb{R}} t^{2n}\,\mu(\mathrm{d}t) < +\infty.$

(2) *There exist $a_1, a_2, \dots, a_{2n}\in\mathbb{R}$ such that*

$$G_\mu(z) = \frac{1}{z} + \frac{a_1}{z^2} + \frac{a_2}{z^3} + \cdots + \frac{a_{2n}}{z^{2n+1}} + o(|z|^{-(2n+1)}) \tag{4.28}$$

for $z = \mathrm{i}y$ as $y\to+\infty$.

(3) $\displaystyle\int_{\mathbb{R}} t^{2n}\,\tau(\mathrm{d}t) < +\infty.$

(4) *There exist $b_1, b_2, \dots, b_{2n}\in\mathbb{R}$ such that*

$$F_\mu(z) = z - b_1 - \frac{b_2}{z} - \cdots - \frac{b_{2n}}{z^{2n-1}} + o(|z|^{-(2n-1)}) \tag{4.29}$$

for $z = \mathrm{i}y$ as $y\to+\infty$.

If the above equivalent conditions are satisfied then the expansions (4.28) and (4.29) hold as $z\to\infty$ nontangentially, i.e., the remainder term $o(|z|^{-(2n+1)})$ in (4.28) is a function $r(z)$ that satisfies $\sphericalangle\lim_{z\to\infty} z^{2n+1} r(z) = 0$, and similar for (4.29). Moreover, it holds that

$$a_\ell = \mathrm{m}_\ell(\mu), \qquad 1\le\ell\le 2n, \tag{4.30}$$

$$b_\ell = \int_{\mathbb{R}} t^{\ell-2}(1+t^2)\tau(\mathrm{d}t), \qquad 2\le\ell\le 2n, \tag{4.31}$$

$$b_1 = a_1 = \mathrm{m}_1(\mu), \tag{4.32}$$

$$b_2 = a_2 - a_1^2 = \mathrm{Var}(\mu). \tag{4.33}$$

Proof. (1) $\Longrightarrow$ (2). The assumption implies that $\int_{\mathbb{R}} |t|^\ell \, \mu(\mathrm{d}t) < \infty$ for $1 \le \ell \le 2n$. Observe that the identity

$$\frac{1}{z-t} = \sum_{\ell=0}^{2n} \frac{t^\ell}{z^{\ell+1}} + \frac{t^{2n+1}}{z^{2n+1}(z-t)} \tag{4.34}$$

holds, which integrates to

$$G_\mu(z) = \sum_{\ell=0}^{2n} \frac{\mathrm{m}_\ell(\mu)}{z^{\ell+1}} + \underbrace{\frac{1}{z^{2n+1}} \int_{\mathbb{R}} \frac{t^{2n+1}}{z-t} \, \mu(\mathrm{d}t)}_{=:\, R_{2n}(z)}. \tag{4.35}$$

By the dominated convergence theorem we can show $R_{2n}(\mathrm{i}y) = o(y^{-2n-1})$ as $y \to +\infty$.

(2) $\Longrightarrow$ (1). We only consider the case $n = 2$ which should well clarify how the general n can be handled. Keeping in mind that $a_1/(\mathrm{i}y)^2$ is real, we observe that

$$y^3 \Im \left[G_\mu(\mathrm{i}y) - \frac{1}{\mathrm{i}y} - \frac{a_1}{(\mathrm{i}y)^2} \right] = y^3 \Im \left[G_\mu(\mathrm{i}y) - \frac{1}{\mathrm{i}y} \right] = \int_{\mathbb{R}} \frac{y^2 t^2}{y^2 + t^2} \, \mu(\mathrm{d}t),$$

where (4.35) for $n = 0$ is used in the second equality. By assumption, the left-hand side above is bounded as $y \to +\infty$. By the monotone convergence theorem we get

$$\int_{\mathbb{R}} t^2 \, \mu(\mathrm{d}t) < +\infty.$$

This implies that $\int_{\mathbb{R}} |t| \, \mu(\mathrm{d}t)$ is also finite; see (4.20). By the established implication (1) $\Longrightarrow$ (2) for $n = 1$, we have

$$G_\mu(\mathrm{i}y) = \frac{1}{\mathrm{i}y} + \frac{\mathrm{m}_1(\mu)}{(\mathrm{i}y)^2} + \frac{\mathrm{m}_2(\mu)}{(\mathrm{i}y)^3} + o(y^{-3}).$$

Since the asymptotic expansion is unique, we have $a_1 = \mathrm{m}_1(\mu)$ and $a_2 = \mathrm{m}_2(\mu)$.

Next, formula (4.35) with $n = 1$ yields

$$y^5 \Im \left[G_\mu(\mathrm{i}y) - \frac{1}{\mathrm{i}y} - \frac{\mathrm{m}_1(\mu)}{(\mathrm{i}y)^2} - \frac{\mathrm{m}_2(\mu)}{(\mathrm{i}y)^3} - \frac{a_3}{(\mathrm{i}y)^4} \right] = y^5 \Im [R_2(\mathrm{i}y)] = -\int_{\mathbb{R}} \frac{y^2 t^4}{y^2 + t^2} \, \mu(\mathrm{d}t).$$

The left side is bounded as $y \to +\infty$ since we have assumed the expansion $G_\mu(\mathrm{i}y) = (\mathrm{i}y)^{-1} + a_1(\mathrm{i}y)^{-2} + a_2(\mathrm{i}y)^{-3} + a_3(\mathrm{i}y)^{-4} + O(y^{-5})$. By the monotone convergence theorem, we obtain condition (1), i.e.,

$$\int_{\mathbb{R}} t^4 \, \mu(\mathrm{d}t) < +\infty$$

as desired. The previous reasoning for showing $a_1 = \mathrm{m}_1(\mu)$ and $a_2 = \mathrm{m}_2(\mu)$ now also works to show that $a_3 = \mathrm{m}_3(\mu)$ and $a_4 = \mathrm{m}_4(\mu)$, and thus (4.30) holds.

(3) $\Longleftrightarrow$ (4). The proof is very similar to the equivalence of (1) and (2). In the course of the proof, formula (4.31) naturally appears. Note that formula (4.14) is helpful.

(2) $\Longleftrightarrow$ (4) is an easy consequence of the relation $G_\mu(z) = 1/F_\mu(z)$ and the geometric series expansion $1/(1-\zeta) = \zeta + \zeta^2 + \cdots$. For example, in case $n = 1$, assuming $G_\mu(z) = z^{-1} + \mathrm{m}_1(\mu) z^{-2} + \mathrm{m}_2(\mu) z^{-3} + o(z^{-3})$, we have

$$\begin{aligned} F_\mu(z) &= \frac{1}{z^{-1} + \mathrm{m}_1(\mu) z^{-2} + \mathrm{m}_2(\mu) z^{-3} + o(z^{-3})} = \frac{z}{1 + \mathrm{m}_1(\mu) z^{-1} + \mathrm{m}_2(\mu) z^{-2} + o(z^{-2})} \\ &= z \left[1 - \left(\frac{\mathrm{m}_1(\mu)}{z} + \frac{\mathrm{m}_2(\mu)}{z^2} + o(z^{-2}) \right) + \left(\frac{\mathrm{m}_1(\mu)}{z} + o(z^{-1}) \right)^2 + o(z^{-2}) \right] \\ &= z - \mathrm{m}_1(\mu) - \frac{\mathrm{m}_2(\mu) - \mathrm{m}_1(\mu)^2}{z} + o(z^{-1}), \qquad z = \mathrm{i}y, \ y \to +\infty. \end{aligned}$$

This also verifies (4.32) and (4.33).

(4.28) *and* (4.29) *as the nontangential limits.* As for (4.28), assuming (1) and fixing $\gamma > 0$, let us prove $z^{2n+1} R_{2n}(z) \to 0$ as $z \to \infty$ within ∇_γ. In order to use the dominated convergence, it suffices to find a bound of the form $1/|z - t| \le C/(1 + |t|)$ with a constant C independent of $z \in \nabla_\gamma \cap \{\Im z > 1\}$. For $t \in [-1, 1]$ we can simply find an upper bound $1/|z - t| \le 1/\Im(z) \le 1 \le 2/(1 + |t|)$. For $|t| \ge 1$ we consider $z = x + \mathrm{i}y \in \nabla_\gamma$ with $y > 1$, and proceed as

$$\begin{aligned} |z - t|^2 &= (x - t)^2 + y^2 \ge x^2 - 2xt + t^2 + \gamma^2 x^2 \\ &= (1 + \gamma^2) \left(x - \frac{t}{1 + \gamma^2} \right)^2 + \frac{1}{1 + \gamma^{-2}} t^2 \ge \frac{1}{1 + \gamma^{-2}} t^2, \end{aligned}$$

so that

$$\frac{1}{|z - t|} \le \frac{\sqrt{1 + \gamma^{-2}}}{|t|} \le \frac{2\sqrt{1 + \gamma^{-2}}}{1 + |t|}.$$

Of course (4.29) can be similarly proved. □

Remark 4.47. a) The coefficients b_n are called the Boolean cumulants of μ that are central notions in "Boolean probability theory", another type of noncommutative probability. For this reason the minus signs are put in formula (4.29). The interested reader is referred to [147].

b) The assumption that the coefficients $a_1, a_2, \dots$ are real is crucial. Indeed, as for the Cauchy distribution (4.38), its Cauchy transform (4.37) has a convergent series expansion

$$G_\mu(z) = \sum_{n\geq 0} \frac{(a-\mathrm{i}b)^n}{z^{n+1}}, \qquad z\in\mathbb{C}^+ \text{ with } |z| > \sqrt{a^2+b^2}.$$

However, the second moment is infinity.

c) Alternatively, one could also prove $\sphericalangle \lim_{z\to\infty} z^{2n+1}R_{2n}(z) = 0$ using Lindelöf's theorem. Indeed, one could decompose $z^{2n+1}R_{2n} = S_{2n}^+ + S_{2n}^-$, where

$$S_{2n}^+(z) = \int_{[0,+\infty)} \frac{t^{2n+1}}{z-t}\,\mu(\mathrm{d}t), \qquad S_{2n}^-(z) = \int_{(-\infty,0)} \frac{t^{2n+1}}{z-t}\,\mu(\mathrm{d}t).$$

Because $\mp S_{2n}^\pm(z)$ are holomorphic maps from $\mathbb{C}^+$ into $\mathbb{C}^+\cup\mathbb{R}$ and the convergence $S_{2n}^\pm(\mathrm{i}y)\to 0$ holds by the dominated convergence, Lindelöf's theorem would imply $\sphericalangle \lim_{z\to\infty} S_{2n}^\pm(z) = 0$.

d) In the special case $n=1$, condition (3) means that the measure $\rho(\mathrm{d}t) := (1+t^2)\tau(\mathrm{d}t)$ is a finite measure. Then the Nevanlinna formula for F_μ can be reduced to

$$F_\mu(z) = z - m + \int_{\mathbb{R}} \frac{\rho(\mathrm{d}t)}{t-z}, \tag{4.36}$$

where $m := \mathrm{m}_1(\mu)$. Note that $\rho(\mathbb{R}) = \mathrm{Var}(\mu)$.

4.6. **Methods to compute Cauchy transforms.** In some situations, we can obtain an explicit formula for the Cauchy transform before knowing the underlying probability measure, e.g. when we solve a functional equation or a differential equation satisfied by the Cauchy transform. In such a situation, we can compute the measure by the Stieltjes inversion formula. The next two examples naturally appear from the study of monotone convolution semigroups and infinitely divisible distributions, see Example 5.17.

Example 4.48. Let us consider the function

$$G(z) = \frac{1}{z-a+\mathrm{i}b} \tag{4.37}$$

where $a\in\mathbb{R}$, $b>0$ are constants. Since G is a holomorphic function from $\mathbb{C}^+$ into $-\mathbb{C}^+$ and $\lim_{z\to\infty} zG(z) = 1$, we have $G = G_\mu$ for some probability measure μ on $\mathbb{R}$. Moreover, G extends to a continuous function $G\colon \mathbb{C}^+\cup\mathbb{R}\to\mathbb{C}^+$. By Corollary 4.34, μ is Lebesgue absolutely continuous on $\mathbb{R}$ with density

$$\frac{\mathrm{d}\mu}{\mathrm{d}t} = -\frac{1}{\pi}\Im\left[\frac{1}{t-a+\mathrm{i}b}\right] = \frac{b}{\pi[(t-a)^2+b^2]}. \tag{4.38}$$

This is the Cauchy distribution. Thus, we have verified that the Cauchy transform of the Cauchy distribution μ is given by (4.37).

Example 4.49. Let $v>0$. The function

$$G(z) := \frac{1}{\sqrt{z^2-2v}}$$

where $\sqrt{w}$ is defined on $\mathbb{C}\setminus[0,+\infty)$ by $\sqrt{w} = |w|^{1/2}e^{(\mathrm{i}/2)\arg w}, 0<\arg w<2\pi$. Then G is a holomorphic function from $\mathbb{C}^+$ into $-\mathbb{C}^+$. One can check that $\mathrm{i}yG(\mathrm{i}y)\to 1$, so that G is the Cauchy transform of a probability measure μ. The function G extends to a continuous function on $\mathbb{C}^+\cup(\mathbb{R}\setminus\{\pm\sqrt{2v}\})$, so that by Corollary 4.34, the underlying probability measure μ is Lebesgue absolutely continuous on $\mathbb{R}\setminus\{\pm\sqrt{2v}\}$ and the density is given by

$$-\frac{1}{\pi}\lim_{\varepsilon\to 0^+}\Im[G(t+\mathrm{i}\varepsilon)] = \frac{1}{\pi\sqrt{2v-t^2}}\,\chi_{(-\sqrt{2v},\sqrt{2v})}(t). \tag{4.39}$$

Since $G(z)$ diverges as $z\to\pm\sqrt{2v}$, one has to check whether μ has an atom at these points or not. Because we have an estimate $|G(z)|\leq C|z\mp\sqrt{2v}|^{-1/2}$ as $z\to\pm\sqrt{2v}$, we have $\mu(\{\pm\sqrt{2v}\}) = \lim_{\varepsilon\to0^+}\mathrm{i}\varepsilon G(\pm\sqrt{2v}+\mathrm{i}\varepsilon) = 0$. Hence, μ is Lebesgue absolutely continuous on $\mathbb{R}$, supported on $(-\sqrt{2v},\sqrt{2v})$ and $\mathrm{d}\mu/\mathrm{d}t$ is given by (4.39), i.e., μ is the arcsine law $\mathrm{A}(0,v)$ with mean 0 and variance v. Another way to check the absence of atoms is to show that (4.39) has total mass 1 by directly computing the integral; then there would be no room to have atoms.

The other direction, computing the Cauchy transform of a given probability measure, is usually harder. For example, if we do not know formula (4.37) but want to compute G_μ for the Cauchy distribution μ, we need to calculate the integral

$$\int_{\mathbb{R}} \frac{1}{z-t}\cdot\frac{b}{\pi[(t-a)^2+b^2]}\,\mathrm{d}t.$$

In this case, the residue theorem allows us to perform the calculations.

When μ has compact support and the moments are explicit, we can sometimes find a closed formula for

$$G_\mu(z) = \sum_{n=0}^\infty \frac{1}{z^{n+1}}\int_{\mathbb{R}} t^n\,\mu(\mathrm{d}t)$$

for large $|z|$, and then perform the analytic continuation to $\mathbb{C}^+$. As an example, one can calculate the Cauchy transform of the semicircle distribution $\mathrm{S}(0,v)$ with mean 0 and variance v as

$$\int_{-2\sqrt{v}}^{2\sqrt{v}} \frac{1}{z-t}\cdot\frac{\sqrt{4v-t^2}}{2\pi v}\,\mathrm{d}t = \frac{z-\sqrt{z^2-4v}}{2v}, \qquad v>0,\ z\in\mathbb{C}^+, \tag{4.40}$$

see [125, Lemma 2.21].

The pushforward of a symmetric distribution around the origin by the map $t \mapsto t^2$ can be calculated from the original Cauchy transform.

Proposition 4.50. *Let ν be a probability measure on $\mathbb{R}$ that is symmetric around 0, i.e., $\nu(B) = \nu(-B)$ holds for all $B \in \mathcal{B}(\mathbb{R})$. Let μ be the pushforward of ν by the map $t \mapsto t^2$. Then we have $z\tilde{G}_\mu(z^2) = G_\nu(z)$ on $\mathbb{C}^+$, where $\tilde{G}_\mu$ is the analytic continuation of G_μ to $\mathbb{C} \setminus [0, +\infty)$ given in Proposition 4.42.*

Proof. The desired formula follows from the straightforward calculations

$$\tilde{G}_\mu(z^2) = \int_{\mathbb{R}} \frac{1}{z^2 - t^2}\,\nu(\mathrm{d}t) = \frac{1}{2z}\int_{\mathbb{R}} \left(\frac{1}{z-t} + \frac{1}{z+t}\right)\nu(\mathrm{d}t) = \frac{1}{z}G_\nu(z). \qquad \square$$

Example 4.51. Let μ be the Marchenko–Pastur distribution

$$\mu(\mathrm{d}t) = \frac{1}{2\pi r}\sqrt{\frac{4r-t}{t}}\,\chi_{(0,4r)}(t)\,\mathrm{d}t, \qquad r > 0,$$

which is the pushforward of the semicircle distribution $\mathrm{S}(0, r)$ by the map $t \mapsto t^2$. Using (4.40) and Proposition 4.50 one obtains

$$G_\mu(z) = \frac{1}{\sqrt{z}} \cdot \frac{\sqrt{z} - \sqrt{z-4r}}{2r} = \frac{z - \sqrt{z^2 - 4rz}}{2rz}, \qquad z \in \mathbb{C}^+.$$

Yet another way is to establish a differential equation for the Cauchy transform by integration by parts.

Example 4.52. Let us consider $G(z) := G_{\mathrm{N}(0,1)}(z) = \int_{\mathbb{R}} \frac{1}{z-t}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,\mathrm{d}t$. By integration by parts we have

$$\begin{aligned} G'(z) &= -\int_{\mathbb{R}} \frac{1}{(z-t)^2} \cdot \frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,\mathrm{d}t = \int_{\mathbb{R}} \frac{1}{z-t} \cdot \frac{1}{\sqrt{2\pi}}(e^{-t^2/2})'\,\mathrm{d}t \\ &= \int_{\mathbb{R}} \frac{1}{z-t} \cdot \frac{1}{\sqrt{2\pi}}(z-t-z)e^{-t^2/2}\,\mathrm{d}t = 1 - zG(z). \end{aligned}$$

To solve the differential equation $G'(z) = 1 - zG(z)$, first we solve the homogeneous equation $H'(z) = -zH(z)$, which has a general solution

$$H(z) = C_1 e^{-z^2/2}.$$

Next we replace the constant C_1 with a function and set $G(z) := f(z)e^{-z^2/2}$. The equation $G'(z) = 1 - zG(z)$ yields $f'(z) = e^{z^2/2}$, so that

$$f(z) = C_2 + \int_{L_z} e^{w^2/2}\,\mathrm{d}w,$$

where C_2 is a constant and the line integral is performed over the half-line $L_z := \{z + \mathrm{i}y : -\infty < y < 0\}$ started at ∞ and terminated at z. To determine the constant C_2, let us consider

$$G(\mathrm{i}y) = e^{y^2/2}\left[C_2 + \mathrm{i}\int_{-\infty}^{y} e^{-t^2/2}\,\mathrm{d}t\right].$$

Since G is a Cauchy transform we must have $G(\mathrm{i}y) \to 0$ as $y \to +\infty$. This forces $C_2 = -\mathrm{i}\sqrt{2\pi}$, so that we obtain

$$G_{\mathrm{N}(0,1)}(z) = e^{-z^2/2}\left[-\mathrm{i}\sqrt{2\pi} + \int_{L_z} e^{w^2/2}\,\mathrm{d}w\right], \qquad z \in \mathbb{C}^+.$$

4.7. **Notes.** The proof of the Nevanlinna formula in Theorem 4.26 is based on the expositions by Akhiezer and Glazman [4, Section 59] and Bhatia [34, Chapter V.4]. The characterization of weak convergence in Proposition 4.36 is an extension of Maassen's result [109, Theorem 2.5]. The part of Proposition 4.39 that a tight family of probability measures satisfies the convergence $\sphericalangle \lim_{z\to\infty} F_\mu(z)/z = 1$ uniformly over μ was proved by Bercovici and Voiculescu in the remark following [32, Proposition 5.1].

The characterization of η-transform in Proposition 4.41 is adopted from [10, Proposition 3.2] and [11, Proposition 2.4]. The characterization of probability measures on $[0, +\infty)$ in terms of η_μ in Proposition 4.44 was given by Belinschi and Bercovici [20, Proposition 2.2] for the purpose of studying multiplicative free convolution. The characterization of probability measures on $[0, +\infty)$ in terms of F_μ in Proposition 4.43 is adopted from [80, Proposition 2.5]. Concerning the characterization of finite even moments, the equivalence (1) $\iff$ (2) in Proposition 4.46 is due to [3, Theorem 3.2.1]. Some of the results in this section can also be found in the books of Mingo and Speicher [113] and Schleissinger [138].

5. Analysis of monotone convolution

Let x, y be monotonically independent real random variables in a unital C^*-probability space. The reciprocal Cauchy transform of μ_x and the shifted moment generating function of x are connected as

$$F_{\mu_x}(z) = 1/M_x(1/z).$$

Accordingly, Theorem 1.27 can be written in the form $F_{\mu_{x+y}}(z) = F_{\mu_x}(F_{\mu_y}(z))$. From the complex-analytic perspective, we can extend the additive monotone convolution to any probability measures on $\mathbb{R}$.

Theorem 5.1. *Let μ, ν be probability measures on $\mathbb{R}$. Then there exists a unique probability measure λ on $\mathbb{R}$ such that*

$$F_\lambda(z) = F_\mu(F_\nu(z)), \qquad z \in \mathbb{C}^+.$$

The measure λ is denoted by $\mu \rhd \nu$ and is called the ***(additive) monotone convolution*** *of μ and ν.*

Proof. Note that $\Im[F_\nu(z)] \geq \Im z > 0$ and in particular $F_\nu(\mathbb{C}^+) \subseteq \mathbb{C}^+$. The function $N(z) := F_\mu(F_\nu(z))$ is therefore a Nevanlinna function. According to Proposition 4.37, it suffices to prove $\lim_{y\to+\infty} N(\mathrm{i}y)/(\mathrm{i}y) = 1$. Proposition 4.37 guarantees that $F_\nu(\mathrm{i}y) = \mathrm{i}y(1+o(1))$ as $y \to +\infty$. From this one can see that for any fixed $\gamma > 0$ there exists $y_0 > 0$ such that $F_\nu(\mathrm{i}y) \in \nabla_\gamma$ for all $y > y_0$. Therefore, by Proposition 4.37 (3), it holds that $\frac{F_\mu(F_\nu(\mathrm{i}y))}{F_\nu(\mathrm{i}y)} \to 1$ and hence

$$\frac{N(\mathrm{i}y)}{\mathrm{i}y} = \frac{F_\mu(F_\nu(\mathrm{i}y))}{F_\nu(\mathrm{i}y)} \cdot \frac{F_\nu(\mathrm{i}y)}{\mathrm{i}y} \to 1. \qquad \square$$

Proposition 5.2. *Let μ, μ_n, ν, ν_n $(n \in \mathbb{N})$ be probability measures on $\mathbb{R}$ such that $\mu_n \to \mu$ and $\nu_n \to \nu$ weakly. Then $\mu_n \rhd \nu_n \to \mu \rhd \nu$ weakly.*

Proof. By Proposition 4.36, it suffices to show the pointwise convergence $F_{\mu_n}(F_{\nu_n}(z)) \to F_\mu(F_\nu(z))$ on $\mathbb{C}^+$. For this we fix $z \in \mathbb{C}^+$, set $w_n := F_{\nu_n}(z)$ and $w := F_\nu(z)$. In the inequality

$$|F_{\mu_n}(w_n) - F_\mu(w)| \leq |F_{\mu_n}(w_n) - F_\mu(w_n)| + |F_\mu(w_n) - F_\mu(w)|,$$

the second term clearly converges to 0 because $w_n \to w$. The first term also converges to zero since the convergence $F_{\mu_n} \to F_\mu$ is locally uniform. $\square$

Example 5.3. The measure $\mu \rhd \delta_a$ is the translation of μ by a. Indeed, the fact $F_{\delta_a}(z) = 1/G_{\delta_a}(z) = z - a$ yields

$$G_{\mu\rhd\delta_a}(z) = G_\mu(F_{\delta_a}(z)) = G_\mu(z-a) = \int_{\mathbb{R}} \frac{1}{z-(t+a)}\,\mu(\mathrm{d}t), \tag{5.1}$$

showing that $\mu \rhd \delta_a$ is the pushforward of μ by the map $t \mapsto t + a$.

On the other hand, the measure $\delta_a \rhd \mu$ is not a translation of μ for generic $a \in \mathbb{R}$ and μ. For example, if μ is the arcsine law $\mathrm{A}(0, v)$ in Example 4.49, then $F_\mu(z) = \sqrt{z^2 - 2v}$ and so $F(z) := F_{\delta_a\rhd\mu}(z) = \sqrt{z^2-2v} - a$. Observe first that F has an analytic extension to $\mathbb{C} \setminus [-\sqrt{2v}, \sqrt{2v}]$, which we denote by the same symbol F. If $a > 0$, then F has a zero at $x = \sqrt{a^2+2v}$, while F has no zero on $(-\infty, -\sqrt{2v})$ as $F(-x) = -\sqrt{x^2-2v} - a < 0$ for $x > \sqrt{2v}$. In view of Proposition 4.42 (iv) and its proof, $\delta_a \rhd \mu$ has an atom at $\sqrt{a^2+2v}$ and its weight is $1/F'(\sqrt{a^2+2v}) = a/\sqrt{a^2+2v}$. By the Stieltjes inversion formula $\delta_a \rhd \mu$ has a Lebesgue absolutely continuous part on $[-\sqrt{2v}, \sqrt{2v}]$ with density

$$\frac{-1}{\pi}\Im\left[\frac{1}{\sqrt{(t+\mathrm{i}0)^2-2v}-a}\right] = \frac{-1}{\pi}\Im\left[\frac{1}{\mathrm{i}\sqrt{2v-t^2}-a}\right] = \frac{\sqrt{2v-t^2}}{\pi(a^2+2v-t^2)}.$$

By symmetry, a similar result holds for $a < 0$, and consequently, we obtain

$$\delta_a \rhd \mu = \frac{|a|}{\sqrt{a^2+2v}}\delta_{\mathrm{sign}(a)\sqrt{a^2+2v}} + \frac{\sqrt{2v-t^2}}{\pi(a^2+2v-t^2)}\chi_{(-\sqrt{2v},\sqrt{2v})}(t)\,\mathrm{d}t, \qquad a \in \mathbb{R},\ v > 0.$$

Example 5.4. Let μ_r be the Marchenko–Pastur distribution with scale parameter $r > 0$ in Example 4.51. For $a \in \mathbb{R}$ we have

$$F_{\delta_a\rhd\mu_r}(z) = \frac{z - 2a + \sqrt{(z-2r)^2 - 4r^2}}{2}.$$

In particular, $\delta_{2r} \rhd \mu_r$ is the semicircle distribution with mean $2r$ and variance r^2:

$$(\delta_{2r} \rhd \mu_r)(\mathrm{d}t) = \frac{1}{2\pi r^2}\sqrt{4r^2 - (t-2r)^2}\,\chi_{[0,4r]}(t)\,\mathrm{d}t;$$

see (4.40) and (5.1).

In a similar vein, we can define multiplicative monotone convolution by extending the formula $\eta_{\sqrt{x}y\sqrt{x}}(z) = \eta_x(\eta_y(z))$ in Theorem 1.29.

Theorem 5.5. *Let μ, ν be probability measures on $[0, +\infty)$ and on $\mathbb{R}$, respectively. Suppose that $\nu \neq \delta_0$. Let $\tilde{\eta}_\mu$ denote the analytic continuation of η_μ to $\mathbb{C} \setminus [0, +\infty)$ as given in Proposition 4.44. Then there exists a unique probability measure λ on $\mathbb{R}$ such that*

$$\eta_\lambda(z) = \tilde{\eta}_\mu(\eta_\nu(z)), \qquad z \in \mathbb{C}^+.$$

The measure λ is denoted by $\mu \circlearrowright \nu$ and is called the ***multiplicative monotone convolution*** *of μ and ν. We also define $\mu \circlearrowright \delta_0 := \delta_0$.*

Proof. If $\mu = \delta_0$ then $\eta_\lambda = 0$ and so $\lambda = \delta_0$. We therefore assume $\mu \neq \delta_0$. Note that the property $\eta_\nu(\mathbb{C}^+) \subseteq \mathbb{C} \setminus [0, +\infty)$ in Proposition 4.41 implies that the composition $\tilde{\eta}_\mu \circ \eta_\nu$ is well defined and is holomorphic on $\mathbb{C}^+$. Moreover, if $\arg \eta_\nu(z) \in [\arg z, \pi]$ then, by Proposition 4.44, $\pi \geq \arg \tilde{\eta}_\mu(\eta_\nu(z)) \geq \arg \eta_\nu(z) \geq \arg z$. If $\arg \eta_\nu(z) \in (\pi, \arg z + \pi]$ then by the symmetry $\tilde{\eta}_\mu(\overline{z}) = \overline{\tilde{\eta}_\mu(z)}$ we deduce that $\pi \leq \arg \tilde{\eta}_\mu(\eta_\nu(z)) \leq \arg \eta_\nu(z) \leq \arg z + \pi$. In any case the inequality $\arg z \leq \arg \eta_\lambda(z) \leq \arg z + \pi$ holds on $\mathbb{C}^+$, and so condition (ii) in Proposition 4.41 holds. It remains to check condition (iii) in Proposition 4.41. As $z \to 0$ within $z \in \nabla_\gamma$, $\eta_\nu(z)$ converges to 0, and moreover, the inequality $\arg z \leq \arg \eta_\nu(z) \leq \arg z + \pi$ implies $\arg \eta_\nu(z) \in (\theta, 2\pi - \theta)$ for some $\theta \in (0, \pi)$. Therefore, using Proposition 4.44 (iii) yields $\tilde{\eta}_\mu(\eta_\nu(z)) \to 0$.

The uniqueness of λ is a consequence of the uniqueness result in Proposition 4.37 and the relation $\eta_\lambda(z) = 1 - zF_\lambda(1/z)$ noted in (4.19). $\square$

Although $\eta_{\delta_0}(z) \equiv 0$ is not contained in the domain of $\tilde{\eta}_\mu$, the above exceptional definition $\mu \circlearrowright \delta_0 := \delta_0$ is natural because multiplicative monotone convolution comes from the distribution of $\sqrt{x}y\sqrt{x}$ and $\nu = \delta_0$ corresponds to $y = 0$. This definition can also be justified from the perspective of continuity.

Proposition 5.6. *Let μ, μ_n $(n \in \mathbb{N})$ be probability measures on $[0, +\infty)$ and ν, ν_n $(n \in \mathbb{N})$ be probability measures on $\mathbb{R}$ such that $\mu_n \to \mu$ and $\nu_n \to \nu$ weakly. Then $\mu_n \circlearrowright \nu_n \to \mu \circlearrowright \nu$ weakly.*

Proof. Observe first that the weak convergence of measures is equivalent to the locally uniform convergence of η-transforms due to (4.19) and Proposition 4.36.

In case $\nu \neq \delta_0$, the proof is the same as Proposition 5.2 because $\eta_\nu(z)$ belongs to $\mathbb{C} \setminus [0,+\infty)$ that is the domain of $\tilde{\eta}_{\mu_n}$ and $\tilde{\eta}_\mu$.

The case $\nu = \delta_0$ needs to be handled separately. In this case, we first extend Proposition 4.39 as follows: for a tight family $\mathscr{P}$ of probability measures on $[0,+\infty)$, it holds that

$$\lim_{\substack{z\to\infty \\ \theta<\arg z<2\pi-\theta}} \sup_{\mu'\in\mathscr{P}} \left|\frac{\tilde{F}_{\mu'}(z)}{z} - 1\right| = 0 \tag{5.2}$$

for each $\theta \in (0,\pi)$. The proof is almost the same; the required modification is that inequality (4.13) holds for all z with $\arg z \in (\theta, 2\pi-\theta)$ and all $t \geq 0$ where $\theta := \arctan\gamma$ and that $|t/(z-t)|$ can be bounded by $t/|z|$ for $\Re(z) < 0$ instead of $t/\Im(z)$.

In our situation, since the weakly convergent family $\{\mu_n : n \in \mathbb{N}\}$ is tight, (5.2) yields

$$\lim_{\substack{z\to 0 \\ \theta<\arg z<2\pi-\theta}} \sup_{n\in\mathbb{N}} |\tilde{\eta}_{\mu_n}(z)| = 0.$$

This estimate and the fact that $\eta_{\nu_n}(z) \to 0$ and $\arg z \leq \arg \eta_{\nu_n}(z) \leq \arg z + \pi$ (if $\nu_n \neq \delta_0$) imply that $\tilde{\eta}_{\mu_n}(\eta_{\nu_n}(z)) \to 0$ as $n \to \infty$ for each $z \in \mathbb{C}^+$. This further implies the weak convergence $\mu_n \circlearrowright \nu_n \to \delta_0$ by Proposition 4.36. □

Example 5.7. In a similar way to Example 5.3, $\mu \circlearrowright \delta_b$ is the dilation of μ by $b \in \mathbb{R}$, i.e., the pushforward of the measure μ by the map $x \mapsto bx$. On the other hand, $\delta_a \circlearrowright \nu$ is different from the dilation for generic $a > 0$ and ν.

In the rest of this section and also in Sections 6 and 7, we focus on additive monotone convolution and simply call it monotone convolution. In many cases similar results can be obtained for multiplicative monotone convolution but are omitted. We come back to multiplicative monotone convolution in Section 8 to explore the eigenvalues of perturbed random matrices.

5.1. Support and moments for monotone convolution.

We study support and moments of monotone (additive) convolution of probability measures.

Definition 5.8. Let S, T be topological spaces. A **Borel kernel** from S to T is a function $\tau\colon S \times \mathcal{B}(T) \to [0,+\infty]$ such that

- $B \mapsto \tau(s,B)$ is a Borel measure for every $s \in S$,
- $s \mapsto \tau(s,B)$ is measurable for every $B \in \mathcal{B}(T)$.

If $S = T$ we simply say τ is a Borel kernel on S. If each $\tau(s,\cdot)$ is a probability measure on T, we call τ a **probability kernel** (from S to T).

The following fact is well known in the theory of Markov processes.

Lemma 5.9. *Let R, S, T be topological spaces. Let ρ be a Borel measure on S, σ be a Borel kernel from R to S and τ be a Borel kernel from S to T. Then the* ***compositions***‡

$$(\rho\tau)(B) := \int_S \rho(\mathrm{d}s)\tau(s,B), \qquad B \in \mathcal{B}(T), \tag{5.3}$$

$$(\sigma\tau)(r,B) := \int_S \sigma(r,\mathrm{d}s)\tau(s,B), \qquad r \in R,\ B \in \mathcal{B}(T), \tag{5.4}$$

define a Borel measure on T and a Borel kernel from R to T, respectively. Moreover, for every measurable function $f\colon S \to [0,+\infty]$,

$$\int_T (\rho\tau)(\mathrm{d}t)f(t) = \int_S \rho(\mathrm{d}s)\left[\int_T \tau(s,\mathrm{d}t)f(t)\right], \tag{5.5}$$

$$\int_T (\sigma\tau)(r,\mathrm{d}t)f(t) = \int_S \sigma(r,\mathrm{d}s)\left[\int_T \tau(s,\mathrm{d}t)f(t)\right], \qquad r \in R. \tag{5.6}$$

Proof. Since $\sigma\tau$ is more general, we only discuss $\sigma\tau$. The σ-additivity of $\sigma\tau$ for the second component follows from the monotone convergence theorem. For the measurability $r \mapsto (\sigma\tau)(r,B)$, we return to the definition of integration and take a sequence of nonnegative simple functions $f_n(s) \uparrow \tau(s,B)$ of the form $f_n(s) = \sum_{i=1}^{m_n} a_{n,i}\chi_{A_{n,i}}(s)$. Then the function $r \mapsto \int_S \sigma(r,\mathrm{d}s) f_n(s) = \sum_{i=1}^{m_n} a_{n,i}\sigma(r,A_{n,i})$ is measurable. Therefore its limit $\int_S \sigma(r,\mathrm{d}s)\tau(s,B)$ is also a measurable function of r.

Formula (5.6) is obvious for linear combinations of indicator functions, and then extends to general f's by the monotone convergence theorem. □

Lemma 5.10. *Let S be a topological space. Let $(\tau_s)_{s\in S}$ be a family of finite Borel measures on $\mathbb{R}$. Then $\tau(s,B) := \tau_s(B)$ is a Borel kernel from S to $\mathbb{R}$ if and only if $s \mapsto G_{\tau_s}(z)$ is measurable for every $z \in \mathbb{C}^+$.*

Proof. If τ is a Borel kernel, then the definition of integral (i.e., approximating the integrand $1/(z-x)$ by simple functions) implies the measurability of $s \mapsto G_{\tau_s}(z)$. Conversely, if $s \mapsto G_{\tau_s}(z)$ is measurable then the Stieltjes inversion formula (Proposition 4.33) implies that $s \mapsto \tau_s(I)$ is measurable for each finite open interval I. By taking the limit, the same holds for infinite intervals I. Since each $\tau(s,\cdot)$ is a finite measure, the set

$$\mathcal{L} := \{B \in \mathcal{B}(\mathbb{R}) : s \mapsto \tau(s,B) \text{ is measurable}\}$$

is a λ-system and contains the π-system $\mathcal{P}$ of all open intervals of $\mathbb{R}$. By the π-λ theorem (Theorem 4.2), $\mathcal{L}$ contains $\sigma(\mathcal{P})$ that coincides with $\mathcal{B}(\mathbb{R})$. Therefore, $s \mapsto \tau(s,B)$ is measurable for any $B \in \mathcal{B}(\mathbb{R})$. □

‡In mathematics, it is more common to write an integrand in front of a measure; however, when discussing composition, the notation such as (5.3) and (5.4) is convenient, see also Section 7.2.

Proposition 5.11. *Let μ, ν be probability measures on $\mathbb{R}$. For $x \in \mathbb{R}$ we define*

$$\tilde{\nu}(x, \mathrm{d}y) := (\delta_x \rhd \nu)(\mathrm{d}y).$$

Then $\tilde{\nu}(\cdot,\cdot)$ is a probability kernel on $\mathbb{R}$ and the identity $\mu \rhd \nu = \mu\tilde{\nu}$ holds, i.e.,

$$(\mu \rhd \nu)(B) = \int_{\mathbb{R}} \mu(\mathrm{d}x)\tilde{\nu}(x, B), \qquad B \in \mathcal{B}(\mathbb{R}). \tag{5.7}$$

Proof. Since $x \mapsto G_{\tilde{\nu}(x,\cdot)}(z) = G_{\delta_x}(F_\nu(z)) = \frac{1}{F_\nu(z)-x}$ is measurable in x, Lemma 5.10 ensures that $\tilde{\nu}(\cdot,\cdot)$ is a probability kernel. Let us denote $\lambda := \mu\tilde{\nu}$. For $z \in \mathbb{C}^+$, we apply formula (5.5) to $f(y) := 1/(z-y)$ (by decomposing f as $f = g_+ - g_- + \mathrm{i}(h_+ - h_-)$) to obtain

$$\begin{aligned} G_\lambda(z) &= \int_{\mathbb{R}} \frac{\lambda(\mathrm{d}y)}{z-y} = \int_{\mathbb{R}} \mu(\mathrm{d}x) \left[\int_{\mathbb{R}} \frac{\tilde{\nu}(x,\mathrm{d}y)}{z-y}\right] \\ &= \int_{\mathbb{R}} \frac{\mu(\mathrm{d}x)}{F_{\tilde{\nu}(x,\cdot)}(z)} = \int_{\mathbb{R}} \frac{\mu(\mathrm{d}x)}{F_\nu(z)-x} = G_\mu(F_\nu(z)), \end{aligned}$$

so that $F_\lambda = F_\mu \circ F_\nu$. This implies $\lambda = \mu \rhd \nu$. □

The probability kernel in Proposition 5.11 is helpful for studying monotone convolution, in particular to see how a property of $\mu \rhd \nu$ is inherited to μ and ν. Below we study the support and moments of monotone convolution. Another remarkable aspect is a connection of this probability kernel to a certain Markov process, which will be explored in Section 7.2.

Proposition 5.12. *Let μ, ν be probability measures on $\mathbb{R}$. Then $\mu \rhd \nu$ has compact support if and only if both μ, ν have compact support. Moreover, if $\mu \rhd \nu$ is supported on $[-R, R]$ then F_ν extends analytically to $\mathbb{C} \setminus [-R, R]$ with $F_\nu(\overline{z}) = \overline{F_\nu(z)}$.*

Proof. If μ and ν have compact support, then there are monotonically independent real random variables x and y in a unital C^*-probability space such that the analytic distributions of x and y are μ and ν, respectively. Since the distribution of $x + y$ coincides with $\mu \rhd \nu$, it has compact support.

Conversely, suppose that $\mu \rhd \nu$ has compact support. Let $\tilde{\nu}$ be the probability kernel defined in Proposition 5.11. There exists some $B = \{x \in \mathbb{R} : |x| > R\}$ such that

$$0 = (\mu \rhd \nu)(B) = \int_{\mathbb{R}} \mu(\mathrm{d}x)\tilde{\nu}(x, B),$$

and so $\tilde{\nu}(x, B) = 0$ for μ-a.e. $x \in \mathbb{R}$. Pick such an $x \in \mathbb{R}$ that $\tilde{\nu}(x, B) = 0$. Then $F_{\tilde{\nu}(x,\cdot)}$ has an analytic continuation to $\mathbb{C} \setminus [-R, R]$ such that $F_{\tilde{\nu}(x,\cdot)}(\overline{z}) = \overline{F_{\tilde{\nu}(x,\cdot)}(z)}$, and the same holds for $F_\nu(z) = F_{\tilde{\nu}(x,\cdot)}(z) + x$. By Proposition 4.42, ν is supported on some compact interval.

To show that μ is compactly supported, it suffices by symmetry to show that the support of μ is bounded from above. We use the expansion in Proposition 4.42 (with the same symbol F_μ for simplicity)

$$F_\nu(z) = z - \sum_{n=1}^{\infty} \frac{b_n}{z^{n-1}}, \qquad |z| > R,$$

which implies $F_\nu(z) = z + O(1)$ as $z \to \infty$. Hence, there is $c > 0$ such that for all $y \in \mathbb{R}$ with $y \geq R + 1$ one has $y - c < F_\nu(y) < y + c$. Since $F_\nu'(y) \geq 1$ on $(R, +\infty)$ (see (4.24)), this implies that for each x with $x \geq R + c + 1$ the function $F_{\tilde{\nu}(x,\cdot)}(y) = F_\nu(y) - x$ has a unique zero y_x on $(x - c, x + c)$. This means that $\tilde{\nu}(x, \cdot)$ has an atom at y_x, so that $\tilde{\nu}(x, (x - c, x + c)) > 0$. Since for $x \geq R + c + 1$ we have $(x - c, x + c) \subseteq B$, so that $\tilde{\nu}(x, B) > 0$. However, $\tilde{\nu}(x, B) = 0$ for μ-a.e. x, i.e., there is an $S \in \mathcal{B}(\mathbb{R})$ such that $\mu(S) = 1$ and $\tilde{\nu}(x, B) = 0$ for all $x \in S$. Therefore, $S \cap (R + c + 1, +\infty) = \emptyset$, showing that the support of μ is bounded from above. □

We extend the combinatorial formula (1.10) for the moments of the sum $x + y$ to the monotone convolution of probability measures.

Proposition 5.13. *Let μ and ν be probability measures on $\mathbb{R}$ and let $n \in \mathbb{N}$. Then the $2n$th moment $\int_{\mathbb{R}} t^{2n}\,(\mu \rhd \nu)(\mathrm{d}t)$ is finite if and only if both $\int_{\mathbb{R}} t^{2n}\,\mu(\mathrm{d}t)$ and $\int_{\mathbb{R}} t^{2n}\,\nu(\mathrm{d}t)$ are finite. Moreover, if all these $2n$th moments are finite, then we have*

$$\mathrm{m}_p(\mu \rhd \nu) = \sum_{\ell=0}^{p} \sum_{\substack{k_0,k_1,\dots,k_\ell \geq 0, \\ k_0+k_1+\cdots+k_\ell = p-\ell}} \mathrm{m}_\ell(\mu)\mathrm{m}_{k_0}(\nu)\mathrm{m}_{k_1}(\nu)\cdots\mathrm{m}_{k_\ell}(\nu), \qquad 1 \leq p \leq 2n. \tag{5.8}$$

In particular,

$$\mathrm{m}_1(\mu \rhd \nu) = \mathrm{m}_1(\mu) + \mathrm{m}_1(\nu), \tag{5.9}$$

$$\mathrm{Var}(\mu \rhd \nu) = \mathrm{Var}(\mu) + \mathrm{Var}(\nu). \tag{5.10}$$

Proof. First we assume that $\int_{\mathbb{R}} t^{2n}\,(\mu \rhd \nu)(\mathrm{d}t)$ is finite. By Proposition 5.11 and Lemma 5.9, we obtain, with notation $\tilde{\nu}(x, \cdot) = (\delta_x \rhd \nu)(\cdot)$,

$$\int_{\mathbb{R}} t^{2n}\,(\mu \rhd \nu)(\mathrm{d}t) = \int_{\mathbb{R}} \mu(\mathrm{d}x) \left[\int_{\mathbb{R}} t^{2n}\,\tilde{\nu}(x, \mathrm{d}t)\right],$$

which implies $\int_{\mathbb{R}} t^{2n}\,\tilde{\nu}(x, \mathrm{d}t) < +\infty$ for μ-a.e. x. We choose such an x. According to Proposition 4.46, $F_{\tilde{\nu}(x,\cdot)}(z)$ has an expansion of the form (4.29), and the same applies to $F_\nu(z) = F_{\tilde{\nu}(x,\cdot)}(z) + x$. Thus we obtain $\int_{\mathbb{R}} t^{2n}\,\nu(\mathrm{d}t) < +\infty$. For the finiteness of $\int_{\mathbb{R}} t^{2n}\,\mu(\mathrm{d}t)$ we use an asymptotic expansion of the inverse function of F_ν. Because we do not use this part later and the proof is rather long, the proof is postponed to Appendix B.

Next we assume that $\int_{\mathbb{R}} t^{2n}\,\mu(\mathrm{d}t)$ and $\int_{\mathbb{R}} t^{2n}\,\nu(\mathrm{d}t)$ are finite. As discussed in the proof of Theorem 5.1, for any fixed $\gamma > 0$ we have $F_\nu(\mathrm{i}y) \in \nabla_\gamma$ for all sufficiently large y. By (4.29) on the domain ∇_γ, we obtain

$$F_\mu(F_\nu(\mathrm{i}y)) = F_\nu(\mathrm{i}y) - b_1 - b_2 G_\nu(\mathrm{i}y) - \cdots - b_{2n} G_\nu(\mathrm{i}y)^{2n-1} + o(|F_\nu(\mathrm{i}y)|^{-(2n-1)}). \tag{5.11}$$

The remainder term above equals $o(y^{-(2n-1)})$ as $F_\nu(\mathrm{i}y) = \mathrm{i}y(1+o(1))$. Expanding $G_\nu(\mathrm{i}y)$ and $F_\nu(\mathrm{i}y)$ in the forms (4.28) and (4.29) respectively, substituting them into (5.11) and recollecting terms shows that for some reals $c_1, c_2, \dots, c_{2n}$

$$F_\mu(F_\nu(z)) = z - c_1 - \frac{c_2}{z} - \cdots - \frac{c_{2n}}{z^{2n-1}} + o(|z|^{-(2n-1)})$$

for $z = \mathrm{i}y$ as $y \to +\infty$. Proposition 4.46 ensures that the $2n$th moment of $\mu \rhd \nu$ is finite.

Finally, formula (5.8) is obtained by expanding the right-hand side of $G_{\mu\rhd\nu}(z) = G_\mu(1/G_\nu(z))$, which is just tracing the calculations (1.11)–(1.15) backwards, where the infinite sum is to be replaced with truncated finite sum with remainder terms and $\varphi(x^\ell), \varphi(y^k)$ are to be replaced with $\mathrm{m}_\ell(\mu), \mathrm{m}_k(\nu)$, respectively. □

5.2. **Convolution semigroups.** In probability theory, time-homogeneous random walk is the sum of independent, identically distributed random variables $(X_n)_{n\ge1}$:

$$S_n := X_1 + X_2 + \cdots + X_n; \qquad S_0 := 0.$$

The distribution μ_n of S_n is the n-fold convolution $\mu * \mu * \cdots * \mu$, where μ is the distribution of X_1. Obviously, we have $\mu_m * \mu_n = \mu_{m+n}$ and $\mu_0 = \delta_0$. A continuous-time analogue of random walk is called a Lévy process, which is characterized by a convolution semigroup $(\mu_t)_{t\ge0}$, i.e., μ_t $(t \ge 0)$ are probability measures on $\mathbb{R}$ such that $\mu_{s+t} = \mu_s * \mu_t$, $s,t \ge 0$, $\mu_0 = \delta_0$ and $t \mapsto \mu_t$ is weakly continuous.

We consider a monotone analogue of Lévy processes. The process itself will be explored later in Section 7. Here we investigate the distributional properties of monotone convolution semigroups with complex-analytic meiiimpthods.

Definition 5.14. A family $(\mu_t)_{t\ge0}$ of probability measures on $\mathbb{R}$, indexed by nonnegative reals t, is called a **monotone convolution semigroup** if

(i) $t \mapsto \mu_t$ is weakly continuous, i.e., for every bounded continuous function f on $\mathbb{R}$, the function $t \mapsto \int_{\mathbb{R}} f(x)\mu_t(\mathrm{d}x)$ is continuous.

(ii) $\mu_{s+t} = \mu_s \rhd \mu_t$ for all $s,t \ge 0$,

(iii) $\mu_0 = \delta_0$, where δ_0 is the delta measure at 0.

In the study of monotone convolution semigroups, a fundamental result is the description of infinitesimal generators. The proof relies heavily on Berkson–Porta's work summarized in Appendix C.

Theorem 5.15. *Let $(\mu_t)_{t\ge0}$ be a monotone convolution semigroup. Let F_t be the reciprocal Cauchy transform of μ_t. Then the limit*

$$A(z) := \lim_{t\to0^+} \frac{F_t(z) - z}{t} \tag{5.12}$$

exists locally uniformly on $\mathbb{C}^+$ and it holds that

(i) *A is a Nevanlinna function such that $\sphericalangle \lim_{z\to\infty} A(z)/z = 0$,*

(ii) $\dfrac{\mathrm{d}}{\mathrm{d}t}F_t(z) = A(F_t(z)), \quad t \ge 0,\ z \in \mathbb{C}^+$, *where* $\dfrac{\mathrm{d}}{\mathrm{d}t}$ *at $t = 0$ is to be interpreted as the right-derivative.*

Conversely, given a function A satisfying condition (i) *above, then equation* (ii) *with initial condition $F_0 = \mathrm{id}_{\mathbb{C}^+}$ has a unique solution $(F_t)_{t\ge0}$ which consists of the reciprocal Cauchy transforms of a monotone convolution semigroup $(\mu_t)_{t\ge0}$.*

*The function A is called the **infinitesimal generator** of (F_t) and also of (μ_t).*

Proof. According to Proposition 4.36, the weak continuity $t \mapsto \mu_t$ ensures the continuity $t \mapsto F_t$ with respect to the locally uniform convergence. The existence of the limit (5.12), the holomorphicity of A, and the differential equation in (ii) will be proven in Theorem C.2. The inequality $\Im[F_{\mu_t}(z)] \ge \Im z$, which follows by (4.17), implies A is a Nevanlinna function.

It remains to show $\sphericalangle \lim_{z\to\infty} A(z)/z = 0$. To begin with, integrating equation (ii) yields

$$F_t(z) = z + \int_0^t A(F_s(z))\,\mathrm{d}s, \qquad z \in \mathbb{C}^+. \tag{5.13}$$

Let $a := \sphericalangle \lim_{z\to\infty} A(z)/z$. Since $s \mapsto \mu_s$ is weakly continuous, for a fixed $t > 0$, the family $\{\mu_s : 0 \le s \le t\}$ is tight. By Proposition 4.39, $\sup_{s\in[0,t]} |F_s(\mathrm{i}y) - \mathrm{i}y| = o(y)$ as $y \to +\infty$. This in particular implies that for any $\gamma > 0$, there is $y_0 > 0$ such that $F_s(\mathrm{i}y) \in \nabla_\gamma$ for all $y > y_0$ and $s \in [0,t]$. Therefore,

$$\sup_{s\in[0,t]} \left| \frac{A(F_s(\mathrm{i}y))}{\mathrm{i}y} - a \right| = \sup_{s\in[0,t]} \left| \frac{A(F_s(\mathrm{i}y))}{F_s(\mathrm{i}y)} \cdot \frac{F_s(\mathrm{i}y)}{\mathrm{i}y} - a \right| \to 0 \quad \text{as} \quad y \to +\infty.$$

Dividing (5.13) by $z = \mathrm{i}y$ and using the obtained uniform convergence, we obtain

$$\frac{F_t(\mathrm{i}y)}{\mathrm{i}y} = 1 + \int_0^t \frac{A(F_s(\mathrm{i}y))}{\mathrm{i}y}\,\mathrm{d}s \to 1 + ta \quad \text{as} \quad y \to +\infty.$$

Since $F_t(\mathrm{i}y)/(\mathrm{i}y) \to 1$, a must be zero.

Conversely, if A is a function described in (i), then Theorem C.3 (note that $A \in \mathscr{G}_1(\mathbb{C}^+)$) guarantees that the equation in (ii) has a unique solution $(F_t)_{t\ge0}$ that forms a one-parameter semigroup of holomorphic self-maps of $\mathbb{C}^+$. To show each F_t is the reciprocal Cauchy transform of a probability measure, it suffices to show $F_t(\mathrm{i}y)/(\mathrm{i}y) \to 1$. Let us use the PDE

$\frac{\partial}{\partial t}F_t(z) = A(z)\frac{\partial}{\partial z}F_t(z)$ that arises by taking the derivative $\mathrm{d}/\mathrm{d}s|_{s=0}$ of the relation $F_{t+s}(z) = F_t(F_s(z))$. The integrated form reads

$$\frac{F_t(z)}{z} = 1 + \frac{A(z)}{z}\int_0^t \frac{\partial F_s}{\partial z}(z)\,\mathrm{d}s. \tag{5.14}$$

By the assumption on A we know that $A(\mathrm{i}y)/(\mathrm{i}y) \to 0$ as $y \to +\infty$. To estimate the integral part, let us write the Nevanlinna formula

$$F_s(z) = a_s z - b_s + \int_{\mathbb{R}} \frac{1+xz}{x-z}\,\tau_s(\mathrm{d}x).$$

Using the continuous function $f(s) := \Im[F_s(\mathrm{i})] = a_s + \tau_s(\mathbb{R})$ we obtain the bound

$$\begin{aligned}\left|\frac{\partial F_s}{\partial z}(\mathrm{i}y)\right| &= \left|a_s + \int_{\mathbb{R}} \frac{1+x^2}{(x-\mathrm{i}y)^2}\,\tau_s(\mathrm{d}x)\right| \\ &\le a_s + \int_{\mathbb{R}} \frac{1+x^2}{x^2+y^2}\,\tau_s(\mathrm{d}x) \le a_s + \tau_s(\mathbb{R}) = f(s), \qquad s \ge 0,\ y \ge 1.\end{aligned} \tag{5.15}$$

Combining this estimate and (5.14) yields

$$\left|\frac{F_t(\mathrm{i}y)}{\mathrm{i}y} - 1\right| \le \left|\frac{A(\mathrm{i}y)}{\mathrm{i}y}\right| \left|\int_0^t f(s)\,\mathrm{d}s \to 0 \quad \text{as} \quad y \to +\infty,\right.$$

thereby $F_t = F_{\mu_t}$ for some probability measure μ_t on $\mathbb{R}$. The continuity of $t \mapsto F_t(z)$, the semigroup relation $F_{s+t} = F_s \circ F_t$, and the initial condition $F_0 = \mathrm{id}_{\mathbb{C}^+}$ imply that $(\mu_t)_{t\ge 0}$ is a monotone convolution semigroup. ☐

Remark 5.16. When proving $a = 0$ in the first part of the above proof, we could also use formula (5.14) instead of (5.13).

The Nevanlinna formula for $A(z)$ in Theorem 5.15 is of the form

$$A(z) = -\gamma + \int_{\mathbb{R}} \frac{1+xz}{x-z}\,\sigma(\mathrm{d}x), \tag{5.16}$$

where $\gamma \in \mathbb{R}$ and σ is a finite Borel measure on $\mathbb{R}$. In this case we write $A(z) = A^{(\gamma,\sigma)}(z)$. This integral formula is referred to as the **monotone Lévy–Khintchine representation** of $(\mu_t)_{t\ge 0}$.

Example 5.17. In the following examples of A, we can explicitly solve the complex ODE in Theorem 5.15 (ii). The function w^β below is defined on $\mathbb{C} \setminus [0,+\infty)$ as $|w|^\beta e^{\mathrm{i}\beta \arg w}$ in such a way that $\arg w \in (0, 2\pi)$.

(a) Let $A(z) = A^{a,0}(z) = -a$ with $a \in \mathbb{R}$. Then $F_t(z) = z - at$ and $\mu_t = \delta_{at}$.

(b) Let $A(z) = -a + \mathrm{i}b$ with $a \in \mathbb{R}$ and $b > 0$. Then $F_t(z) = z - (a - \mathrm{i}b)t$ and the measure μ_t is the Cauchy distribution

$$\mu_t(\mathrm{d}x) = \frac{1}{\pi} \cdot \frac{bt}{(x-at)^2 + (bt)^2}\,\mathrm{d}x, \qquad t > 0,$$

see Example 4.48. Note that the measure σ for A can be computed from the Stieltjes inversion (see Theorem 4.26) and we obtain $\sigma(\mathrm{d}x) = b[\pi(1+x^2)]^{-1}\,\mathrm{d}x$.

(c) Let $A(z) = A^{0,\delta_0}(z) = -\frac{1}{z}$. Then $F_t(z) = \sqrt{z^2 - 2t}$. The measure μ_t is the arcsine law $\mathrm{A}(0,t)$, i.e.,

$$\mu_t(\mathrm{d}x) = \frac{1}{\pi\sqrt{2t - x^2}}\,\chi_{(-\sqrt{2t},\sqrt{2t})}(x)\,\mathrm{d}x, \qquad t > 0.$$

see Example 4.49. This measure has appeared in the monotone CLT, see Theorem 3.18.

(d) Let $A(z) = -\frac{1}{z-a}$ with $a \in \mathbb{R}$. Then $F_t(z) = a + \sqrt{(z-a)^2 - 2t}$. The corresponding measure μ_t equals $\delta_{-a} \rhd \mathrm{A}(0,t) \rhd \delta_a$. From Example 5.3 we obtain

$$\mu_t(\mathrm{d}x) = \frac{|a|}{\sqrt{a^2+2t}}\delta_{a - \mathrm{sign}(a)\sqrt{a^2+2t}} + \frac{\sqrt{2t - (x-a)^2}}{\pi(a^2 + 2t - (x-a)^2)}\,\chi_{(a-\sqrt{2t},a+\sqrt{2t})}(x)\,\mathrm{d}x, \qquad t > 0.$$

This measure belongs to a so-called *free Meixner family* and appears in a limit theorem on a (weakly) monotone Fock space, see [57].

(e) Let $A(z) = e^{\mathrm{i}\alpha\rho\pi} z^{1-\alpha}$, where $\alpha \in (0,2]$ and $\rho \in [0,1] \cap [1 - 1/\alpha, 1/\alpha]$. Then F_t is given by $F_t(z) = (z^\alpha + te^{\mathrm{i}\alpha\rho\pi})^{1/\alpha}$. The corresponding distribution μ_t is called a monotone stable distribution. This monotone convolution semigroup is characterized by the condition $D_c(\mu_t) = \mu_{c^\alpha t}$ for all $c, t > 0$, where $D_c(\nu)$ is the push-forward of ν by the map $x \mapsto cx$. In the case $\alpha = 2$, μ_t is the arcsine law $\mathrm{A}(0, t/2)$, and in the case $\alpha = 1$, μ_t is a Cauchy distribution. This distribution for $\alpha \ne 2$ appears in a generalization of the monotone CLT of the form $(x_1 + x_2 + \cdots + x_N)/a_N$ where $a_N > 0$ is a deterministic sequence going to $+\infty$ and $(x_i)_{i=1}^\infty$ is monotonically iid but x_i has infinite second moment [152]. See [88] and references therein for further information.

In what follows, we characterize the monotone convolution semigroups having compact support and finite moments up to an even order. We also clarify the connection between convolution semigroups and monotone cumulants defined in Remark 3.13.

Proposition 5.18. *Let $(\mu_t)_{t\ge 0}$ be a monotone convolution semigroup with infinitesimal generator $A^{(\gamma,\sigma)}$. The following are equivalent:*

(1) *μ_t has compact support at every $t \ge 0$;*

(2) *μ_t has compact support at some $t > 0$;*

(3) *σ has compact support.*

If the above conditions are fulfilled then there are constants $C_1, C_2 > 0$ such that μ_t is supported on $[-C_1 - C_2 t, C_1 + C_2 t]$ for all $t \ge 0$.

Proof. (1) $\Longrightarrow$ (2) is obvious.

(2) $\Longrightarrow$ (3). Suppose that μ_{t_0} is supported on a compact interval $[-R, R]$ for some $t_0 > 0$ and $R > 0$. For each $t \in (0, t_0)$, μ_t is also compactly supported according to Proposition 5.12 applied to $\mu_{t_0} = \mu_{t_0 - t} \rhd \mu_t$. Moreover, the same proposition shows $F_t := F_{\mu_t}$ has an analytic continuation to $\mathbb{C} \setminus [-R, R]$ taking real values on $\mathbb{R} \setminus [-R, R]$, and the same holds for the Nevanlinna function $A_t(z) := [F_t(z) - z]/t$, $t > 0$. The latter means that the Nevanlinna formula of A_t is of the form

$$A_t(z) = -\gamma_t + \int_{\mathbb{R}} \frac{1 + xz}{x - z}\, \sigma_t(\mathrm{d}x), \qquad z \in \mathbb{C}^+,$$

where σ_t is supported on $[-R, R]$. According to Theorem 5.15, the function A_t converges to $A^{(\gamma,\sigma)}$ on $\mathbb{C}^+$ as $t \to 0^+$. By Proposition 4.29, the extended measure $\widehat{\sigma}_t$ on $\widehat{\mathbb{R}}$ with zero mass at ∞ converges weakly to a finite measure $\widehat{\sigma}$ on $\widehat{\mathbb{R}}$. Since the measure $\widehat{\sigma}_t$ is uniformly supported on the compact interval $[-R, R]$, the original measure σ_t also converges weakly to a finite measure on $\mathbb{R}$. The limit measure coincides with σ as A_t converges to $A^{(\gamma,\sigma)}$. This implies that σ is also supported on $[-R, R]$.

(3) $\Longrightarrow$ (1). Suppose that σ is supported on $[-R, R]$. Let $\rho(\mathrm{d}x) := (1 + x^2)\sigma(\mathrm{d}x)$. Then we can write

$$A(z) := A^{(\gamma,\sigma)}(z) = -a + \int_{[-R,R]} \frac{1}{x - z}\, \rho(\mathrm{d}x), \tag{5.17}$$

where $a := \gamma + \int_{\mathbb{R}} x\, \sigma(\mathrm{d}x)$. Denote by $\tilde{A}$ the analytic continuation of A to $\mathbb{C} \setminus [-R, R]$ given by the right-hand side of (5.17). The idea is to solve the differential equation $\frac{\mathrm{d}}{\mathrm{d}t}\tilde{F}_t(z) = \tilde{A}(\tilde{F}_t(z))$, $\tilde{F}_0(z) = z$, by the usual Picard's iteration, and show that the solution is holomorphic on $\mathbb{C} \setminus [-R_t, R_t]$ for some $R_t > 0$ with $\tilde{F}_t(\overline{z}) = \overline{\tilde{F}_t(z)}$. Beware that the existence of a unique solution $(\tilde{F}_t(z))_{t \ge 0}$ is already known for $z \in \mathbb{C} \setminus \mathbb{R}$ in Theorem 5.15, but for $z \in \mathbb{R} \setminus [-R, R]$ a solution may not exist globally as $\tilde{F}_t(z)$ might hit $+R$ or $-R$ in finite time.

Observe first that there exists $C > 0$ such that

$$|\tilde{A}(z)| \le C \quad \text{for all } z \in \mathbb{C} \text{ with } |z| > R + 1. \tag{5.18}$$

Let $F_t^n(z)$, $n = 0, 1, 2, \dots$, be recursively defined by

$$F_t^n(z) = z + \int_0^t \tilde{A}(F_s^{n-1}(z))\, \mathrm{d}s, \qquad F_t^0(z) := z. \tag{5.19}$$

All F_t^n are well defined holomorphic functions on $|z| > R + Ct + 2$. To see this, it suffices to show that $|F_t^n(z)| \ge R + 2$ for all $n \in \mathbb{N}$, $t \ge 0$ and $|z| > R + Ct + 2$. Indeed, supposing the claim is the case for $F_t^{n-1}(z)$, we have for all $t \ge 0$ and $|z| > R + Ct + 2$

$$|F_t^n(z)| \ge |z| - \int_0^t |\tilde{A}(F_s^{n-1}(z))|\, \mathrm{d}s \ge R + Ct + 2 - Ct = R + 2$$

as desired.

We can then easily show that $|F_t^n(z)| \le |z| + Ct$ for all $|z| > R + Ct + 2$, $t \ge 0$, and $n \in \mathbb{N}$. Since $(F_t^n(z))_{n \ge 1}$ is known to converge whenever $z \in \mathbb{C} \setminus \mathbb{R}$ (one can also show this directly by Picard's iteration; a more general setting will be treated in Theorem 6.11 (i)), by Vitali's theorem, for each fixed $t \ge 0$, the functions F_t^n converge locally uniformly to a holomorphic function $\tilde{F}_t$ on $|z| > R + Ct + 2$ with $\tilde{F}_t(\overline{z}) = \overline{\tilde{F}_t(z)}$. By the dominated convergence theorem, passing to the limit in (5.19) yields

$$\tilde{F}_t(z) = z + \int_0^t \tilde{A}(\tilde{F}_s(z))\, \mathrm{d}s, \qquad \tilde{F}_0(z) = z$$

for all $t \ge 0$ and $|z| > R + Ct + 2$. Therefore, we have constructed an analytic continuation $\tilde{F}_t$ of F_t on the domain $\mathbb{C} \setminus [-R - Ct - 2, R + Ct + 2]$. By Proposition 4.42, every μ_t has compact support.

Moreover, the above proof of (3) $\Longrightarrow$ (1) shows that μ_t is supported on $[-R - Ct - 2, R + Ct + 2]$ because $|\tilde{F}_t(z)| \ge R + 2$ holds for all $|z| > R + Ct + 2$ and hence $\tilde{F}_t$ has no zeros on $\mathbb{R} \setminus [-R - Ct - 2, R + Ct + 2]$; see Proposition 4.42 (iv). ☐

Remark 5.19. The above proof (2) $\Longrightarrow$ (3) actually shows that if μ_t is supported on $[-R_t, R_t]$ $(t > 0)$ then σ is supported on $\bigcap_{t>0}[-R_t, R_t]$.

Proposition 5.20. *Let $(\mu_t)_{t \ge 0}$ be a monotone convolution semigroup with infinitesimal generator $A^{(\gamma,\sigma)}$. If σ is supported on a compact interval $[-R, R]$, then A has a convergent series expansion*

$$A(z) = -\sum_{n=1}^{\infty} \frac{\alpha_n}{z^{n-1}}, \qquad |z| > R, \tag{5.20}$$

and the nth monotone cumulant of μ_t coincides with $t\alpha_n$ for all $t \ge 0$ and $n \in \mathbb{N}$.

Proof. Observe that

$$A(z) = -\gamma + \int_{\mathbb{R}} \left(\frac{1 + x^2}{x - z} - t \right) \sigma(\mathrm{d}x) = -a - \sum_{n \ge 0} \frac{1}{z^{n+1}} \int_{\mathbb{R}} x^n (1 + x^2)\, \sigma(\mathrm{d}x),$$

where $a := \gamma + \int_{\mathbb{R}} x\, \sigma(\mathrm{d}x)$, so that the sequence $(\alpha_n)_{n \ge 1}$ in (5.20) is given by

$$\alpha_1 := a, \qquad \alpha_n := \int_{\mathbb{R}} x^{n-2} (1 + x^2)\, \sigma(\mathrm{d}x), \qquad n \ge 2.$$

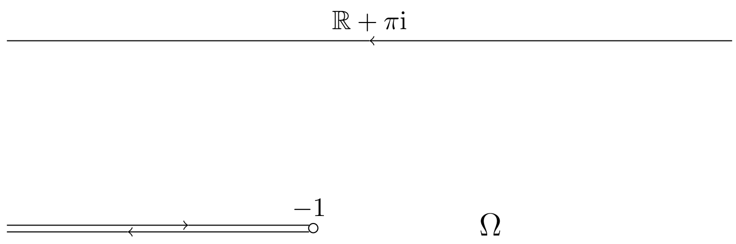

FIGURE 5. The range $\Omega = H(\mathbb{C}^+)$

Taking the derivative of $F_{t+s}(z) = F_t(F_s(z))$ with respect to s at 0 yields the first order PDE

$$\frac{\partial}{\partial t}F_t(z) = A(z)\frac{\partial}{\partial z}F_t(z),$$

which is equivalent to $\frac{\partial}{\partial t}G_t(z) = A(z)\frac{\partial}{\partial z}G_t(z)$ and so its integrated form

$$G_t(z) = \frac{1}{z} + A(z)\int_0^t \frac{\partial}{\partial z}G_s(z)\,\mathrm{d}s, \qquad t \geq 0. \tag{5.21}$$

As a result of Lemma 4.31, the map $(s,z) \mapsto \frac{\partial}{\partial z}G_s(z)$ is continuous, so that we may interchange the integral and $\frac{\partial}{\partial z}$ in (5.21). Let $\mathrm{m}_n(t)$ be the nth moment of μ_t. In view of Proposition 5.18, for each $T > 0$, the series

$$G_t(z) = \sum_{n\geq 0}\frac{\mathrm{m}_n(t)}{z^{n+1}}$$

converges uniformly on $[0,T]\times\{z : |z| > C_1 + C_2T\}$. This shows that the map $t \mapsto \mathrm{m}_n(t) = \frac{1}{2\pi \mathrm{i}}\int_{|z|=R} z^n G_t(z)\,\mathrm{d}z$ is continuous on $[0,T]$, where $R := C_1 + C_2T + 1$. The above arguments allow us to perform series expansions of functions in (5.21) to obtain

$$\sum_{n\geq 0}\frac{\mathrm{m}_n(t)}{z^{n+1}} = \frac{1}{z} + \left(\sum_{n=1}^{\infty}\frac{-\alpha_n}{z^{n-1}}\right)\left(\sum_{n=0}^{\infty}\frac{-(n+1)}{z^{n+2}}\int_0^t \mathrm{m}_n(s)\,\mathrm{d}s\right).$$

Comparing the coefficient of $\frac{1}{z^{n+1}}$ and by Cauchy's product formula, we obtain $\mathrm{m}_0(t) \equiv 1$, $\mathrm{m}_n(0) = 0$ for $n \in \mathbb{N}$, and

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{m}_n(t) = \sum_{\ell=0}^{n-1}(\ell+1)\alpha_{n-\ell}\mathrm{m}_\ell(t), \qquad n \geq 1,\ t \geq 0. \tag{5.22}$$

This is exactly the recursion for $\mathrm{m}_n^{\rhd}(t)$ in Proposition 3.15, so that $\mathrm{m}_n(t) = \mathrm{m}_n^{\rhd}(t)$ and α_n is the nth monotone cumulant of μ_1.

The above argument around (5.22) works for the rescaled monotone convolution semigroup $\tilde{\mu}_t := \mu_{st}$, $t \geq 0$, where $s > 0$ is fixed. Then $\tilde{\mathrm{m}}_n(t) := \mathrm{m}_n(st)$ is the nth moment of $\tilde{\mu}_t$ and the recursion (5.22) implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\mathrm{m}}_n(t) = \sum_{\ell=0}^{n-1}(\ell+1)s\alpha_{n-\ell}\tilde{\mathrm{m}}_\ell(t), \qquad n \geq 1,\ t \geq 0.$$

showing that $(s\alpha_n)_{n\geq 1}$ is the sequence of monotone cumulants of $\tilde{\mu}_1 = \mu_s$. ☐

Example 5.21 (Monotone Poisson distribution)**.** Let $A(z) = -1 + \frac{1}{1-z} = -\sum_{n\geq 1}\frac{1}{z^{n-1}}$. From Propositions 5.18 and 5.20, the corresponding monotone convolution semigroup $(\mu_t)_{t\geq 0}$ consists of compactly supported measures and $\kappa_n(\mu_t) = t$. Therefore μ_t has the same moment sequence as $\mathrm{Poi}^{\rhd}(t)$ in Theorem 3.19. Since μ_t has compact support, we conclude $\mu_t = \mathrm{Poi}^{\rhd}(t)$ by Proposition A.4.

By a standard ODE technique, the solution $(F_t)_{t\geq 0}$ to the ODE in Theorem 5.15 (ii) is given by the implicit formula $H(F_t(z)) = H(z) + t$, where

$$H(z) = C + \int_{\mathrm{i}}^{z}\frac{\mathrm{d}w}{A(w)} = \log z - z,$$

where $\log z$ is the principal branch and the arbitrary constant C is suitably selected. Since $H'(z) = 1/A(z)$ has negative imaginary part on $\mathbb{C}^+$, according to Lemma B.1, H is injective. We determine the range $H(\mathbb{C}^+)$. This is basically determined by $H(\mathbb{R})$ (see discussions below). By calculus, we see that

- $H\restriction_{(0,1]}$ is strictly increasing with range $(-\infty,-1]$,
- $H\restriction_{[1,+\infty)}$ is strictly decreasing with the same range $(-\infty,-1]$,
- $H\restriction_{(-\infty,0)}$ is injective with range $\mathbb{R} + \pi\mathrm{i}$.

This observation shows that $H(\mathbb{C}^+)$ is the region $\Omega := \{z \in \mathbb{C} : \Im(z) < \pi\} \setminus (-\infty,-1]$ (see Figure 5). Indeed, considering the behavior $H(w) = -w + o(w)$ as $w \to \infty$, every point $z \in \Omega$ has rotation number 1 with respect to the closed simple curve

$$\{H(w)\}_{w\in C}, \qquad C := [-R,-1/R] \cup \{(1/R)e^{-\mathrm{i}\theta} : -\pi \leq \theta \leq 0\} \cup [1/R, R] \cup \{Re^{\mathrm{i}\theta} : 0 \leq \theta \leq \pi\}$$

for sufficiently large $R > 1$, so that $z \in H(\mathbb{C}^+)$ by the argument principle. We can easily see from $\Im \log z < \pi$ that $H(\mathbb{C}^+) \subseteq \{z \in \mathbb{C} : \Im(z) < \pi\}$. We can also see that any $x \in (-\infty, -1]$ does not belong to $H(\mathbb{C}^+)$; indeed, if $z = re^{i\theta} \in \mathbb{C}^+$ and $H(z) \in \mathbb{R}$ then the condition $\Im[H(z)] = 0$ implies $r = \theta/\sin\theta$. Then the function

$$f(\theta) := \Re[H(re^{i\theta})] = \log\frac{\theta}{\sin\theta} - \frac{\theta\cos\theta}{\sin\theta}$$

has the positive derivative $f'(\theta) = [(\theta - \sin\theta\cos\theta)^2 + \sin^4\theta]/[\theta\sin^2\theta]$, so that $f(\theta) > f(+0) = -1$.

Note that Ω is invariant under the positive shifts $z \mapsto z + t$, which is actually a direct consequence of $H(F_t(z)) = H(z) + t$. Then the formula $F_t(z) = H^{-1}(H(z) + t)$ is well defined. Since H is analytic and injective on $(-\infty, 0)$, the function F_t extends to $(-\infty, 0)$, which is analytic and takes values in $(-\infty, 0)$. This implies that μ_t is supported on $[0, +\infty)$. This fact can also be deduced from Theorem 3.19 because the weak convergence limit of measures on $[0, +\infty)$ is supported on $[0, +\infty)$ as well.

Since H^{-1} has singularity at -1, let β_t be the unique solution $x > 1$ to $H(x) + t = -1$ and α_t the unique solution $0 < x < 1$ to $H(x) + t = -1$. Then F_t has an analytic extension to $(0, \alpha_t) \cup (\beta_t, +\infty)$ taking values in $(0, +\infty) \setminus \{1\}$. Therefore, μ_t is supported on $\{0\} \cup [\alpha_t, \beta_t]$. We can actually show that F_t extends to a continuous function on $\mathbb{C}^+ \cup \mathbb{R}$. For $x \in (\alpha_t, \beta_t)$ we can see that $H(x) + t \in \Omega$, so that $F_t(x) \in \mathbb{C}^+$. If $x = \alpha_t$ or $x = \beta_t$ then $F_t(x) = 1$. This implies by the Stieltjes inversion that μ_t has a continuous density function $p_t(x)$ on $[\alpha_t, \beta_t]$, positive on (α_t, β_t) and vanishing at the edges.

Finally, let us study F_t at 0. From $F_t(z) = H^{-1}(H(z) + t)$ we see that $\lim_{x\to 0^-} F_t(x) = 0$. The exponential form of $H(F_t(z)) = H(z) + t$ reads $F_t(z)e^{-F_t(z)} = ze^{-z+t}$. This has a unique analytic solution $F_t(z)$ at 0 having a convergent power series $F_t(z) = e^t z + O(z^2)$, $z \to 0$ with real coefficients. Therefore, F_t is analytic at 0 and $G_{\mu_t}(z) = e^{-t}/z + O(1/z^2)$, $z \to 0$, showing that $\mu_t(\{0\}) = e^{-t}$. Altogether, we have

$$\mu_t(\mathrm{d}x) = \mathrm{Poi}^{\rhd}(t)(\mathrm{d}x) = e^{-t}\delta_0(\mathrm{d}x) + p_t(x)\chi_{[\alpha_t,\beta_t]}(x)\,\mathrm{d}x,$$

where p_t is continuous on $[\alpha_t, \beta_t]$, positive in the interior and vanishing at the edges. In fact, $p_t(x)$ can be expressed with the Lambert W function as

$$p_t(x) = \frac{1}{\pi}\Im\left[\frac{1}{W_{-1}(-xe^{-x+t})}\right],$$

see the original article [121] and [26] for further information.

Theorem 5.22. *Let $(\mu_t)_{t\ge 0}$ be a monotone convolution semigroup, $A = A^{(\gamma,\sigma)}$ be its infinitesimal generator, and $n \in \mathbb{N}$. The following statements are equivalent:*

(1) $\int_{\mathbb{R}} x^{2n}\,\mu_t(\mathrm{d}x) < +\infty$ *for some $t > 0$;*

(2) $\int_{\mathbb{R}} x^{2n}\,\mu_t(\mathrm{d}x) < +\infty$ *for all $t > 0$;*

(3) $\int_{\mathbb{R}} x^{2n}\,\sigma(\mathrm{d}x) < +\infty$.

Moreover, if the above conditions are satisfied then for all $t \ge 0$ the monotone cumulants of μ_t up to order $2n$ are given by

$$\kappa_1(\mu_t) = t\left(\gamma + \int_{\mathbb{R}} x\,\sigma(\mathrm{d}x)\right),$$
$$\kappa_p(\mu_t) = t\int_{\mathbb{R}} x^{p-2}(1+x^2)\,\sigma(\mathrm{d}x), \qquad 2 \le p \le 2n.$$

Proof. Throughout the proof we use the simplified symbols $F_t := F_{\mu_t}$, $G_t := G_{\mu_t}$, and $\mathrm{m}_p(t) := \int_{\mathbb{R}} x^p\,\mu_t(\mathrm{d}x)$.

(1) $\Longrightarrow$ (2) is a direct consequence of Proposition 5.13 and of the semigroup relation $\mu_t = \mu_{t-s} \rhd \mu_s$, $0 \le s \le t$; note that we do not need the part $\int_{\mathbb{R}} x^{2n}\,(\mu \rhd \nu)(\mathrm{d}x) < +\infty \Longrightarrow \int_{\mathbb{R}} x^{2n}\,\mu(\mathrm{d}x) < +\infty$, whose proof has been postponed to Appendix.

(2) $\Longrightarrow$ (3)—*Step 1.* We show that $t \mapsto \mathrm{m}_p(t)$ is locally integrable with respect to the Lebesgue measure. For this purpose, we will show more strongly that $\mathrm{m}_p(t)$ is a polynomial in t. The starting point is the observation that $t \mapsto \mathrm{m}_p(t)$ is measurable for any $1 \le p \le 2n$. Formula in Proposition 5.13 yields

$$\mathrm{m}_p(t+s) = \mathrm{m}_p(t) + \mathrm{m}_p(s) + \sum_{\ell=1}^{p-1} \sum_{\substack{k_0,k_1,\dots,k_\ell \ge 0\\ k_0+k_1+\cdots+k_\ell = p-\ell}} \mathrm{m}_\ell(t)\mathrm{m}_{k_0}(s)\cdots\mathrm{m}_{k_\ell}(s) \tag{5.23}$$

for $1 \le p \le 2n$. As shown in (3.6), the polynomials $\mathrm{m}_0^{\rhd}(t), \mathrm{m}_1^{\rhd}(t), \mathrm{m}_2^{\rhd}(t), \dots,$ satisfy the same relation

$$\mathrm{m}_p^{\rhd}(t+s) = \mathrm{m}_p^{\rhd}(t) + \mathrm{m}_p^{\rhd}(s) + \sum_{\ell=1}^{p-1} \sum_{\substack{k_0,k_1,\dots,k_\ell \ge 0\\ k_0+k_1+\cdots+k_\ell = p-\ell}} \mathrm{m}_\ell^{\rhd}(t)\mathrm{m}_{k_0}^{\rhd}(s)\cdots\mathrm{m}_{k_\ell}^{\rhd}(s) \tag{5.24}$$

for $p \in \mathbb{N}$. Here we select (A, φ) and $x_i \in A$ in Proposition 3.15 so that $\varphi(x_i^p) = \mathrm{m}_p(1)$, $0 \le p \le 2n$. We will show that $\mathrm{m}_p(t) = \mathrm{m}_p^{\rhd}(t)$ for all $t \ge 0$ and $1 \le p \le 2n$ by induction on p.

For $p = 1$, formula (5.23) is just $\mathrm{m}_1(t+s) = \mathrm{m}_1(t) + \mathrm{m}_1(s)$, i.e., Cauchy's functional equation. Since $t \mapsto \mathrm{m}_1(t)$ is measurable, due to Proposition D.1 we have $\mathrm{m}_1(t) = \mathrm{m}_1(1)t = \varphi(x_1)t = \mathrm{m}_1^{\rhd}(t)$. Assume that $\mathrm{m}_i(t) = \mathrm{m}_i^{\rhd}(t)$ for all $t \ge 0$ and $1 \le i \le p-1$. Subtracting (5.24) from (5.23) yields

$$\mathrm{m}_p(t+s) - \mathrm{m}_p^{\rhd}(t+s) = [\mathrm{m}_p(t) - \mathrm{m}_p^{\rhd}(t)] + [\mathrm{m}_p(s) - \mathrm{m}_p^{\rhd}(s)],$$

which is again Cauchy's functional equation. Therefore $\mathrm{m}_p(t) - \mathrm{m}_p^{\triangleright}(t) = (\mathrm{m}_p(1) - \mathrm{m}_p^{\triangleright}(1))t$. Since $\mathrm{m}_p(1) = \mathrm{m}_p(\mu_1) = \varphi(x_i^p) = \mathrm{m}_p^{\triangleright}(1)$, we conclude $\mathrm{m}_p(t) = \mathrm{m}_p^{\triangleright}(t)$ as desired.

(2) $\Longrightarrow$ (3)—*Step 2.* We fix an arbitrary $T > 0$. From equality (5.21) and the asymptotic expansion (4.35), we get as $z = \mathrm{i}y$, $y \to +\infty$,

$$\begin{aligned} A(z) &= \frac{G_T(z) - z^{-1}}{\int_0^T \frac{\partial}{\partial z} G_t(z)\,\mathrm{d}t} \\ &= \frac{-\left[\mathrm{m}_1(T) + \mathrm{m}_1(T) z^{-1} + \cdots + \mathrm{m}_{2n}(T) z^{-2n+1} + o(|z|^{-2n+1})\right]}{1 + 2z^{-1}\int_0^T \mathrm{m}_1(t)\,\mathrm{d}t + \cdots + (2n+1) z^{-2n}\int_0^T \mathrm{m}_{2n}(t)\,\mathrm{d}t - z^2 \int_0^T R_t'(z)\,\mathrm{d}t}, \end{aligned} \tag{5.25}$$

where $R_t'(z)$ is the z-derivative of the remainder term in (4.35), i.e.,

$$R_t'(z) := -\frac{2n+1}{z^{2n+2}} \int_{\mathbb{R}} \frac{x^{2n+1}}{z - x}\,\mu_t(\mathrm{d}x) - \frac{1}{z^{2n+1}} \int_{\mathbb{R}} \frac{x^{2n+1}}{(z-x)^2}\,\mu_t(\mathrm{d}x).$$

Since $\mathrm{m}_{2n}(t)$ is a polynomial in t, we have $\int_0^T \left(\int_{\mathbb{R}} x^{2n}\,\mu_t(\mathrm{d}x)\right)\mathrm{d}t < +\infty$, which readily implies that $\int_0^T R_t'(\mathrm{i}y)\,\mathrm{d}t = o(y^{-(2n+2)})$ by the dominated convergence theorem. Therefore, using the geometric series expansion $1/(1+\zeta) = 1 - \zeta + \zeta^2 + \cdots$ to the denominator of (5.25) and recollecting terms, we find reals $\alpha_1, \alpha_2, \dots, \alpha_{2n}$ such that

$$A(z) = -\alpha_1 - \frac{\alpha_2}{z} - \cdots - \frac{\alpha_{2n}}{z^{2n-1}} + Q(z) \tag{5.26}$$

where $Q(\mathrm{i}y) = o(y^{-(2n-1)})$ as $y \to +\infty$. By Proposition 4.46, we have $\mathrm{m}_{2n}(\sigma) < +\infty$.

(3) $\Longrightarrow$ (1). Since $\mathrm{m}_{2n}(\sigma) < +\infty$, again Proposition 4.46 yields the expansion (5.26) in each sector domain, i.e., $Q(z) = o(|z|^{-(2n-1)})$ holds as $z \to \infty$ within ∇_γ for each fixed $\gamma > 0$. The integral equation (5.13) reads

$$F_t(z) = z - \alpha_1 t - \sum_{\ell=1}^{2n-1} \alpha_{\ell+1} \int_0^t F_s(z)^{-\ell}\,\mathrm{d}s + \int_0^t Q(F_s(z))\,\mathrm{d}s. \tag{5.27}$$

As discussed in the proof of Theorem 5.22, thanks to the tightness of $\{\mu_t : t \in [0,T]\}$ for each fixed $T > 0$ and $\gamma > 0$, there is $y_0 > 0$ such that $\{F_t(\mathrm{i}y) : 0 \le t \le T,\ y > y_0\} \subseteq \nabla_\gamma$, and also the asymptotic behavior $F_t(\mathrm{i}y) = \mathrm{i}y(1 + o(1))$ holds uniformly on $t \in [0,T]$. From this observation, we can deduce the following uniform estimates over $0 \le t \le T$:

$$\int_0^t Q(F_s(\mathrm{i}y))\,\mathrm{d}s = o(y^{-(2n-1)}), \tag{5.28}$$

$$\int_0^t F_s(\mathrm{i}y)^{-\ell}\,\mathrm{d}s = t(\mathrm{i}y)^{-\ell}(1 + o(1)). \tag{5.29}$$

Plugging (5.28) and (5.29) for $\ell = 1$ into (5.27) yields

$$F_t(\mathrm{i}y) = z - \alpha_1 t - \frac{\alpha_2 t}{\mathrm{i}y} + o(y^{-1}),$$

which is again uniform, i.e., the modulus of the remainder term $o(y^{-1})$ is bounded by a function $f(y)$ independent of $t \in [0,T]$ such that $yf(y) \to 0$ as $y \to +\infty$. Plugging this improved estimate into (5.27) then gives

$$F_t(\mathrm{i}y) = \mathrm{i}y - \alpha_1 t - \frac{\alpha_2 t}{\mathrm{i}y} - \frac{\alpha_3 t + (\alpha_1\alpha_2 t^2)/2}{(\mathrm{i}y)^2} + o(y^{-2}).$$

Repeating these arguments amounts to

$$F_t(\mathrm{i}y) = \mathrm{i}y - b_1(t) - \frac{b_2(t)}{\mathrm{i}y} - \cdots - \frac{b_{2n}(t)}{(\mathrm{i}y)^{2n-1}} + o(y^{-(2n-1)}), \qquad y \to +\infty.$$

for some polynomials $b_1(t), b_2(t), \dots, b_{2n}(t)$ with real coefficients. From Proposition 4.46 we conclude $\int_{\mathbb{R}} x^{2n}\,\mu_t(\mathrm{d}x) < +\infty$ for all $0 \le t \le T$.

Formulas for monotone cumulants. In the proof of Step 2 above, we have used

$$G_t(z) - \frac{1}{z} = A(z) \int_0^t \frac{\partial}{\partial z} G_s(z)\,\mathrm{d}s.$$

Substituting the truncated Laurent series for $G_t(z)$ and $A(z)$ and comparing the coefficients yields exactly the relations (5.22) up to the order $2n$. Then the remaining proof is identical to the proof of Proposition 5.20. □

5.3. **Infinitely divisible distributions.** The concept of infinitely divisible distribution is closely related to convolution semigroups. In probability theory, infinitely divisible distributions naturally appear as the limit distributions of sums of independent random variables that are "uniformly" small [76]. We consider the monotone analogue.

Definition 5.23. A probability measure μ on $\mathbb{R}$ is said to be **monotonically infinitely divisible** if for every $N \in \mathbb{N}$ there exists a probability measure $\mu_{1/N}$ such that μ is the N-fold monotone convolution of $\mu_{1/N}$, i.e.,

$$\mu = (\mu_{1/N})^{\triangleright N} := \underbrace{\mu_{1/N} \triangleright \mu_{1/N} \triangleright \cdots \triangleright \mu_{1/N}}_{N \text{ times}}.$$

It is obvious that each member of a monotone convolution semigroup $(\mu_t)_{t \ge 0}$ is monotonically infinitely divisible, as μ_t is the N-fold monotone convolution of $\mu_{t/N}$. This observation can be enhanced to the following.

Theorem 5.24. *Let μ be a probability measure on $\mathbb{R}$ with determinate moment sequence. The following are equivalent.*

(1) *μ is monotonically infinitely divisible.*

(2) *There is a monotone convolution semigroup $(\mu_t)_{t\geq 0}$ such that $\mu_1 = \mu$.*

(3) *The sequence $(\kappa_n(\mu))_{n\geq 2}$ of monotone cumulants from order two is positive semi-definite: for every $p \in \mathbb{N}$ and $c_1, c_2, \dots, c_p \in \mathbb{R}$ one has*

$$\sum_{i,j=1}^{p} c_i c_j \kappa_{i+j}(\mu) \geq 0.$$

(4) *There exist a sequence of probability measures $(\nu_N)_{N\geq 1}$ with finite moments of all orders and a sequence of positive integers $(\ell_N)_{N\geq 1}$ such that $\lim_{N\to\infty} \ell_N = \infty$ and*

$$\lim_{N\to\infty} \int_{\mathbb{R}} x^n \, (\nu_N)^{\rhd \ell_N}(\mathrm{d}x) = \int_{\mathbb{R}} x^n \, \mu(\mathrm{d}x), \qquad n \in \mathbb{N}.$$

Proof. (2) $\Longrightarrow$ (1) is obvious from $\mu = \mu_1 = (\mu_{1/N})^{\rhd N}$ for all $N \in \mathbb{N}$.

(1) $\Longrightarrow$ (4) is also obvious as one can select $\nu_N := \mu_{1/N}$ and $\ell_N := N$.

(4) $\Longrightarrow$ (3). We show that

$$\kappa_n(\mu) = \lim_{N\to\infty} \ell_N \mathrm{m}_n(\nu_N), \qquad n \in \mathbb{N}. \tag{5.30}$$

To see this first we observe that

$$\kappa_n((\nu_N)^{\rhd \ell_N}) = \ell_N \kappa_n(\nu_N).$$

The left-hand side converges to $\kappa_n(\mu)$ since $\kappa_n((\nu_N)^{\rhd \ell_N})$ is a polynomial in the moments $\mathrm{m}_p((\nu_N)^{\rhd \ell_N})$ $(1 \leq p \leq n)$ that converge to $\mathrm{m}_p(\mu)$ by the assumption. In particular, $\lim_{N\to\infty} \kappa_n(\nu_N) = 0$. On the other hand, recall that

$$\mathrm{m}_n(\nu_N) = \kappa_n(\nu_N) + Q_n^{\rhd}(\kappa_1(\nu_N), \dots, \kappa_{n-1}(\nu_N)),$$

where the universal polynomial $Q_n^{\rhd}$ has no constant or linear term. Therefore, we obtain

$$\begin{aligned}\lim_{N\to\infty} \ell_N \mathrm{m}_n(\nu_N) &= \lim_{N\to\infty} \ell_N \kappa_n(\nu_N) + \lim_{N\to\infty} \ell_N Q_n^{\rhd}(\kappa_1(\nu_N), \dots, \kappa_{n-1}(\nu_N)) \\ &= \lim_{N\to\infty} \ell_N \kappa_n(\nu_N) = \kappa_n(\mu)\end{aligned}$$

as desired. From (5.30) we obtain

$$\begin{aligned}\sum_{i,j=1}^{p} c_i c_j \kappa_{i+j}(\mu) &= \lim_{N\to\infty} \ell_N \sum_{i,j=1}^{p} c_i c_j \int_{\mathbb{R}} x^{i+j} \, \nu_N(\mathrm{d}x) \\ &= \lim_{N\to\infty} \ell_N \int_{\mathbb{R}} \left| \sum_{i=1}^{p} c_i x^i \right|^2 \nu_N(\mathrm{d}x) \geq 0.\end{aligned}$$

(3) $\Longrightarrow$ (2). By the assumption and by Theorem A.1, there exists a finite Borel measure ρ on $\mathbb{R}$ with finite moments of all orders such that

$$\kappa_n(\mu) = \int_{\mathbb{R}} x^{n-2} \, \rho(\mathrm{d}x), \qquad n \geq 2.$$

We set

$$A(z) := -\kappa_1(\mu) + \int_{\mathbb{R}} \frac{1}{x - z} \, \rho(\mathrm{d}x), \qquad z \in \mathbb{C}^+.$$

Since this is a Nevanlinna function with $\sphericalangle \lim_{z\to\infty} A(z)/z = 0$, by Theorem 5.15, there corresponds a monotone convolution semigroup $(\mu_t)_{t\geq 0}$. Theorem 5.22 ensures that every μ_t has finite moments of all orders. By the last statement of Theorem 5.22, the monotone cumulant $\kappa_n(\mu_t)$ coincides with $t\kappa_n(\mu)$ for all $t \geq 0$ and $n \in \mathbb{N}$; in particular, $\kappa_n(\mu_1) = \kappa_n(\mu)$. This means that μ_1 and μ have the same moment sequence. Since μ has a determinate moment sequence, we conclude that $\mu_1 = \mu$. ☐

Example 5.25. Condition (3) means that the symmetric matrix $(\kappa_{i+j}(\mu))_{i,j=1}^{p}$ is positive semi-definite for all $p \in \mathbb{N}$. This condition can be used as a test for monotonically infinite divisibility.

(a) One can check that the determinant of the matrix $(\kappa_{i+j}(\mathrm{S}(0,1)))_{i,j=1}^{5}$ is negative; the recurrence relation in Example 3.14 would be helpful for the calculation. This implies that the standard semicircle distribution is not monotonically infinitely divisible.

(b) The matrix $(\kappa_{i+j}(\mathrm{N}(0,1)))_{i,j=1}^{p}$ is surprisingly positive semi-definite up to $p = 200$ according to simulations. However, whether $\mathrm{N}(0,1)$ is monotonically infinitely divisible or not is still unknown. In terms of holomorphic dynamics, the question is whether one can find a compositional semigroup $(F_t)_{t\geq 0}$ of reciprocal Cauchy transforms such that $F_0 = \mathrm{id}_{\mathbb{C}^+}$ and $F_1 = F_{\mathrm{N}(0,1)}$. This kind of question is natural in complex analysis but known to be hard.

Example 5.26. Condition (4) and Proposition A.7 imply that the sequence of probability measures $((\nu_N)^{\rhd \ell_N})_{N\geq 1}$ weakly converges to μ. Thus monotonically infinitely divisible distributions are obtained as the limits of the distribution of the sum of monotonically iid random variables. Special cases include the monotone CLT and Poisson's law of small numbers as follows.

(a) Let ν be a probability measure with mean 0, second moment 1 and finite moments of higher orders. Let ν_N be the pushforward of ν by the map $x \mapsto x/\sqrt{N}$. Then the monotone CLT (Theorem 3.18) says that $(\nu_N)^{\triangleright N}$ converges weakly to the arcsine law $\mathrm{A}(0,1)$.

(b) Let $\lambda > 0$ and $\nu_N := (1 - \frac{\lambda}{N})\delta_0 + \frac{\lambda}{N}\delta_1$ for $N > \lambda$. Then monotone Poisson's law of small numbers (Theorem 3.19) says that the measure $(\nu_N)^{\triangleright N}$ converges weakly to the monotone Poisson distribution $\mathrm{Poi}^{\triangleright}(\lambda)$.

Remark 5.27. As a further generalization of the limit theorem in condition (4), one could also consider the monotone convolution of non-identical distributions of the form $\nu_{N,1} \triangleright \nu_{N,2} \triangleright \cdots \triangleright \nu_{N,\ell_N}$. Then the limit distribution would not necessarily be monotonically infinitely divisible, see [70] for further details.

5.4. **Notes.** The definition of additive monotone convolution of probability measures in Theorem 5.1 is due to Muraki [120, Definition 3.2]. Example 5.4 is a special case of examples in [117, Section 8]. Młotkowski introduced "Fuss-Catalan distributions", which behave nicely with respect to monotone convolution [116, Proposition 4.5]. The definition of multiplicative monotone convolution in Theorem 5.5 appeared in Arizmendi and Hasebe [10, Proposition 3.2] but its weak continuity in Proposition 5.6 is a new result. Both convolutions have unbounded operator models constructed by Franz [69]. Note that for multiplicative monotone convolution, Franz's model was restricted to the case where both measures are supported on $[0, +\infty)$, but the same technique is applicable to the general case of Theorem 5.5.

Theorem 5.15 was due to Muraki [120]. Our proof heavily depends on Berkson–Porta's work while the original proof was more straightforward. The exposition of Theorem 5.22 followed [80, Theorem 4.8]. Theorem 5.24 is based on [82, Theorem 8.5]. Theorem 5.24 for compactly supported μ was due to Muraki [120, Section 5] except condition (3). Belinschi proved the equivalence of (1) and (2) in Theorem 5.24 for arbitrary probability measures, as well as the uniqueness of the monotone convolution semigroup into which μ embeds [23]. The aspect of limit theorems, i.e., condition (4) of Theorem 5.24, was further studied by Anshelevich and Williams without assuming the existence of moments [7].

An analogue of Theorem 5.24 in free probability is known [125, Theorem 13.16]; however, the proof of Theorem 5.24 is more complicated even if μ has compact support. The main difficulty is the absence of a priori bounds for monotone cumulants of the form $|\kappa_n(\mu)| \le C^n$ for compactly supported measures μ; compare with the bounds for free cumulants [125, Lemma 13.13]. In the case of monotonically infinitely divisible distributions with compact support, this bound comes a posteriori as a result of Propositions 5.18, 5.20 and Theorem 5.24. For general probability measures μ with compact support, it is still unknown whether an exponential bound $|\kappa_n(\mu)| \le C^n$ holds for some $C > 0$ or not.

Monotone convolution semigroups for multiplicative convolutions are also studied in the literature; see e.g. [31, 68]. There is a certain parallelism between additive and multiplicative cases, which was systematically studied in [6].

A remarkable feature of additive and multiplicative monotone convolutions is a connection to additive free convolution $\boxplus$ and multiplicative free convolution $\boxtimes$: for probability measures λ on $[0, +\infty)$ and μ, ν on $\mathbb{R}$ there exist probability measures $\rho = \rho_{\mu,\nu}$ and $\sigma = \sigma_{\lambda,\mu}$ on $\mathbb{R}$ such that

$$\mu \boxplus \nu = \mu \triangleright \rho, \tag{5.31}$$

$$\lambda \boxtimes \mu = \lambda \circlearrowright \sigma; \tag{5.32}$$

see [21, 35] for additive convolution and [11] for multiplicative convolution. These relations have been used for the study of regularity properties of free convolutions, see e.g. [11, 24, 94]. Formulas (5.31) and (5.32) have an elegant interpretation in terms of graph products; see Accardi, Lenczewski and Sałapata [2] (additive case) and Lenczewski [103] (multiplicative case). Jekel and Liu's tree independence also allows an interpretation. In the context of Loewner theory, the monotone convolution hemigroups associated with free convolution hemigroups are studied e.g. in [70, 84, 92, 140]. Other notable connections between free probability and monotone probability can be found in Franz [68], Skoufranis [144], Cébron, Dahlqvist, Gabriel and Gilliers [41, 43], and Mingo and Tseng [114].

6. Monotone convolution hemigroups and Loewner theory

As already mentioned, convolution semigroups correspond to Lévy processes that are continuous-time analogues of random walk of identically distributed increments. We turn our attention to random walk whose increments are still independent but not necessarily identically distributed. Let $(X_i)_{i=1}^\infty$ be independent, $\mathbb{R}$-valued random variables. We denote by ν_i the distribution of X_i. Let us consider the random walk

$$S_n := X_1 + X_2 + \cdots + X_n; \qquad S_0 := 0.$$

The distribution of S_n is given by $\mu_n := \nu_1 * \nu_2 * \cdots * \nu_n$. For $n > m$ we have the relation $\mu_n = \mu_m * \nu_{m+1} * \cdots * \nu_n$, which involves ν_i's. To obtain a closed relation of distributions, it is more convenient to consider the increments

$$S_{m,n} := S_n - S_m, \qquad 0 \le m \le n.$$

Let $\mu_{m,n}$ be the law of $S_{m,n}$. The obvious identity $S_{\ell,m} + S_{m,n} = S_{\ell,n}$ gives rise to the distributional relation

$$\mu_{\ell,m} * \mu_{m,n} = \mu_{\ell,n}, \qquad 0 \le \ell \le m \le n, \tag{6.1}$$

$$\mu_{n,n} = \delta_0, \qquad n \in \mathbb{N}_0. \tag{6.2}$$

The law of X_i can be recovered from the two parameter family $(\mu_{m,n})_{0 \le m \le n}$ as $\mu_{i-1,i}$. Conversely, given a family $(\mu_{m,n})_{0 \le m \le n}$ of probability measures on $\mathbb{R}$ with relations (6.1) and (6.2), it comes from a random walk.

The above discrete-time setup can be well extended to the continuous-time case. A family $(\mu_{s,t})$ of probability measures, index by real numbers $0 \le s \le t < \infty$, is called a convolution hemigroup if $\mu_{t,t} = \delta_0$ and $\mu_{s,t} * \mu_{t,u} = \mu_{s,u}$ for all $0 \le s \le t \le u$ and $(s,t) \mapsto \mu_{s,t}$ is weakly continuous. A convolution hemigroup corresponds to a stochastic process called an additive process or a process with independent increments. The reader is referred to [136] for further information. Replacing the convolution $*$ with monotone convolution, we are lead to the following.

Definition 6.1. Let $\triangle := \{(s,t) \in \mathbb{R}^2 : 0 \le s \le t < +\infty\}$. A family $(\mu_{s,t})_{(s,t)\in\triangle}$ of probability measures on $\mathbb{R}$ is called a **monotone convolution hemigroup** if

(i) $\triangle \ni (s,t) \mapsto \mu_{s,t}$ is weakly continuous, i.e., for every bounded continuous function f on $\mathbb{R}$, the function $(s,t) \mapsto \int_{\mathbb{R}} f(x)\mu_{s,t}(\mathrm{d}x)$ is continuous.

(ii) $\mu_{s,u} = \mu_{s,t} \triangleright \mu_{t,u}$ for all $0 \le s \le t \le u < +\infty$,

(iii) $\mu_{s,s} = \delta_0$ for all $s \ge 0$.

If a monotone convolution hemigroup satisfies $\mu_{s,t} = \mu_{0,t-s}$ for all $0 \le s \le t$, then it is reduced to the convolution semigroup: $\mu_{0,s} \triangleright \mu_{0,t} = \mu_{0,s+t}$ holds. Conversely, given a monotone convolution semigroup (μ_t), the measures $\mu_{s,t} := \mu_{t-s}$ form a convolution hemigroup. Thus monotone convolution hemigroups generalize semigroups.

A monotone convolution hemigroup $(\mu_{s,t})$ can be described by its reciprocal Cauchy transforms.

Definition 6.2. A family of holomorphic self-maps $(F_{s,t})_{(s,t)\in\triangle}$ on $\mathbb{C}^+$ is called a **$\mathscr{P}$-reverse evolution family** ($\mathscr{P}$-REF for short)[§] if

(R1) $\sphericalangle \lim_{z\to\infty} F_{s,t}(z)/z = 1$ for all $(s,t) \in \triangle$,

(R2) $\triangle \ni (s,t) \mapsto F_{s,t}$ is continuous with respect to the locally uniform convergence on $\mathbb{C}^+$,

(R3) $F_{s,t} \circ F_{t,u} = F_{s,u}$ for all $0 \le s \le t \le u < +\infty$,

(R4) $F_{s,s}(z) = z$ for all $s \ge 0$ and $z \in \mathbb{C}^+$.

We often impose the additional condition that

(R5) for each $(s,t) \in \triangle$ there exist $m_{s,t} \in \mathbb{R}$ and $v_{s,t} \in [0,+\infty)$ such that

$$F_{s,t}(z) = z - m_{s,t} + \frac{v_{s,t}}{z} + o(z^{-1}), \qquad z = \mathrm{i}y,\ y \to +\infty, \tag{6.3}$$

and the maps $(s,t) \mapsto m_{s,t}$ and $(s,t) \mapsto v_{s,t}$ are continuous.

We call $(F_{s,t})$ satisfying (R1)–(R5) a **$\mathscr{P}_2$-REF**. Moreover, if $m_{s,t} = 0$ for all (s,t), we call $(F_{s,t})$ a **$\mathscr{P}_2^0$-REF**.

Remark 6.3. a) The continuity $(s,t) \mapsto F_{s,t}$ is equivalent to the weaker condition that the map $\triangle \ni (s,t) \mapsto F_{s,t}(z) \in \mathbb{C}^+$ is continuous for all $z \in \mathbb{C}^+$, thanks to Proposition 4.36.

b) In fact, the continuities of $(s,t) \mapsto m_{s,t}$ and $(s,t) \mapsto v_{s,t}$ in condition (R5) follow from the other conditions. This fact, however, requires rather long arguments and we refer the interested reader to [84].

c) There is a one-to-one correspondence between the set of the monotone convolution hemigroups and the set of the $\mathscr{P}$-REFs. From Proposition 4.46, the $\mathscr{P}_2$-REFs exactly correspond to the monotone convolution hemigroups with finite first and second moments continuous with respect to (s,t). Let $(F_{s,t})_{(s,t)\in\triangle}$ be a $\mathscr{P}_2$-REF and $(\mu_{s,t})_{(s,t)\in\triangle}$ be the associated monotone convolution hemigroup. Formulas (5.9) and (5.10) together with (4.32) and (4.33) show that, with notation in (6.3),

$$m_{s,u} = \mathrm{m}_1(\mu_{s,u}) = \mathrm{m}_1(\mu_{s,t}) + \mathrm{m}_1(\mu_{t,u}) = m_{s,t} + m_{t,u},$$
$$v_{s,u} = \mathrm{Var}(\mu_{s,u}) = \mathrm{Var}(\mu_{s,t}) + \mathrm{Var}(\mu_{t,u}) = v_{s,t} + v_{t,u}, \qquad 0 \le s \le t \le u.$$

d) If $(F_{s,t})$ is a $\mathscr{P}_2$-REF, then from (4.36) each $F_{s,t}$ has an integral formula

$$F_{s,t}(z) = z - m_{s,t} + \int_{\mathbb{R}} \frac{1}{x - z}\,\rho_{s,t}(\mathrm{d}x),$$

where $\rho_{s,t}$ is a finite Borel measure such that $\rho_{s,t}(\mathbb{R}) = v_{s,t}$. We can see that the map $(s,t) \mapsto \rho_{s,t}$ is weakly continuous. Let $(s_n,t_n), (s,t) \in \triangle$ and $(s_n,t_n) \to (s,t)$. Let $\rho_n := \rho_{s_n,t_n}$ and $\rho := \rho_{s,t}$. If $\rho = 0$ then $\rho_n(\mathbb{R}) = v_{s_n,t_n} \to v_{s,t} = 0$, so that $\rho_n \to 0$ weakly. If $\rho \ne 0$ then $v_{s,t} > 0$ and we set $\overline{\rho} := \rho/v_{s,t}$ and $\overline{\rho}_n := \rho/v_{s_n,t_n}$. As $G_{\rho_n}(z) = z - m_{s_n,t_n} - F_{s_n,t_n}(z)$ converges to $G_\rho(z)$, the Cauchy transform of the normalized measure $G_{\overline{\rho}_n}(z)$ also converges to $G_{\overline{\rho}}(z)$ for each $z \in \mathbb{C}^+$. Therefore, by Proposition 4.36, $\overline{\rho}_n$ converges weakly to $\overline{\rho}$, which in turn implies that ρ_n converges weakly to ρ.

e) A family $(F_{s,t})_{(s,t)\in\triangle}$ of holomorphic self-maps of $\mathbb{C}^+$ satisfying (R2)–(R4) is called a reverse evolution family. Such a family of holomorphic self-maps is well developed in Loewner theory, e.g. in [39]. The reason of the term “reverse” is that from the viewpoint of dynamics on $\mathbb{C}^+$, the alternative condition $F_{t,u} \circ F_{s,t} = F_{s,u}$ has a more natural interpretation that a point z at time s arrives at the point $F_{s,t}(z)$ at time t and then $F_{s,t}(z)$ arrives at the point $F_{t,u}(F_{s,t}(z))$ at time u, which coincides with $F_{s,u}(z)$. Such a family is called an evolution family.

Moreover, we consider a one-parameter family of holomorphic functions, which turns out to have a one-to-one correspondence with the $\mathscr{P}$-REFs.

Definition 6.4. A family $(F_t)_{t\ge0}$ of holomorphic self-maps of $\mathbb{C}^+$ is called a **$\mathscr{P}$-decreasing Loewner chain** ($\mathscr{P}$-DLC for short) if the following conditions (L1)–(L5) are satisfied:

(L1) $\sphericalangle \lim_{z\to\infty} F_t(z)/z = 1$ for every $t \ge 0$,

(L2) $t \mapsto F_t$ is continuous with respect to the locally uniform convergence,

(L3) F_t is injective on $\mathbb{C}^+$ for each $t \ge 0$,

[§]$\mathscr{P}$ stands for the fact that each function $F_{s,t}$ corresponds to a probability measure.

(L4) the range $F_t(\mathbb{C}^+)$ is non-increasing with respect to $t \geq 0$,

(L5) $F_0(z) = z$ for all $z \in \mathbb{C}^+$.

We also consider the condition that

(L6) for every $t \geq 0$ there exist $m_t \in \mathbb{R}, v_t \in [0, +\infty)$ such that

$$F_t(z) = z - m_t + \frac{v_t}{z} + o(z^{-1}), \qquad z = \mathrm{i}y, \ y \to +\infty, \tag{6.4}$$

and also the functions $t \mapsto m_t$ and $t \mapsto v_t$ are continuous.

We call (F_t) a $\mathscr{P}_2$**-DLC** if (L1)–(L6) are satisfied. Moreover, if $m_t = 0$ for all $t \geq 0$ then we call (F_t) a $\mathscr{P}_2^0$**-DLC**.

Example 6.5. (a) Let $(\mu_t)_{t\geq 0}$ be a monotone convolution semigroup and $f\colon [0, +\infty) \to \mathbb{R}$ be a continuous nondecreasing function. Then $\mu_{s,t} := \mu_{f(t)-f(s)}$ form a monotone convolution hemigroup.

(b) Recall from (4.40) that the semicircle distribution $\mathrm{S}(0, t)$ has the reciprocal Cauchy transform

$$F_t(z) = \frac{z + \sqrt{z^2 - 4t}}{2}.$$

One can check that $(F_t)_{t\geq 0}$ is a $\mathscr{P}_2^0$-DLC. First of all, F_t is of the form (6.4) with $m_t = 0$ and $v_t = t$. It remains to check conditions (L3) and (L4) since the others are easy. Condition (L3) can be directly confirmed: assuming $z, w \in \mathbb{C}^+$ and $F_t(z) = F_t(w)$, we obtain $z = w$ after algebraic calculations. In fact, we can show more strongly that F_t extends to a continuous injective function $\tilde{F}_t\colon \mathbb{C}^+ \cup \mathbb{R} \to \mathbb{C}^+ \cup \mathbb{R}$. Regarding condition (L4), from the previous consideration and Carathéodory's theorem for Jordan domains, the boundary of the domain $F_t(\mathbb{C}^+)$ is $\tilde{F}_t(\mathbb{R})$. For $x > 2\sqrt{t}$, the point $\tilde{F}_t(x) = (x + \sqrt{x^2 - 4t})/2$ moves over the half-line $[2\sqrt{t}, +\infty)$. For $x < -2\sqrt{t}$, note that the point $\sqrt{(x + 0\mathrm{i})^2 - 4t}$ has argument π by the definition of square root, so that $\tilde{F}_t(x) = (x - \sqrt{x^2 - 4t})/2$ and its trajectory is the half-line $(-\infty, -2\sqrt{t})$. For $|x| \leq 2\sqrt{t}$ by the Stieltjes inversion we have $\tilde{F}_t(x) = (x + \mathrm{i}\sqrt{4t - x^2})/2$, which moves over the semi-circle $\{u + \mathrm{i}v : u^2 + v^2 = 4t, \ v > 0\}$. In conclusion,

$$F_t(\mathbb{C}^+) = \{u + \mathrm{i}v \in \mathbb{C}^+ : u^2 + v^2 > 4t\},$$

which is decreasing with respect to $t \geq 0$.

This section establishes a correspondence between $\mathscr{P}_2$-REFs and $\mathscr{P}_2$-DLCs, and an integro-differential/integral equation for them, which generalizes the differential equation known in case the time-dependence is absolutely continuous.

6.1. **Reverse evolution families and Loewner chains.** We establish a bijection between the $\mathscr{P}_2$-REFs and $\mathscr{P}_2$-DLCs.

Lemma 6.6. *Let $\mathbb{C}^+_\beta := \{z \in \mathbb{C}^+ : \Im(z) > \beta\}$ for $\beta > 0$. Let μ be a probability measure with finite second moment. Then for any $\beta > 0$ one has*

$$F_\mu(z) = z + O(1), \qquad z \to \infty, \ z \in \mathbb{C}^+_\beta. \tag{6.5}$$

Moreover, F_μ is injective on $\mathbb{C}^+_\sigma$, where $\sigma := \sqrt{\mathrm{Var}(\mu)}$, and the range $F_\mu(\mathbb{C}^+_\sigma)$ contains $\mathbb{C}^+_{2\sigma}$.

Proof. Recall from (4.36) that we have a Nevanlinna formula of the form

$$F_\mu(z) = z - m + \int_{\mathbb{R}} \frac{\rho(\mathrm{d}x)}{x - z}, \tag{6.6}$$

where $m = \mathrm{m}_1(\mu)$, ρ is a finite Borel measure and $\rho(\mathbb{R}) = \mathrm{Var}(\mu) = \sigma^2$. For $z \in \mathbb{C}^+_\beta$, the function $1/|x - z|$ is bounded by $1/\beta$, so that the dominated convergence theorem yields that the integral term in (6.6) goes to zero, and hence (6.5) holds true.

For every $z, w \in \mathbb{C}^+_\sigma$ with $z \neq w$ we have

$$\begin{aligned} |F_\mu(z) - F_\mu(w)| &= |z - w| \left| 1 + \int_{\mathbb{R}} \frac{\rho(\mathrm{d}x)}{(x - z)(x - w)} \right| \\ &\geq |z - w| \left[1 - \int_{\mathbb{R}} \frac{\rho(\mathrm{d}x)}{|x - z||x - w|} \right] \\ &\geq |z - w| \left[1 - \frac{\rho(\mathbb{R})}{\Im(z)\Im(w)} \right] > 0, \end{aligned}$$

thereby verifying the injectivity.

For all z with $\Im(z) = \sigma$ we have

$$\Im[F_\mu(z)] = \Im(z) + \int_{\mathbb{R}} \frac{\Im(z)}{(x - \Re(z))^2 + (\Im(z))^2} \rho(\mathrm{d}x) \leq \Im(z) + \frac{\sigma^2}{\Im(z)} = 2\sigma.$$

Combining this estimate and (6.5) yields that for every $w \in \mathbb{C}^+_{2\sigma}$ the curve $\{F_\mu(z) : z \in C\}$, where C is the boundary of a large square D in $\mathbb{C}^+_\sigma$ that has horizontal edges $[-R, R] + \mathrm{i}\sigma$ and $[-R, R] + \mathrm{i}(\sigma + 2R)$, surrounds the point w. By the argument principle, the equation $F_\mu(z) = w$ has a solution $z \in D$. $\square$

Proposition 6.7. *Let $(F_{s,t})$ be a $\mathscr{P}_2$-REF. Then each map $F_{s,t}$ is injective on $\mathbb{C}^+$.*

Proof. Let $z, w \in \mathbb{C}^+$ with $z \neq w$ and $0 \le s \le t$. We select $\varepsilon > 0$ so that $z, w \in \mathbb{C}^+_\varepsilon$ and take $s = s_0 < s_1 < s_2 < \cdots < s_n = t$ such that $\min_{0\le i\le n-1} |v_{s_i,s_{i+1}}| < \varepsilon^2$, which is possible by the continuity of $(r,u) \mapsto v_{r,u}$ and the fact $v_{r,r} = 0$. Let

$$\begin{aligned} z_k &:= F_{s_k,s_{k+1}} \circ \cdots \circ F_{s_{n-1},s_n}(z), \\ w_k &:= F_{s_k,s_{k+1}} \circ \cdots \circ F_{s_{n-1},s_n}(w), \qquad 0 \le k \le n-1. \end{aligned}$$

Note that $F_{s,t}(z) = z_0$ and $F_{s,t}(w) = w_0$. By the inequality $\Im[F_{s_{n-1},s_n}(z)] \ge \Im(z)$ (see (4.18)) and by Lemma 6.6, the points z_{n-1} and w_{n-1} are distinct and lie in $\mathbb{C}^+_\varepsilon$; recall that $v_{s,t}$ is the variance of the probability measure associated with $F_{s,t}$. By induction on k in the decreasing direction, we can prove z_k, w_k are distinct and lie in $\mathbb{C}^+_\varepsilon$ for all $k = n-1, n-2, \dots, 0$, so that $F_{s,t}(z) = z_0 \neq w_0 = F_{s,t}(w)$. □

Theorem 6.8. *There is a one-to-one correspondence between $\mathscr{P}_2$-REFs and $\mathscr{P}_2$-DLCs given by the maps*

$$(F_{s,t})_{(s,t)\in\triangle} \mapsto (F_{0,t})_{t\ge 0}, \tag{6.7}$$

$$(F_t)_{t\ge 0} \mapsto (F_s^{-1} \circ F_t)_{(s,t)\in\triangle}. \tag{6.8}$$

Proof. Given $(F_{s,t})$, let $F_t := F_{0,t}$. Conditions (L1), (L2) and (L6) are obvious. By Proposition 6.7, each F_t is injective on $\mathbb{C}^+$. Moreover, $F_t \circ F_{t,u} = F_u$ holds for $0 \le t \le u$, which implies that $F_u(\mathbb{C}^+) = F_t(F_{t,u}(\mathbb{C}^+)) \subseteq F_t(\mathbb{C}^+)$, i.e., $t \mapsto F_t(\mathbb{C}^+)$ is non-increasing. Thus $(F_t)_{t\ge 0}$ is a $\mathscr{P}_2$-DLC.

Conversely, given a $\mathscr{P}_2$-DLC $(F_t)_{t\ge 0}$, the assumption $F_s(\mathbb{C}^+) \supseteq F_t(\mathbb{C}^+), s \le t$ allows us to define the composed map $F_{s,t} := F_s^{-1} \circ F_t$ as a self-map of $\mathbb{C}^+$. It is well known that the inverse map of a holomorphic function is also holomorphic, so that $F_{s,t}$ is holomorphic. Conditions (R3) and (R4) are obvious.

Condition (R1). For each $s \ge 0$, recall from Lemma 6.6 that the subdomain $F_s(\mathbb{C}^+_\sigma)$ for the inverse map F_s^{-1} contains $\mathbb{C}^+_{2\sigma}$, where $\sigma^2 := v_s$ is the variance of the underlying probability measure. From (6.5), $w \to \infty$ implies $z = F_s^{-1}(w) \to \infty$, so that we have

$$\lim_{\substack{w\to\infty \\ w\in\mathbb{C}^+_{2\sigma}}} \frac{F_s^{-1}(w)}{w} = \lim_{\substack{z\to\infty \\ z\in\mathbb{C}^+_{\sigma}}} \frac{z}{F_s(z)} = 1.$$

Then

$$\sphericalangle \lim_{z\to\infty} \frac{F_{s,t}(z)}{z} = \sphericalangle \lim_{z\to\infty} \frac{F_s^{-1}(F_t(z))}{F_t(z)} \cdot \frac{F_t(z)}{z} = 1,$$

which verifies condition (R1).

Condition (R2). It suffices to show the continuity of $(s,t) \mapsto F_{s,t}(z)$ at each $z_0 \in \mathbb{C}^+$, see Remark 6.3. We use the Lagrange inversion formula. Let $(s_0, t_0) \in \triangle$. Since $F_{t_0}(z_0) \in F_{s_0}(\mathbb{C}^+)$, there exists an open disk D such that $\overline{D} \subseteq \mathbb{C}^+$ and $F_{t_0}(z_0) \in F_{s_0}(D)$. By the continuity of $t \mapsto F_t$ and since $F_{s_0}(D)$ is open, we can find open intervals $I \ni s_0$ and $J \ni t_0$ such that for all $s \in I$ and $t \in J$, the closed curve $F_s(\partial D)$ surrounds $F_t(z_0)$ and so $F_t(z_0) \in F_s(D)$. Therefore, by the Lagrange inversion formula,

$$F_{s,t}(z_0) = F_s^{-1}(F_t(z_0)) = \frac{1}{2\pi \mathrm{i}} \int_{\partial D} \frac{w F_s'(w)}{F_s(w) - F_t(z_0)} \,\mathrm{d}w, \qquad s \in I,\ t \in J,\ s \le t.$$

The continuity $(s,t) \mapsto F_{s,t}(z_0)$ at (s_0,t_0) is now a consequence of the dominated convergence theorem.

Condition (R5). If we denote by μ_t and $\mu_{s,t}$ the underlying probability measures for F_t and $F_{s,t}$, respectively, then $\mu_s \triangleright \mu_{s,t} = \mu_t$. Since μ_t has finite second moment due to (6.4), $\mu_{s,t}$ also has finite second moment from Proposition 5.13. Therefore, formula (6.3) holds and from Remark 6.3 c) the numbers $m_{s,t}$ and $v_{s,t}$ satisfy

$$\begin{aligned} m_{s,t} &= \mathrm{m}_1(\mu_{s,t}) = \mathrm{m}_1(\mu_t) - \mathrm{m}_1(\mu_s) = m_t - m_s, \\ v_{s,t} &= \mathrm{Var}(\mu_{s,t}) = \mathrm{Var}(\mu_t) - \mathrm{Var}(\mu_s) = v_t - v_s. \end{aligned}$$

This implies the continuity of $(s,t) \mapsto (m_{s,t}, v_{s,t})$, so that (R5) holds true. □

Example 6.9. Let F_t be the reciprocal Cauchy transform of the semicircle distribution $\mathrm{S}(0,t)$ for $t > 0$ and $F_0 := \mathrm{id}_{\mathbb{C}^+}$. The family $(F_t)_{t\ge 0}$ is a $\mathscr{P}_2^0$-DLC, see Example 6.5. This can be more easily shown from Theorem 6.8 by finding the corresponding REF. First, a formal algebraic calculation yields a formal inverse function $F_s^{-1}(z) = z + s/z$. Therefore, the corresponding REF should be $F_s^{-1} \circ F_t$, which is

$$F_{s,t}(z) := \frac{1}{2}\left(1 + \frac{s}{t}\right) z + \frac{1}{2}\left(1 - \frac{s}{t}\right) \sqrt{z^2 - 4t}, \qquad 0 \le s \le t.$$

We can check this is a Nevanlinna function with $F_{s,t}(\mathrm{i}y)/(\mathrm{i}y) \to 1$, so that it is the reciprocal Cauchy transform of a probability measure $\mu_{s,t}$. We can also check that $F_{s,t}(z) = z - (t-s)/z + O(z^{-2}),\ z \to \infty$ and $F_{s,t} \circ F_{t,u} = F_{s,u}$, $0 \le s \le t \le u$. Therefore $(F_{s,t})$ is a $\mathscr{P}_2^0$-REF, and so $F_t = F_{0,t}$ form a $\mathscr{P}_2^0$-DLC.

6.2. **Integral/Integro-differential equations.** Infinitesimal descriptions are helpful to better understand reverse evolution families; later we will see in Section 7.1 that an infinitesimal description is useful for constructing an operator model for monotone additive processes. For example, suppose that the limit

$$A(s,z) = \lim_{h\to 0^+} \frac{F_{s-h,s}(z) - z}{h}$$

exists. Then taking the derivative of $F_{s,u} = F_{s,t} \circ F_{t,u}$ with respect to s at $s = t$ yields the non-autonomous ODE

$$\frac{\partial F_{t,u}}{\partial t} = A(t, F_{t,u}(z)).$$

In the following, we establish a refined description of $\mathscr{P}_2^0$-REFs. In general, $F_{s,t}$ need not be differentiable in time, but still an integral/integro-differential equation holds. As a key lemma we use the following version of Radon-Nikodym's theorem that generalizes Lebesgue's differentiation theorem. The reader is referred to [62, Theorem 2, Section 1.6] or [110, Theorem 5.8] for a proof. In the latter reference, the result is called the Lebesgue–Besicovitch differentiation theorem.

Lemma 6.10. *Let μ and ν be locally finite Borel measures on $\mathbb{R}$. Suppose that μ is absolutely continuous with respect to ν. Then the limit*

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x) := \lim_{h\to 0^+} \frac{\mu(x-h,x+h)}{\nu((x-h,x+h))} \in [0,+\infty)$$

exists at ν-a.e. x, and it serves as a Radon-Nikodym derivative, i.e.,

$$\mu(B) = \int_B \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x)\,\nu(\mathrm{d}x), \qquad B \in \mathcal{B}(\mathbb{R}).$$

Theorem 6.11. *Let $(F_{s,t})_{(s,t)\in\triangle}$ be a $\mathscr{P}_2^0$-REF having the asymptotic behavior (6.3). Let τ be the Lebesgue-Stieltjes measure on $[0,+\infty)$ associated with the non-decreasing continuous function $t \mapsto v_{0,t}$. There exists a probability kernel $\dot\rho$ from $[0,+\infty)$ to $\mathbb{R}$ such that for all $0 \le s \le t$*

$$F_{s,t}(z) = z + \int_s^t \left[\int_{\mathbb{R}} \frac{1}{x - F_{a,t}(z)}\,\dot\rho(a,\mathrm{d}x)\right]\tau(\mathrm{d}a), \tag{6.9}$$

$$F_{s,t}(z) = z + \int_s^t \frac{\partial F_{s,b}}{\partial z}(z)\left[\int_{\mathbb{R}} \frac{1}{x-z}\,\dot\rho(b,\mathrm{d}x)\right]\tau(\mathrm{d}b). \tag{6.10}$$

Either of (6.9) and (6.10) implies the uniqueness of $\dot\rho$ in the sense that if another probability kernel $\dot\sigma$ exists, then we must have $\dot\rho(t,\cdot) = \dot\sigma(t,\cdot)$ for τ-a.e. $t \ge 0$.

Proof. From Remark 6.3 d) each $F_{s,t}$ is of the form

$$F_{s,t}(z) = z + \int_{\mathbb{R}} \frac{\rho_{s,t}(\mathrm{d}x)}{x-z}$$

and $v_{s,t} = \rho_{s,t}(\mathbb{R})$. We set $v_t := v_{0,t}$. The fact $F_{0,0} = \mathrm{id}_{\mathbb{C}^+}$ implies $v_0 = 0$. From Remark 6.3 c), we have $v_s + v_{s,t} = v_t$; in particular, $t \mapsto v_t$ is non-decreasing. Note that the measure τ is atomless as $t \mapsto v_t$ is continuous.

Let us consider the finite Borel measure $\overline{\rho}_{r,s}$ on $\widehat{\mathbb{R}}$ defined by

$$\overline{\rho}_{r,s}(\{\infty\}) = 0, \qquad \overline{\rho}_{r,s}\restriction_{\mathbb{R}} := \begin{cases} \frac{\rho_{r,s}}{v_{r,s}}, & \text{if } v_{r,s} > 0, \\ 0, & \text{otherwise.}\end{cases}$$

Since $\overline{\rho}_{r,s}(\widehat{\mathbb{R}}) \le 1$, for each $s \ge 0$ we can find a sequence $h_n \to 0^+$ such that $\overline{\rho}_{s-h_n,s+h_n}$ converges weakly to a Borel measure $\overline{\rho}_s$ on $\widehat{\mathbb{R}}$ with $\overline{\rho}_s(\widehat{\mathbb{R}}) \le 1$; be aware that h_n depends on s. The following convergence holds whenever $v_{s+h_n} > v_{s-h_n}$ for all $n \in \mathbb{N}$:

$$\frac{F_{s-h_n,s+h_n}(z) - z}{v_{s+h_n} - v_{s-h_n}} = \int_{\mathbb{R}} \frac{\overline{\rho}_{s-h_n,s+h_n}(\mathrm{d}x)}{x-z} \to \int_{\mathbb{R}} \frac{\overline{\rho}_s(\mathrm{d}x)}{x-z} = -G_{\overline{\rho}_s\restriction_{\mathbb{R}}}(z), \qquad n \to \infty. \tag{6.11}$$

Proof of (6.10). Observe that

$$\begin{aligned} F_{r,t}(z) - F_{r,s}(z) &= F_{r,s}(F_{s,t}(z)) - F_{r,s}(z) \\ &= [F_{s,t}(z) - z]\left[1 + \int_{\mathbb{R}} \frac{\rho_{r,s}(\mathrm{d}x)}{(x-z)(x-F_{s,t}(z))}\right] \end{aligned} \tag{6.12}$$

and

$$|F_{s,t}(z) - z| = \left|\int_{\mathbb{R}} \frac{\rho_{s,t}(\mathrm{d}x)}{x-z}\right| \le \frac{v_t - v_s}{\Im(z)}.$$

Therefore, for any fixed $r \ge 0$ and $z \in \mathbb{C}^+$, the function $f(s) := F_{r,s}(z)$ satisfies

$$|f(t) - f(s)| \le C_{t,z}(v_t - v_s), \qquad r \le s \le t,$$

where $C_{t,z} := 1/\Im(z) + v_t/\Im(z)^3$. In particular, for any fixed $T > 0$, f is of bounded variations on $[r,T]$, and so there exists a complex Borel measure ν on $(r,T]$ such that $\nu((s,t]) = f(t) - f(s)$ for $r \le s \le t \le T$. The inequality $|\nu((s,t])| \le C_{T,z}\tau((s,t])$ extends to

$$|\nu(B)| \le C_{T,z}\tau(B), \qquad B \in \mathcal{B}((r,T]). \tag{6.13}$$

To show (6.13), we consider the set $\mathcal{M} := \{B \in \mathcal{B}((r,T]) : |\nu(B)| \le C_{T,z}\tau(B)\}$, which is a monotone class. We also consider the algebra $\mathcal{A}$ consisting of the empty set and finite unions of disjoint intervals of the form $(s,t]$ ($r \le s \le t \le T$). It is easy to check that $\mathcal{M}$ contains $\mathcal{A}$. By the monotone class theorem (Theorem 4.3), $\mathcal{M}$ contains $\sigma(\mathcal{A}) = \mathcal{B}((r,T])$.

Because of the arbitrariness of T, we can extend ν to a complex measure on $[r,+\infty)$ with domain the set of bounded Borel subsets. Inequality (6.13) implies that ν is absolutely continuous with respect to τ. By Lemma 6.10, the limit

$$\mathrm{D}^{v(s)}F_{r,s}(z) := \lim_{h\to 0^+} \frac{\nu((s-h,s+h))}{\tau((s-h,s+h))} = \lim_{h\to 0^+} \frac{F_{r,s+h}(z) - F_{r,s-h}(z)}{v_{s+h} - v_{s-h}} \in \mathbb{C} \tag{6.14}$$

exists at τ-a.e. $s \in (r,+\infty)$. Let $J_{r,z}$ be the set of all $s \in (r,+\infty)$ such that this limit exists. Dividing (6.12) by $v_t - v_s$, replacing (s,t) with $(s-h,s+h)$ and passing to the limit $h \to 0^+$ yields

$$\lim_{h\to 0^+} \frac{F_{s-h,s+h}(z) - z}{v_{s+h} - v_{s-h}} = \frac{\mathrm{D}^{v(s)}F_{r,s}(z)}{1 + \int_{\mathbb{R}}(x-z)^{-2}\,\rho_{r,s}(\mathrm{d}x)} = \frac{\mathrm{D}^{v(s)}F_{r,s}(z)}{\frac{\partial}{\partial z}F_{r,s}(z)}, \qquad s \in J_{r,z},\ z \in \mathbb{C}^+. \tag{6.15}$$

Note here that the weak convergence $\rho_{r,s-h} \to \rho_{r,s}$ shown in Remark 6.3 d) was used in the first equality above.

On the other hand, at every $s \in J_{r,z}$, the inequality $v_{s+h} > v_{s-h}$ holds for all $h > 0$, and so we can deduce from (6.11) that

$$\lim_{h\to 0^+} \frac{F_{s-h,s+h}(z) - z}{v_{s+h} - v_{s-h}} = \lim_{h\to 0^+} \int_{\mathbb{R}} \frac{\overline{\rho}_{s-h,s+h}(\mathrm{d}x)}{x-z} = \int_{\mathbb{R}} \frac{\overline{\rho}_s(\mathrm{d}x)}{x-z}, \qquad s \in J_{r,z},\ z \in \mathbb{C}^+. \tag{6.16}$$

Combining (6.15) and (6.16) yields

$$\mathrm{D}^{v(s)} F_{r,s}(z) = \frac{\partial F_{r,s}}{\partial z}(z) \int_{\mathbb{R}} \frac{1}{x-z}\, \overline{\rho}_s(\mathrm{d}x), \qquad s \in J_{r,z},\ z \in \mathbb{C}^+. \tag{6.17}$$

Applying Lemma 6.10 to the measure ν and $B = [r,t]$ and using (6.17), we get

$$\begin{aligned} F_{r,t}(z) - z &= \int_r^t \mathrm{D}^{v(s)} F_{r,s}(z)\, \tau(\mathrm{d}s) \\ &= \int_r^t \frac{\partial F_{r,s}}{\partial z}(z) \left[\int_{\mathbb{R}} \frac{1}{x-z}\, \overline{\rho}_s(\mathrm{d}x)\right] \tau(\mathrm{d}s), \qquad z \in \mathbb{C}^+,\ 0 \le r \le t. \end{aligned} \tag{6.18}$$

Here we take any countable subset $C \subseteq \mathbb{C}^+$ having an accumulation point in $\mathbb{C}^+$, e.g. $C = \{\mathrm{i} + 1/n : n \in \mathbb{N}\}$, and set $J := \bigcap_{z\in C} J_{0,z}$. For any $s \in J$ and $z \in C$ the convergence in (6.16) holds. Therefore, by Proposition 4.29, the convergence in (6.16) holds for all $z \in \mathbb{C}^+$ and $s \in J$. This implies that $J \ni s \mapsto G_{\overline{\rho}_s\restriction_{\mathbb{R}}}(z)$ is measurable for each $z \in \mathbb{C}^+$, so that Lemma 5.10 implies that $J \ni s \mapsto \overline{\rho}_s(B)$ is measurable for each $B \in \mathcal{B}(\mathbb{R})$.

By the dominated convergence theorem applied to (6.18) (the estimates in (5.15) are helpful), we can see that

$$v_t = \lim_{y\to+\infty} [F_{0,t}(\mathrm{i}y) - \mathrm{i}y]\mathrm{i}y = \int_0^t \overline{\rho}_s(\mathbb{R})\tau(\mathrm{d}s).$$

Since $\overline{\rho}_s(\mathbb{R}) \le 1$ and $\tau([0,t]) = v_t$, we must have $\overline{\rho}_s(\mathbb{R}) = 1$ for τ-a.e. $s \ge 0$. We can then define $\dot\rho(s,\cdot) := \overline{\rho}_s\restriction_{\mathbb{R}}$ whenever $\overline{\rho}_s(\mathbb{R}) = 1$ and $s \in J$, and otherwise define $\dot\rho(s,\cdot) := \delta_0$. This is a probability kernel. Now formula (6.18) yields the desired formula (6.10).

Proof of (6.9). Observe that for $0 \le r \le s \le t$

$$F_{s,t}(z) - F_{r,t}(z) = F_{s,t}(z) - F_{r,s}(F_{s,t}(z)) = \int_{\mathbb{R}} \frac{\rho_{r,s}(\mathrm{d}x)}{F_{s,t}(z) - x}, \tag{6.19}$$

and so $|F_{s,t}(z) - F_{r,t}(z)| \le (v_s - v_r)/\Im(z)$. As before, for any (t,z) there exists a complex Borel measure μ on $[0,t]$ such that $\mu((r,s]) = F_{s,t}(z) - F_{r,t}(z)$ for $0 \le r \le s \le t$ and

$$|\mu(B)| \le \frac{\tau(B)}{\Im(z)}, \qquad B \in \mathcal{B}((0,t]). \tag{6.20}$$

Inequality (6.20) implies that μ is absolutely continuous with respect to τ. By Lemma 6.10 the limit

$$\mathrm{D}_{v(s)} F_{s,t}(z) := \lim_{h\to 0^+} \frac{F_{s+h,t}(z) - F_{s-h,t}(z)}{v_{s+h} - v_{s-h}} \in \mathbb{C} \tag{6.21}$$

exists at τ-a.e. $s \in (0,t)$. At any $s \in J$ where this limit exists, (6.19) and (6.16) yield

$$\mathrm{D}_{v(s)} F_{s,t}(z) = \lim_{h\to 0^+} \int_{\widehat{\mathbb{R}}} \frac{\overline{\rho}_{s-h,s+h}(\mathrm{d}x)}{F_{s+h,t}(z) - x} = \int_{\widehat{\mathbb{R}}} \frac{\overline{\rho}_s(\mathrm{d}x)}{F_{s,t}(z) - x} = \int_{\mathbb{R}} \frac{\overline{\rho}_s(\mathrm{d}x)}{F_{s,t}(z) - x}. \tag{6.22}$$

The last equality holds because the integrand vanishes at $x = \infty$. The second equality holds because of the continuity of $r \mapsto F_{r,t}$, i.e., in the triangular inequality

$$\begin{aligned} &\left|\int_{\widehat{\mathbb{R}}} \frac{\overline{\rho}_{s-h,s+h}(\mathrm{d}x)}{F_{s+h,t}(z) - x} - \int_{\widehat{\mathbb{R}}} \frac{\overline{\rho}_s(\mathrm{d}x)}{F_{s,t}(z) - x}\right| \\ &\le \underbrace{\left|\int_{\widehat{\mathbb{R}}} \frac{\overline{\rho}_{s-h,s+h}(\mathrm{d}x)}{F_{s+h,t}(z) - x} - \int_{\widehat{\mathbb{R}}} \frac{\overline{\rho}_{s-h,s+h}(\mathrm{d}x)}{F_{s,t}(z) - x}\right|}_{=:\, I_h^1} + \underbrace{\left|\int_{\widehat{\mathbb{R}}} \frac{\overline{\rho}_{s-h,s+h}(\mathrm{d}x)}{F_{s,t}(z) - x} - \int_{\widehat{\mathbb{R}}} \frac{\overline{\rho}_s(\mathrm{d}x)}{F_{s,t}(z) - x}\right|}_{=:\, I_h^2} \end{aligned}$$

the second term I_h^2 tends to zero due to (6.16) that holds for $s \in J$ and $z \in \mathbb{C}^+$, and the first term I_h^1 also converges to zero because

$$I_h^1 \le \int_{\widehat{\mathbb{R}}} \frac{|F_{s+h,t}(z) - F_{s,t}(z)|}{|F_{s+h,t}(z) - x||F_{s,t}(z) - x|}\, \overline{\rho}_{s-h,s+h}(\mathrm{d}x) \le \frac{|F_{s+h,t}(z) - F_{s,t}(z)|}{\Im(z)^2}.$$

By Lemma 6.10 and (6.22), we have

$$F_{s,t}(z) - z = -\int_s^t \mathrm{D}_{v(a)} F_{a,t}(z)\, \tau(\mathrm{d}a) = \int_s^t \left[\int_{\mathbb{R}} \frac{\overline{\rho}_a(\mathrm{d}x)}{x - F_{a,t}(z)}\right] \tau(\mathrm{d}a),$$

which is nothing but (6.9).

Uniqueness of $\dot\rho$. For example, we assume that (6.9) holds for $\dot\rho$ and $\dot\sigma$. We fix $T > 0$ and $z \in \mathbb{C}^+$ for some time. The complex measure

$$\lambda(B) := \int_B \left[\int_{\mathbb{R}} \frac{1}{x - F_{s,T}(z)}\, \dot\rho(s,\mathrm{d}x)\right] \tau(\mathrm{d}s), \qquad B \in \mathcal{B}([0,T]),$$

is absolutely continuous with respect to τ, so by Lemma 6.10 and uniqueness of Radon-Nikodym derivative, we obtain

$$\int_{\mathbb{R}} \frac{1}{x - F_{s,T}(z)}\, \dot\rho(s, \mathrm{d}x) = \lim_{h\to 0+} \frac{\lambda((s-h, s+h))}{\tau((s-h, s+h))} = \lim_{h\to 0+} \frac{F_{s-h,T}(z) - F_{s+h,T}(z)}{\tau((s-h, s+h))}$$

at τ-a.e. s. The same formula holds for $\dot\sigma$, so we obtain $G_{\dot\rho(s,\cdot)} = G_{\dot\sigma(s,\cdot)}$ on $F_{s,T}(\mathbb{C}^+)$ for τ-a.e. $s \in [0, T]$. By the identity theorem, the equality $G_{\dot\rho(s,\cdot)} = G_{\dot\sigma(s,\cdot)}$ holds on $\mathbb{C}^+$ and then we have $\dot\rho(s, \cdot) = \dot\sigma(s, \cdot)$ for τ-a.e. $s \in [0, T]$. A similar idea works for the case when we assume (6.10) instead. □

We also verify the converse direction: given τ and $\dot\rho$, solving these equations gives a unique $\mathscr{P}_2^0$-REF. The uniqueness is formulated in a stronger form.

Theorem 6.12. *Let τ be an atomless locally finite Borel measure on $[0, +\infty)$, $\dot\rho$ be a probability kernel from $[0, +\infty)$ to $\mathbb{R}$, and*

$$A(t, z) := \int_{\mathbb{R}} \frac{1}{x - z}\, \dot\rho(t, \mathrm{d}x).$$

(i) *For each fixed $t \ge 0$ and $z \in \mathbb{C}^+$, the integral equation*

$$f(s) = z + \int_s^t A(r, f(r))\, \tau(\mathrm{d}r), \qquad s \in [0, t], \tag{6.23}$$

has a unique solution $f(s) = f(s; t, z) \in \mathbb{C}^+$, $s \in [0, t]$, such that $[0, t] \ni s \mapsto f(s) \in \mathbb{C}^+$ is continuous.

(ii) *For each fixed $s \ge 0$, the integro-differential equation*

$$h(t, z) = z + \int_s^t \frac{\partial h}{\partial z}(r, z) A(r, z)\, \tau(\mathrm{d}r), \qquad t \in [s, +\infty),\ z \in \mathbb{C}^+, \tag{6.24}$$

has a unique solution $h(t, z) = h(t, z; s) \in \mathbb{C}^+$, $t \in [s, +\infty)$, $z \in \mathbb{C}^+$, such that $t \mapsto h(t, z)$ is continuous for each fixed z and $z \mapsto h(t, z)$ is holomorphic for each fixed t.

Moreover, $f(s; t, z) = h(t, z; s)$ holds for all $(s, t) \in \triangle$, $z \in \mathbb{C}^+$, and $F_{s,t}(z) := f(s; t, z)$ forms a $\mathscr{P}_2^0$-REF.

Proof. Uniqueness of a solution to (6.23). Let $t \ge 0$ and $z \in \mathbb{C}^+$ be fixed. Suppose that continuous functions $f_1, f_2 \colon [0, t] \to \mathbb{C}^+$ satisfy (6.23). We easily obtain for $F(s) := f_1(s) - f_2(s)$

$$\begin{aligned} |F(s)| &\le \int_s^t \left[\int_{\mathbb{R}} \frac{|f_1(r) - f_2(r)|}{|x - f_1(r)||x - f_2(r)|}\, \dot\rho(r, \mathrm{d}x)\right] \tau(\mathrm{d}r) \\ &\le \frac{1}{\Im(z)^2} \int_s^t |F(r)|\, \tau(\mathrm{d}r), \qquad s \in [0, t]. \end{aligned}$$

Iterating this inequality yields

$$\begin{aligned} |F(s)| &\le \frac{1}{\Im(z)^{2n}} \int_{s \le s_1 \le s_2 \le \cdots \le s_n \le t} |F(s_n)|\, \tau^{\otimes n}(\mathrm{d}s_1 \mathrm{d}s_2 \cdots \mathrm{d}s_n) \\ &\le \frac{\tau([s, t])^n}{n!\Im(z)^{2n}} \sup_{r \in [0,t]} |F(r)|, \qquad s \in [0, t],\ n \in \mathbb{N}, \end{aligned} \tag{6.25}$$

where we used the fact

$$\tau^{\otimes n}(\{(s_1, s_2, \dots, s_n) : s \le s_1 \le s_2 \le \cdots \le s_n \le t\}) = \frac{\tau([s, t])^n}{n!}.$$

Since the right-hand side of (6.25) tends to zero as $n \to \infty$, we conclude $F \equiv 0$.

Existence of a solution to (6.23). The proof is based on Picard's iteration. We recursively define $F^0, F^1, F^2, \dots,$ by $F^0_{s,t}(z) \equiv z$ and

$$F^n_{s,t}(z) = z + \int_s^t A(r, F^{n-1}_{r,t}(z))\, \tau(\mathrm{d}r), \qquad (s, t) \in \triangle,\ z \in \mathbb{C}^+,\ n \in \mathbb{N}. \tag{6.26}$$

Inductive arguments easily show that $F^n_{s,t}$ is holomorphic in $\mathbb{C}^+$ and satisfies $\Im[F^n_{s,t}(z)] \ge \Im(z)$. We then obtain from (6.26)

$$\begin{aligned} |F^n_{s,t}(z)| &\le |z| + \frac{\tau([s, t])}{\Im(z)}, \\ |F^n_{s,t}(z) - F^n_{s',t'}(z)| &\le \frac{\tau([s, t]\Delta[s', t'])}{\Im(z)}, \end{aligned}$$

where $A\Delta B$ is the symmetric difference $(A \setminus B) \cup (B \setminus A)$. In particular, for each fixed $z \in \mathbb{C}^+$, the sequence of functions $f^z_n \colon \triangle \to \mathbb{C}^+$, $f^z_n(s, t) := F^n_{s,t}(z)$ is uniformly bounded and equicontinuous on each compact subset of $\triangle$; note that the latter follows from the (uniform) continuity of the function $t \mapsto \tau([0, t])$. We can therefore use Arzéla-Ascoli's theorem to find a subsequence $(f^z_{n(k)})_{k \ge 1}$ that converges to a function $f^z \colon \triangle \to \mathbb{C}^+$ locally uniformly. Passing to the limit in (6.26), the limit function $F_{s,t}(z) := f^z(s, t)$ satisfies equation (6.9) and hence $f(s) := F_{s,t}(z)$ satisfies (6.23).

The solution to (6.23) *forms a $\mathscr{P}_2^0$-REF.* As already proved, $F^n_{s,t}$ is a Nevanlinna function for each (s, t), so that its pointwise limit $F_{s,t}$ is also a Nevanlinna function; see Proposition 4.29. Moreover, the integral equation (6.9) and the dominated convergence theorem, together with the bound $1/|x - F_{a,t}(z)| \le 1/\Im(z)$, yield $\lim_{y\to+\infty} F_{s,t}(\mathrm{i}y)/(\mathrm{i}y) = 1$ and

$$\lim_{y\to+\infty} \mathrm{i}y[F_{s,t}(\mathrm{i}y) - \mathrm{i}y] = \tau([s, t]).$$

It remains to show $F_{s,t} \circ F_{t,u} = F_{s,u}$. For each fixed $z \in \mathbb{C}^+$ and $0 \le t \le u$, let $F_1, F_2 \colon [0,u] \to \mathbb{C}^+$ be defined by $F_1(s) := F_{s,u}(z)$ and

$$F_2(s) := \begin{cases} F_{s,t}(F_{t,u}(z)), & s \in [0,t], \\ F_{s,u}(z), & s \in (t,u], \end{cases}$$

which is continuous. Recalling the equation

$$F_{s,t}(z) = z + \int_s^t A(r, F_{r,t}(z))\, \tau(\mathrm{d}r), \qquad z \in \mathbb{C}^+,\ 0 \le s \le t,$$

we have, for $s \in [0,t]$,

$$\begin{aligned} F_2(s) &= F_{t,u}(z) + \int_s^t A(r, F_2(r))\, \tau(\mathrm{d}r) \\ &= z + \int_t^u A(r, F_{r,u}(z))\, \tau(\mathrm{d}r) + \int_s^t A(r, F_2(r))\, \tau(\mathrm{d}r) \\ &= z + \int_s^u A(r, F_2(r))\, \tau(\mathrm{d}r), \end{aligned}$$

and for $s \in (t,u]$,

$$F_2(s) = F_{s,u}(z) = z + \int_s^u A(r, F_{r,u}(z))\, \tau(\mathrm{d}r) = z + \int_s^u A(r, F_2(r))\, \tau(\mathrm{d}r).$$

Therefore, F_2 satisfies exactly the same equation satisfied by F_1. By the trajectory-wise uniqueness, we conclude that $F_1 = F_2$ on $[0,u]$.

Existence of a solution to (6.24). We already constructed a $\mathscr{P}_2^0$-REF $(F_{s,t})$ that solves (6.9). On the other hand, by Theorem 6.11, there exists a probability kernel $\dot\sigma$ for which (6.9) and (6.10) hold, where $\dot\rho$ is replaced by $\dot\sigma$. In the same theorem the uniqueness of $\dot\rho$ is verified, so that $\dot\rho(t,\cdot) = \dot\sigma(t,\cdot)$ for τ-a.e. $t \ge 0$. Thus, $(F_{s,t})$ is also a solution to (6.10), so that $h(t,z) := F_{s,t}(z)$ satisfies (6.24).

Uniqueness of a solution to (6.24). Let $s \ge 0$ be fixed. Let $h_1, h_2 \colon [s,+\infty) \times \mathbb{C}^+ \to \mathbb{C}^+$ be solutions to (6.24) with prescribed assumptions. Note then that $h_i, \partial_z h_i$ are continuous on $[s,+\infty) \times \mathbb{C}^+$ thanks to Lemma 4.31; in particular, the integral in (6.24) is well defined. Since $|A(t,z)| \le 1/\Im(z)$, the function $H(t,z) := h_1(t,z) - h_2(t,z)$ satisfies

$$|H(t,z)| \le \frac{1}{\Im(z)} \int_s^t \left| \frac{\partial H}{\partial z}(r,z) \right| \tau(\mathrm{d}r). \tag{6.27}$$

By Cauchy's integral formula we obtain

$$\left| \frac{\partial H}{\partial z}(t,z) \right| = \frac{1}{2\pi} \left| \int_{C(z,\varepsilon)} \frac{H(t,w)}{(w-z)^2}\, \mathrm{d}w \right| \le \frac{1}{2\pi\varepsilon^2} \int_{C(z,\varepsilon)} |H(t,w)|\, |\mathrm{d}w|, \tag{6.28}$$

where $C(z,\varepsilon)$ is the circle centered at z with radius $\varepsilon \in (0, \Im z)$. Combining (6.27) and (6.28) gives

$$|H(t,z)| \le \frac{1}{2\pi\varepsilon^2 \Im(z)} \int_{[s,t] \times C(z,\varepsilon)} |H(r_1,w_1)|\, \tau(\mathrm{d}r_1) |\mathrm{d}w_1|. \tag{6.29}$$

Choosing $\varepsilon = \Im z/(2n)$, $n \in \mathbb{N}$, and iterating this inequality n times yields

$$|H(t,z)| \le \left(\frac{1}{2\pi\varepsilon^2} \right)^n \int_{[s,t]^n_{\ge} \times W_n} \frac{|H(r_n,w_n)|}{\Im(z)\Im(w_1)\cdots\Im(w_{n-1})}\, \tau^{\otimes n}(\mathrm{d}r_1 \cdots \mathrm{d}r_n) |\mathrm{d}w_1| \cdots |\mathrm{d}w_n|,$$

where $[s,t]^n_{\ge} := \{(r_1, r_2, \dots, r_n) \in [s,t]^n : r_1 \ge r_2 \ge \cdots \ge r_n\}$ and $W_n := \{(w_1, w_2, \dots, w_n) \in (\mathbb{C}^+)^n : w_1 \in C(z,\varepsilon),\ w_2 \in C(w_1,\varepsilon),\ \dots,\ w_n \in C(w_{n-1},\varepsilon)\}$. Since $w_1, w_2, \dots, w_n$ belong to the compact subset $K_z := \{w \in \mathbb{C}^+ : |w - z| \le \Im(z)/2\}$, by setting $M_{t,z} := \sup_{r \in [s,t], w \in K_z} |H(r,w)|$ we obtain

$$|H(t,z)| \le \frac{M_{t,z}}{[2\pi\varepsilon^2(\Im(z)/2)]^n} \cdot \frac{\tau([s,t])^n}{n!} \cdot (2\pi\varepsilon)^n = \frac{M_{t,z} n^n}{n!} \left(\frac{4\tau([s,t])}{\Im(z)^2} \right)^n.$$

By Stirling's formula, for sufficiently large n we have $n! \ge \frac{\sqrt{2\pi n}}{2}(n/e)^n$, so that

$$|H(t,z)| \le \frac{2M_{t,z}}{\sqrt{2\pi n}} \underbrace{\left(\frac{4e\tau([s,t])}{\Im(z)^2} \right)^n}_{=: (\alpha_{t,z})^n}.$$

If we take t close enough to s such that $\alpha_{t,z} < 1$, say for $s < t < s + \delta$, then letting $n \to \infty$ we obtain $H(t,z) = 0$. Then (6.29) reads

$$|H(t,z)| \le \frac{1}{2\pi\varepsilon^2 \Im(z)} \int_{[s+\delta,t] \times C(z,\varepsilon)} |H(r_1,w_1)|\, \tau(\mathrm{d}r_1) |\mathrm{d}w_1|.$$

Repeating the above calculations, we can prove $H(t,z) = 0$ for all $s + \delta < t < s + \delta + \delta'$. Actually, we can take $\delta' = \delta$ up to a fixed time $T > s$, so this procedure shows $H(t,z) = 0$ for all $t \ge s$. The reason we can choose $\delta' = \delta$ is that $t \mapsto \tau([0,t])$ is uniformly continuous on any fixed interval $[0,T]$, so that for any $\eta > 0$ there exists $\delta > 0$ such that $\tau([a,b]) < \eta$ whenever $a, b \in [s,T]$ and $|a - b| < \delta$. □

Corollary 6.13. *Let τ be an atomless locally finite Borel measure on $[0,+\infty)$ and $\dot\rho$ be a probability kernel from $[0,+\infty)$ to $\mathbb{R}$. Then there exists a unique $\mathscr{P}_2^0$-DLC $(F_t)_{t\ge0}$ such that*

$$F_t(z) = z + \int_0^t \frac{\partial F_s}{\partial z}(z)\left[\int_{\mathbb{R}} \frac{1}{x-z}\,\dot\rho(s,\mathrm{d}x)\right]\tau(\mathrm{d}s).$$

Overall, there is a one-to-one correspondence between the following four kinds of sets:

- monotone convolution hemigroups $(\mu_{s,t})_{(s,t)\in\triangle}$ such that each $\mu_{s,t}$ has vanishing mean and finite second moment that is continuous with respect to (s,t);
- $\mathscr{P}_2^0$-REFs;
- $\mathscr{P}_2^0$-DLCs;
- pairs $(\dot\rho,\tau)$ of a probability kernel $\dot\rho$ from $[0,+\infty)$ to $\mathbb{R}$ and an atomless locally finite Borel measure τ on $[0,+\infty)$.

We call $(\dot\rho,\tau)$ the **generator** of the other three objects. Actually, the generators can also be defined for $\mathscr{P}_2$-REFs, $\mathscr{P}_2$-DLCs and monotone convolution hemigroups with finite second moments; see [84].

In light of the generator, we offer a sufficient condition for a monotone convolution hemigroup to have locally uniform compact support.

Proposition 6.14. *Let $\dot\rho$ be a probability kernel from $[0,+\infty)$ to $\mathbb{R}$ and τ be an atomless locally finite Borel measure on $[0,+\infty)$. Let $(\mu_{s,t})$ be the corresponding monotone convolution hemigroup. Suppose that for every $T>0$ there exists $R_T>0$ such that $\dot\rho(t,\cdot)$ is supported on $[-R_T,R_T]$ for all $t\in[0,T]$. Then for every $T>0$ there exists $R'_T>0$ such that $\mu_{s,t}$ is supported on $[-R'_T,R'_T]$ for all $0\le s\le t\le T$.*

Proof. The proof is analogous to the proof of Proposition 5.18, part (3) $\Longrightarrow$ (1). For example, one can replace (5.18) with

$$|\tilde A(t,z)| \le C, \qquad |z| > R_T+1,\ t\in[0,T].$$

The details are omitted. □

6.3. **Notes.** C. Loewner introduced Loewner chains in 1923 to attack the Bieberbach conjecture, which lead to the positive solution by de Branges in 1985; the reader interested in the history is referred to the monograph [16] and the survey article [98]. Loewner theory has also found applications to other fields; in particular, applications to SLE (Stochastic Loewner Evolution) made a significant success in physics and probability theory [102].

The results in Section 6 are adopted from Hasebe, Hotta and Murayama [84]. Proposition 6.7 and Theorem 6.8 hold for more general REFs and DLCs, but we restricted the results to the $\mathscr{P}_2$-ones to simplify the proof. Lemma 6.6 was proved by Maassen [109, Lemma 2.4]. Example 6.9 of the REF associated with the semicircle distributions was give by Biane [35]. The original proof of Theorem 6.11 given in [84] was a reduction of time-continuous Loewner chains to the absolutely continuous ones that are already well studied by Goryainov and Ba [77] and Bauer [19]. Our proof is rather different and is more self-contained. Our proof of the uniqueness of a solution to equation (6.10) is different from Bauer's short proof in the absolutely continuous case. Actually, we could give a similar proof to Bauer's but that would require a "$\mathrm{D}_{v(s)}$- ($\mathrm{D}^{v(s)}$-)calculus", e.g., the Leibniz formula and the derivative of composite functions, which also require proofs. To avoid such an argument, we gave a tricky proof based on Picard's iteration.

In the absolutely continuous case, a more general Loewner theory has been established by Bracci, Contreras, Díaz-Madrigal and Gumenyuk [39, 50]. Schleißinger [140], Franz, Hasebe and Schleißinger [70], and Jekel [92] also proved results analogous to Theorems 6.11 and 6.12 in different but absolutely continuous setups.

At present, for a technical reason, we need the assumption of finite second moment to establish the integral/integro-differential equations. On the other hand, similar and more complete results were established for multiplicative monotone convolution hemigroups on the unit circle by Hasebe and Hotta [83].

Schleißinger studied comb products of non-identical rooted graphs and obtained an approximation of (continuous-time) monotone convolution hemigroups by discrete-time ones [137].

7. Monotone additive processes

In probability theory, an additive process is a continuous-time stochastic process whose increments are independent but may have time-dependent distributions. We define and construct a monotone additive process.

Definition 7.1. Let (A,φ) be a unital C^*-probability space. A family of real random variables $(x_t)_{t\ge0}$ in A is called a **monotone additive process**, or a process of monotonically independent increments, if the following conditions are satisfied.

(i) $x_0 = 0$.

(ii) $\triangle \ni (s,t) \mapsto \mu_{x_t-x_s}$ is weakly continuous.

(iii) for every $n\in\mathbb{N}$ and reals $0=t_0<t_1<\cdots<t_n$, the elements (called increments)

$$x_{t_1}-x_{t_0},\ x_{t_2}-x_{t_1},\ \ldots\ ,\ x_{t_n}-x_{t_{n-1}}$$

are monotonically independent.

Proposition 7.2. *Let $(x_t)_{t\ge0}$ be a monotone additive process in a unital C^*-probability space. Let $\mu_{s,t} := \mu_{x_t-x_s}$, the analytic distribution of x_t-x_s, for $0\le s\le t$. Then $(\mu_{s,t})_{(s,t)\in\triangle}$ is a monotone convolution hemigroup.*

Proof. It is obvious that $\mu_{t,t} = \delta_0$. The weak continuity holds by definition. From the decomposition $x_u - x_s = (x_t - x_s) + (x_u - x_t)$ and since $x_t - x_s$ and $x_u - x_t$ are monotonically independent, we have

$$\mu_{s,u} = \mu_{x_u - x_s} = \mu_{x_t - x_s} \rhd \mu_{x_u - x_t} = \mu_{s,t} \rhd \mu_{t,u}, \qquad 0 \le s \le t \le u. \qquad \square$$

A question is given a monotone convolution hemigroup, does there exist a monotone additive process that realizes the hemigroup? If the given hemigroup contains probability measures with unbounded support, the process cannot be realized in a unital C^*-probability space. We will therefore consider monotone convolution hemigroups consisting of probability measures with compact support. Then we can indeed construct a monotone additive process on a unital C^*-probability space. In fact, several constructions are known. Two of them are presented below.

7.1. **A construction on monotone Fock spaces.** Here we define a continuous monotone Fock space on which a monotone additive process can be canonically constructed. Suppose that $(\mu_{s,t})_{(s,t)\in\triangle}$ is a monotone convolution hemigroup of mean zero and finite second moment with generator $(\dot\rho, \tau)$ such that $\dot\rho$ is supported on compact subsets locally uniformly, i.e., for every $T > 0$ there exists $R_T > 0$ such that $\dot\rho(t, \cdot)$ is supported on $[-R_T, R_T]$ for all $t \in [0, T]$. From Proposition 6.14, $(\mu_{s,t})$ is compactly supported locally uniformly. Let $(F_{s,t})$ be the associated $\mathscr{P}_2^0$-REF. Recall that equation (6.9) holds.

For notational convenience, let $\mathbb{R}_+ := [0, +\infty)$ and Θ be the Borel measure on $\mathbb{R}_+ \times \mathbb{R}$ defined by $\Theta(\mathrm{d}t\mathrm{d}x) := \dot\rho(t, \mathrm{d}x)\tau(\mathrm{d}t)$, i.e.,

$$\Theta(B) = \int_{[0,+\infty)} \left[\int_{\mathbb{R}} \chi_B(t, x)\, \dot\rho(t, \mathrm{d}x)\right] \tau(\mathrm{d}t), \qquad B \in \mathcal{B}(\mathbb{R}_+ \times \mathbb{R}).$$

Let $(\mathbb{R}_+)^n_> := \{(t_1, t_2, \dots, t_n) \in (\mathbb{R}_+)^n : t_1 > t_2 > \cdots > t_n \ge 0\}$. We restrict the measure $\Theta^{\otimes n}$ on $(\mathbb{R}_+ \times \mathbb{R})^n \simeq (\mathbb{R}_+)^n \times \mathbb{R}^n$ to the subset $(\mathbb{R}_+)^n_> \times \mathbb{R}^n$ and define

$$\begin{aligned} H_n &:= L^2((\mathbb{R}_+)^n_> \times \mathbb{R}^n, \Theta^{\otimes n}) \\ &= \left\{ f : (\mathbb{R}_+)^n_> \times \mathbb{R}^n \to \mathbb{C} \;\middle|\; \int_{(\mathbb{R}_+)^n_> \times \mathbb{R}^n} |f(\boldsymbol{t}, \boldsymbol{x})|^2\, \Theta^{\otimes n}(\mathrm{d}\boldsymbol{t}\mathrm{d}\boldsymbol{x}) < +\infty \right\} \end{aligned}$$

equipped with the inner product

$$\langle f, g\rangle := \int_{(\mathbb{R}_+)^n_> \times \mathbb{R}^n} \overline{f(\boldsymbol{t}, \boldsymbol{x})} g(\boldsymbol{t}, \boldsymbol{x})\, \Theta^{\otimes n}(\mathrm{d}\boldsymbol{t}\mathrm{d}\boldsymbol{x}).$$

The algebraic monotone Fock space associated to Θ is the pre-Hilbert space

$$F^0_>(\Theta) := \mathbb{C}\Omega \oplus \bigoplus_{n=1}^{\infty} H_n,$$

where Ω is a unit vector and the direct sum is the algebraic one, i.e., each element of $F^0_>(\Theta)$ is a finite sum of elements of H_i's and $\mathbb{C}\Omega$. We also write $H_0 = \mathbb{C}\Omega$. The **monotone Fock space** is the completion of $F^0_>(\Theta)$ that can be defined as

$$F_>(\Theta) := \left\{ (h_n)_{n\in\mathbb{N}_0} : h_n \in H_n\ (n \in \mathbb{N}_0),\ \sum_{n\ge0} \|h_n\|^2_{H_n} < +\infty \right\}.$$

We consider the C^*-probability space $(\mathbb{B}(F_>(\Theta)), \varphi)$, where $\varphi = \langle \Omega, \cdot\, \Omega\rangle$. The identity operator on $F_>(\Theta)$ is denoted as $\mathbf{1}$. Let us first introduce three kinds of operators on the algebraic monotone Fock space, and then extend them to the monotone Fock space.

Definition 7.3. The symbols $\boldsymbol{t} = (t_1, t_2, \dots, t_n) \in (\mathbb{R}_+)^n_>$ and $\boldsymbol{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ are employed below.

(i) **Creation operator** $a^*(f)$, $f \in H_1$: its restriction to H_n is a map to H_{n+1} defined by

$$\begin{aligned} [a^*(f)h]((t, \boldsymbol{t}), (x, \boldsymbol{x})) &:= f(t, x) h(\boldsymbol{t}, \boldsymbol{x}), \qquad t > t_1,\ x \in \mathbb{R},\ h \in H_n,\ n \in \mathbb{N}, \\ a^*(f)\Omega &:= f. \end{aligned}$$

(ii) **Annihilation operator** $a(f)$, $f \in H_1$: its restriction to H_{n+1} is a map into H_n defined by

$$\begin{aligned} [a(f)h](\boldsymbol{t}, \boldsymbol{x}) &:= \int_{(t_1,+\infty)\times\mathbb{R}} \overline{f(t, x)} h((t, \boldsymbol{t}), (x, \boldsymbol{x}))\, \Theta(\mathrm{d}t\mathrm{d}x), \qquad h \in H_{n+1},\ n \in \mathbb{N}, \\ a(f)h &:= \langle f, h\rangle_{H_1} \Omega, \qquad h \in H_1, \\ a(f)\Omega &:= 0. \end{aligned}$$

(iii) **Gauge operator** $\lambda(g)$, $g \in L^\infty(\Theta)$: its restriction to H_n is a map into itself given by

$$\begin{aligned} [\lambda(g)h](\boldsymbol{t}, \boldsymbol{x}) &:= g(t_1, x_1) h(\boldsymbol{t}, \boldsymbol{x}), \qquad h \in H_n,\ n \in \mathbb{N}, \\ \lambda(g)\Omega &:= 0. \end{aligned}$$

Note that $f \mapsto a^*(f)$ is linear while $f \mapsto a(f)$ is antilinear.

Proposition 7.4. *For any $f \in H_1$ and $g \in L^\infty(\Theta)$, we have*

$$\|a^*(f)\| = \|f\|_{H_1}, \qquad \|a(f)\| = \|f\|_{H_1}, \qquad \|\lambda(g)\| = \|g\|_{L^\infty}.$$

Proof. On H_0, we have $\|a^*(f)\Omega\|_{F^0_>(\Theta)} = \|f\|_{F^0_>(\Theta)} = \|f\|_{H_1}\|\Omega\|_{F^0_>(\Theta)}$. For $h \in H_n$, $n \geq 1$, we have by Fubini's theorem

$$\begin{aligned}\|a^*(f)h\|^2_{F^0_>(\Theta)} &= \int_{(\mathbb{R}_+)^n_> \times \mathbb{R}^n} |h(\boldsymbol{t},\boldsymbol{x})|^2 \left(\int_{t_1}^\infty |f(t,x)|^2 \Theta(\mathrm{d}t\mathrm{d}x)\right) \Theta^{\otimes n}(\mathrm{d}\boldsymbol{t}\mathrm{d}\boldsymbol{x}) \\ &\leq \|f\|^2_{H_1} \int_{(\mathbb{R}_+)^n_> \times \mathbb{R}^n} |h(\boldsymbol{t},\boldsymbol{x})|^2\, \Theta^{\otimes n}(\mathrm{d}\boldsymbol{t}\mathrm{d}\boldsymbol{x}) \\ &= \|f\|^2_{H_1}\|h\|^2_{F^0_>(\Theta)}.\end{aligned}$$

Therefore, by linearity we have $\|a^*(f)h\|_{F^0_>(\Theta)} \leq \|f\|_{H_1}\|h\|_{F^0_>(\Theta)}$ for all $h \in F^0_>(\Theta)$. The equality holds for $h \in H_0$ and hence $\|a^*(f)\| = \|f\|_{H_1}$.

The formula $\|a(f)\| = \|f\|_{H_1}$ can also be shown by similar estimates. This is also a consequence of the following Proposition 7.5.

For the gauge operator, we can easily show

$$\|\lambda(g)h\|_{F^0_>(\Theta)} \leq \|g\|_{L^\infty}\|h\|_{F^0_>(\Theta)}, \qquad h \in H_n,\ n \in \mathbb{N}_0,$$

and so $\|\lambda(g)\| \leq \|g\|_{L^\infty}$. It is a standard result in functional analysis that the equality holds because $\lambda(g)$ is a multiplication operator on each H_n, $n \in \mathbb{N}$. □

The previous boundedness allows us to extend these operators to bounded operators on $F_>(\Theta)$, which we still denote by the same symbols.

Proposition 7.5. *The creation operator $a^*(f)$ is the adjoint of the annihilation operator $a(f)$ for all $f \in H_1$.*

Proof. By linearity and continuity it suffices to show $\langle a^*(f)g, h\rangle = \langle g, a(f)h\rangle$ for all $g \in H_n$, $h \in H_m$, $m, n \in \mathbb{N}_0$. We may assume $m = n+1$ since otherwise these inner products are all zero. For $n \geq 1$ we have

$$\begin{aligned}\langle a^*(f)g, h\rangle &= \int_{(\mathbb{R}_+)^{n+1}_> \times \mathbb{R}^{n+1}} \overline{f(t,x)g(\boldsymbol{t},\boldsymbol{x})} h((t,\boldsymbol{t}),(x,\boldsymbol{x}))\, \Theta^{\otimes(n+1)}(\mathrm{d}t\mathrm{d}\boldsymbol{t}\mathrm{d}x\mathrm{d}\boldsymbol{x}) \\ &= \int_{(\mathbb{R}_+)^n_> \times \mathbb{R}^n} \overline{g(\boldsymbol{t},\boldsymbol{x})} \left[\int_{(t_1,\infty)\times\mathbb{R}} \overline{f(t,x)} h((t,\boldsymbol{t}),(x,\boldsymbol{x}))\, \Theta(\mathrm{d}t\mathrm{d}x)\right] \Theta^{\otimes n}(\mathrm{d}\boldsymbol{t}\mathrm{d}\boldsymbol{x}) \\ &= \langle g, a(f)h\rangle.\end{aligned}$$

For $n = 0$ we have

$$\langle a^*(f)\Omega, h\rangle = \langle f, h\rangle = \langle \Omega, a(f)h\rangle.$$ □

Proposition 7.6. *The following formulas hold for $f, k \in H_1$ and $g, h \in L^\infty(\Theta)$:*

$$a(f)\lambda(g) = a(f\overline{g}), \tag{7.1}$$

$$\lambda(g)a^*(k) = a^*(gk), \tag{7.2}$$

$$\lambda(g)\lambda(h) = \lambda(gh), \tag{7.3}$$

$$a(f)a^*(k) = \langle f, k\rangle p_\Omega + \lambda(I_\Theta(\overline{f}k)), \tag{7.4}$$

where p_Ω is the orthogonal projection onto $\mathbb{C}\Omega$ and $I_\Theta\colon L^1(\Theta) \to L^\infty(\Theta)$ is defined by

$$[I_\Theta(v)](t,x) := \int_{(t,\infty)\times\mathbb{R}} v(s,y)\, \Theta(\mathrm{d}s\mathrm{d}y),$$

which is constant on the second variable x.

Proof. These formulas can be checked by straightforward calculations. Here we only show the last formula. On $H_0 = \mathbb{C}\Omega$, we have

$$a(f)a^*(k)\Omega = a(f)k = \langle f, k\rangle\Omega = [\langle f, k\rangle p_\Omega + \lambda(I_\Theta(\overline{f}k))]\Omega,$$

and for $l \in H_n$, $n \in \mathbb{N}$, we have

$$\begin{aligned}[a(f)a^*(k)l](\boldsymbol{t},\boldsymbol{x}) &= \left[\int_{(t_1,+\infty)\times\mathbb{R}} \overline{f(t,x)}k(t,x)\, \Theta(\mathrm{d}t\mathrm{d}x)\right] l(\boldsymbol{t},\boldsymbol{x}) \\ &= [I_\Theta(\overline{f}k)](t_1)l(\boldsymbol{t},\boldsymbol{x}) \\ &= [\langle f, k\rangle p_\Omega + \lambda(I_\Theta(\overline{f}k))]l(\boldsymbol{t},\boldsymbol{x}).\end{aligned}$$ □

We will show that the family of operators $(x_t)_{t\geq 0}$ defined by

$$x_t := a(\chi_{[0,t]\times\mathbb{R}}) + a^*(\chi_{[0,t]\times\mathbb{R}}) + \lambda(\chi_{[0,t]} \otimes \mathrm{id}_\mathbb{R}) \tag{7.5}$$

is an additive monotone process that has the generator $(\dot{\rho}, \tau)$. For notational conciseness we set

$$a_{s,t} := a(\chi_{(s,t]\times\mathbb{R}}), \qquad a^*_{s,t} := a^*(\chi_{(s,t]\times\mathbb{R}}), \qquad \lambda_{s,t} := \lambda(\chi_{(s,t]} \otimes \mathrm{id}_\mathbb{R}).$$

As a first step, we calculate the resolvent of $x_t - x_s$. The following lemma is substantially based on the integral equation developed in Section 6.2.

Lemma 7.7. *Let* $(F_{s,t})_{(s,t)\in\Delta}$ *be the* $\mathscr{P}_2^0$*-REF associated with* $(\dot\rho,\tau)$*. We fix* $0\le s\le t$ *and* $z\in\mathbb{C}\setminus\mathbb{R}$ *and define a function* $\tilde F\in L^\infty(\Theta)$ *by*

$$\tilde F(r) := \begin{cases} F_{s,t}(z), & 0\le r\le s,\\ F_{r,t}(z), & s\le r\le t,\\ z, & r\ge t\end{cases}$$

and the operator

$$\tilde\lambda := \lambda(F_{s,t}(z) - \tilde F\otimes\chi_{\mathbb{R}} + \chi_{(s,t]}\otimes \mathrm{id}_{\mathbb{R}}).$$

Then we have for sufficiently large $|z|$

$$[z\mathbf{1}-(x_t-x_s)]^{-1} = [\mathbf{1}-(F_{s,t}(z)\mathbf{1}-\tilde\lambda)^{-1}a_{s,t}^*]^{-1}(F_{s,t}(z)\mathbf{1}-\tilde\lambda)^{-1}[\mathbf{1}-a_{s,t}(F_{s,t}(z)\mathbf{1}-\tilde\lambda)^{-1}]^{-1}.$$

Proof. We can directly compute the inverse operator

$$(F_{s,t}(z)\mathbf{1}-\tilde\lambda)^{-1} = \frac{1}{F_{s,t}(z)}p_\Omega + \lambda\left(\frac{1}{\tilde F\otimes\chi_{\mathbb{R}}-\chi_{(s,t]}\otimes\mathrm{id}_{\mathbb{R}}}\right). \tag{7.6}$$

Using the relations in Proposition 7.6 and $p_\Omega a_{s,t}^*=0$ we obtain

$$\begin{aligned} a_{s,t}(F_{s,t}(z)\mathbf{1}-\tilde\lambda)^{-1}a_{s,t}^* &= a_{s,t}a^*\left(\frac{\chi_{(s,t]}}{\tilde F\otimes\chi_{\mathbb{R}}-\chi_{(s,t]}\otimes\mathrm{id}_{\mathbb{R}}}\right)\\ &= \left[\int_{(s,t]\times\mathbb{R}}\frac{\Theta(\mathrm{d}r\mathrm{d}x)}{F_{r,t}(z)-x}\right]p_\Omega + \lambda\left(I_\Theta\left(\frac{\chi_{(s,t]}}{\tilde F\otimes\chi_{\mathbb{R}}-\chi_{(s,t]}\otimes\mathrm{id}_{\mathbb{R}}}\right)\right). \end{aligned}\tag{7.7}$$

This can be simplified more because, thanks to the integral equation in (6.9), it holds that

$$\int_{(s,t]\times\mathbb{R}}\frac{\Theta(\mathrm{d}r\mathrm{d}x)}{F_{r,t}(z)-x} = z - F_{s,t}(z). \tag{7.8}$$

Similarly, we have

$$I_\Theta\left(\frac{\chi_{(s,t]}}{\tilde F\otimes\chi_{\mathbb{R}}-\chi_{(s,t]}\otimes\mathrm{id}_{\mathbb{R}}}\right)(r) = z-\tilde F(r). \tag{7.9}$$

Indeed, for $r\le s$ we have

$$\begin{aligned} I_\Theta\left(\frac{\chi_{(s,t]}}{\tilde F\otimes\chi_{\mathbb{R}}-\chi_{(s,t]}\otimes\mathrm{id}_{\mathbb{R}}}\right)(r) &= \int_{(r,\infty)\times\mathbb{R}}\frac{\chi_{(s,t]}(u)}{\tilde F(u)-\chi_{(s,t]}(u)x}\,\Theta(\mathrm{d}u\mathrm{d}x)\\ &= \int_{(s,t]}\left[\int_{\mathbb{R}}\frac{1}{F_{u,t}(z)-x}\,\dot\rho(u,\mathrm{d}x)\right]\tau(\mathrm{d}u)\\ &= -F_{s,t}(z)+z = z-\tilde F(r). \end{aligned}$$

The other cases $s\le r\le t$ and $r\ge t$ can be computed analogously, and consequently, formula (7.9) holds. Combining (7.7)–(7.9) together, we obtain

$$\begin{aligned} a_{s,t}(F_{s,t}(z)\mathbf{1}-\tilde\lambda)^{-1}a_{s,t}^* &= [z-F_{s,t}(z)]p_\Omega + \lambda(z-\tilde F\otimes\chi_{\mathbb{R}})\\ &= [z-F_{s,t}(z)]\mathbf{1}+\tilde\lambda-\lambda_{s,t}. \end{aligned}\tag{7.10}$$

Formula (7.10) allows us to obtain

$$\begin{aligned} z\mathbf{1}-(x_t-x_s) &= z\mathbf{1}-a_{s,t}^*-a_{s,t}-\lambda_{s,t}\\ &= (F_{s,t}(z)\mathbf{1}-\tilde\lambda)-a_{s,t}^*-a_{s,t}+(z\mathbf{1}-F_{s,t}(z)\mathbf{1}+\tilde\lambda-\lambda_{s,t})\\ &= (F_{s,t}(z)\mathbf{1}-\tilde\lambda)-a_{s,t}^*-a_{s,t}+a_{s,t}(F_{s,t}(z)\mathbf{1}-\tilde\lambda)^{-1}a_{s,t}^*\\ &= [\mathbf{1}-a_{s,t}(F_{s,t}(z)\mathbf{1}-\tilde\lambda)^{-1}](F_{s,t}(z)\mathbf{1}-\tilde\lambda)[\mathbf{1}-(F_{s,t}(z)\mathbf{1}-\tilde\lambda)^{-1}a_{s,t}^*]. \end{aligned}$$

Taking the inverse of the above formula yields the desired formula. Note here that $F_{s,t}(z)=z+o(z)$ and the norm of $\tilde\lambda$ is uniformly bounded as $z\to\infty$, so that all the inverses exist as bounded operators. ☐

Proposition 7.8. *The distribution of* x_t-x_s *is* $\mu_{s,t}$ *for all* $0\le s\le t$*. Moreover, let* $A_{[0,s)}^\circ := \{a^*(f),a(f),\lambda(g) : f\in L^2([0,s)\times\mathbb{R},\Theta),\ g\in L^\infty([0,s)\times\mathbb{R},\Theta)\}$ *and* $n\in\mathbb{N}$*. For any* $y,y'\in A_{[0,s)}^\circ$*, we have*

$$y(x_t-x_s)^n y' = \varphi[(x_t-x_s)^n]yy', \tag{7.11}$$

$$y(x_t-x_s)^n\Omega = \varphi[(x_t-x_s)^n]y\Omega. \tag{7.12}$$

Proof. Lemma 7.7 implies for z with large $|z|$

$$[z\mathbf{1}-(x_t-x_s)]^{-1} = \sum_{j,k=0}^\infty [(F_{s,t}(z)\mathbf{1}-\tilde\lambda)^{-1}a_{s,t}^*]^j(F_{s,t}(z)\mathbf{1}-\tilde\lambda)^{-1}[a_{s,t}(F_{s,t}(z)\mathbf{1}-\tilde\lambda)^{-1}]^k, \tag{7.13}$$

and so

$$\begin{aligned}&\varphi([z\mathbf{1}-(x_t-x_s)]^{-1})\\&=\sum_{j,k=0}^{\infty}\langle[a_{s,t}(\overline{F_{s,t}(z)}\mathbf{1}-\tilde{\lambda}^*)^{-1}]^j\Omega,(F_{s,t}(z)\mathbf{1}-\tilde{\lambda})^{-1}[a_{s,t}(F_{s,t}(z)\mathbf{1}-\tilde{\lambda})^{-1}]^k\Omega\rangle.\end{aligned}$$

Here from (7.6) we deduce that

$$a_{s,t}(F_{s,t}(z)\mathbf{1}-\tilde{\lambda})^{-1}\Omega=0. \tag{7.14}$$

Therefore, only $j=k=0$ gives a nonzero contribution, i.e.,

$$\varphi([z\mathbf{1}-(x_t-x_s)]^{-1})=\langle\Omega,(F_{s,t}(z)\mathbf{1}-\tilde{\lambda})^{-1}\Omega\rangle=\frac{1}{F_{s,t}(z)}=:G_{s,t}(z), \tag{7.15}$$

showing that the analytic distribution of x_t-x_s equals $\mu_{s,t}$.

We turn to the proof of (7.11). Using (7.6) we can check

$$(F_{s,t}(z)\mathbf{1}-\tilde{\lambda})^{-1}y'=G_{s,t}(z)y'$$

for any y' that is either creation, annihilation or gauge operators supported on $[0,s)$. The point is that when multiplying a function h in H_n by y', the resulting function $y'h$ vanishes when the first variable t_1 is larger than s. Then the next operator $(F_{s,t}(z)\mathbf{1}-\tilde{\lambda})^{-1}$ acts as

$$\frac{1}{F_{s,t}(z)}p_\Omega+\lambda\left(\frac{1}{F_{s,t}(z)}\right)=G_{s,t}(z)\mathbf{1}$$

because $\chi_{(s,t]}\otimes\mathrm{id}_{\mathbb{R}}$ is always zero on the support of the function $y'h$. From a similar consideration of support, we can further obtain

$$a_{s,t}(F_{s,t}(z)\mathbf{1}-\tilde{\lambda})^{-1}y'=G_{s,t}(z)a_{s,t}y'=0. \tag{7.16}$$

Taking the adjoint and complex conjugate yields

$$y(F_{s,t}(z)\mathbf{1}-\tilde{\lambda})^{-1}a_{s,t}^*=0. \tag{7.17}$$

Combining (7.13), (7.16) and (7.17) we obtain

$$y[z\mathbf{1}-(x_t-x_s)]^{-1}y'=G_{s,t}(z)yy'.$$

Expanding the formula into Laurent series and comparing the coefficients of z^{-n-1} we obtain the desired (7.11). The last formula (7.12) can be obtained by similar calculations, combining (7.13), (7.14) and (7.17). ☐

Theorem 7.9. *The family of operators $(x_t)_{t\ge0}$ defined by (7.5) is a monotone additive process such that the analytic distributions of the increments x_t-x_s $(0\le s\le t)$ form a monotone convolution hemigroup associated with the generator $(\dot{\rho},\tau)$.*

Proof. It remains to prove the independent increment property. A fully general description requires heavy notation, so let us consider $0=t_0<t_1<t_2<t_3$, $m,n,p,q\in\mathbb{N}$ and calculate the example

$$\varphi[(x_{t_2}-x_{t_1})^m(x_{t_3}-x_{t_2})^n(x_{t_1}-x_{t_0})^p(x_{t_2}-x_{t_1})^q]. \tag{7.18}$$

Observe that $x_{t_2}-x_{t_1}$ and $x_{t_1}-x_{t_0}$ are linear combinations of elements in $A^\circ_{[0,t_2)}$ in Proposition 7.8. Then (7.11) implies

$$(x_{t_2}-x_{t_1})^m(x_{t_3}-x_{t_2})^n(x_{t_1}-x_{t_0})^p=\varphi[(x_{t_3}-x_{t_2})^n](x_{t_2}-x_{t_1})^m(x_{t_1}-x_{t_0})^p,$$

which obviously implies (7.18) equals

$$\varphi[(x_{t_3}-x_{t_2})^n]\varphi[(x_{t_2}-x_{t_1})^m(x_{t_1}-x_{t_0})^p(x_{t_2}-x_{t_1})^q]. \tag{7.19}$$

Recall here that $\varphi=\langle\Omega,\cdot\,\Omega\rangle$. We in turn use (7.12) to obtain

$$(x_{t_1}-x_{t_0})^p(x_{t_2}-x_{t_1})^q\Omega=\varphi[(x_{t_2}-x_{t_1})^q](x_{t_1}-x_{t_0})^p\Omega,$$

which implies (7.19) equals

$$\varphi[(x_{t_3}-x_{t_2})^n]\varphi[(x_{t_2}-x_{t_1})^q]\varphi[(x_{t_2}-x_{t_1})^m(x_{t_1}-x_{t_0})^p].$$

To compute the last factor $\varphi[(x_{t_2}-x_{t_1})^m(x_{t_1}-x_{t_0})^p]$, we can move the operators to the left side of the inner product as adjoints, and then we apply (7.12).

The general case can be shown analogously. For the interested reader, we note what has to be shown: Let $n\in\mathbb{N}$ and $0=t_0<t_1<\cdots<t_n$. For each interval $I(k):=(t_{k-1},t_k]$ we denote $x_{I(k)}:=x_{t_k}-x_{t_{k-1}}$ for notational conciseness. Let $j_1,j_2,\dots,j_m\in[n]$ with $j_1\ne j_2$, $j_2\ne j_3$, …, $j_{m-1}\ne j_m$ and $P_i\in\mathbb{C}_0[x]$, $i\in[m]$, be polynomials without a constant term. In the setting above, if $\ell\in[m]$ is such that $j_{\ell-1}<j_\ell>j_{\ell+1}$ then

$$\begin{aligned}&\varphi[P_1(x_{I(j_1)})P_2(x_{I(j_2)})\cdots P_m(x_{I(j_m)})]\\&=\varphi[P_\ell(x_{I(j_\ell)})]\varphi[P_1(x_{I(j_1)})\cdots P_{\ell-1}(x_{I(j_{\ell-1})})P_{\ell+1}(x_{I(j_{\ell+1})})\cdots P_m(x_{I(j_m)})].\end{aligned}$$

☐

Example 7.10. In the case $\dot{\rho}(t,\mathrm{d}x)=\delta_0(\mathrm{d}x)$ and $\tau(\mathrm{d}t)=\mathrm{d}t$, the Hilbert space H_n is isomorphic to $L^2((\mathbb{R}_+)^n_>,\mathrm{d}t_1\mathrm{d}t_2\cdots\mathrm{d}t_n)$. The function $\mathrm{id}_{\mathbb{R}}$ is zero and hence $\lambda_{s,t}=0$ and $x_t=a^*(\chi_{[0,t]})+a(\chi_{[0,t]})$. This is called a **monotone Brownian motion**. The distribution of x_t-x_s is the arcsine law $\mathrm{A}(0,t-s)$ that appeared in the monotone CLT.

The above proof has heavily relied on the resolvent of the increment x_t-x_s. We can show monotone independence for bigger subalgebras by combinatorial methods. For an interval $I\subseteq[0,+\infty)$, let A_I be the (∗-)subalgebra of $\mathbb{B}(F_>(\Theta))$ generated by the set of operators

$$A_I^\circ:=\{a^*(f),a(f),\lambda(g):f\in L^2(I\times\mathbb{R},\Theta),\ g\in L^\infty(I\times\mathbb{R},\Theta)\}.$$

Lemma 7.11. *Let $I := [s, +\infty)$ be a half-axis for some $s \geq 0$. The $*$-subalgebra $A_I + \mathbb{C}\mathbf{1}$ coincides with the linear span of the elements of the form*

$$w = a^*(f_1)a^*(f_2)\cdots a^*(f_m)[\lambda(h) + \alpha\mathbf{1}]a(g_1)a(g_2)\cdots a(g_n), \tag{7.20}$$

where $f_i, g_i \in L^2(I \times \mathbb{R}, \Theta)$, $h \in L^\infty(I \times \mathbb{R}, \Theta)$, $m, n \in \mathbb{N}_0$, $\alpha \in \mathbb{C}$.

Proof. Let B_I denote the linear span of the elements (7.20). Obviously, $B_I \subseteq A_I + \mathbb{C}\mathbf{1}$. To show the opposite inclusion, since B_I contains the generator set A_I° for A_I, it suffices to show that B_I is a $*$-subalgebra. Moreover, for this it suffices to show that $xB_I, B_I x \subseteq B_I$ for any $x \in A_I^\circ$. In addition, since B_I and A_I° are closed under $*$, only showing $xB_I \subseteq B_I$ suffices. We check this case-by-case.

The inclusion $xB_I \subseteq B_I$ is obvious for $x = a^*(f)$ and for $x = \lambda(g)$; the latter is because of (7.2) and (7.3). Therefore, it remains to check that $a(f)w \in B_I$, where w is the operator (7.20).

Case $m \geq 2$. Using Proposition 7.6 and $p_\Omega a^*(f_2) = 0$ we get

$$a(f)a^*(f_1)a^*(f_2) = [\langle f, f_1\rangle p_\Omega + \lambda(k)]a^*(f_2) = \lambda(k)a^*(f_2) = a^*(kf_2),$$

where $k := I_\Theta(\overline{f} f_1)$. Since kf_2 is supported on $I \times \mathbb{R}$, it follows that $a(f)w \in B_I$.

Case $m = 1$. Combining the decomposition $\mathbf{1} = p_\Omega + \lambda(\chi_{\mathbb{R}_+\times\mathbb{R}})$ and the previous calculations yields

$$a(f)a^*(f_1) = [\langle f, f_1\rangle p_\Omega + \lambda(k)] = \langle f, f_1\rangle \mathbf{1} + \lambda(k - \langle f, f_1\rangle \chi_{\mathbb{R}_+\times\mathbb{R}}).$$

Observe that $\tilde{k} := k - \langle f, f_1\rangle \chi_{\mathbb{R}_+\times\mathbb{R}}$ is supported on $I \times \mathbb{R}$. Setting $\beta := \langle f, f_1\rangle$, the first three letters of the word $a(f)w$ equals

$$a(f)a^*(f_1)[\lambda(h) + \alpha\mathbf{1}] = [\lambda(\tilde{k}) + \beta\mathbf{1}][\lambda(h) + \alpha\mathbf{1}] = \lambda(\tilde{k}h + \alpha\tilde{k} + \beta h) + \alpha\beta\mathbf{1},$$

so that $a(f)w \in B_I$.

Case $m = 0$. From Proposition 7.6, the first two letters of the word $a(f)w$ are computed as

$$a(f)[\lambda(h) + \alpha\mathbf{1}] = a(f\overline{h} + \overline{\alpha} f),$$

so that $a(f)w \in B_I$. □

Proposition 7.12. *Let $s > 0$. For any $x, x' \in A_{[0,s)}$ and $y \in A_{[s,+\infty)} + \mathbb{C}\mathbf{1}$, we have*

$$xyx' = \varphi(y)xx', \tag{7.21}$$

$$xy\Omega = \varphi(y)x\Omega. \tag{7.22}$$

Proof. The second formula follows from the first one because (7.4) shows $p_\Omega \in A_{[0,s)}$ unless $L^2([0, s) \times \mathbb{R}, \Theta) = \{0\}$, in which case the second formula is obvious. It suffices to show (7.21) for $y = w$ in the form (7.20) and $x, x' \in A_{[0,s)}^\circ$.

Case $m \geq 1$. Then $y\Omega = 0$ and hence $\varphi(y) = 0$. Since x is one of $a^*(f), a(f), \lambda(g)$ where f, g are supported on $[0, s) \times \mathbb{R}$ and f_1 is supported on $[s, +\infty) \times \mathbb{R}$, we can conclude $xa^*(f_1) = 0$, so that $xy = 0$.

Case $m = 0$ and $n \geq 1$. We have $y\Omega = 0$ and hence $\varphi(y) = 0$. On the other hand, the argument symmetric with the previous case $m \geq 1$ shows $xyx' = 0$.

Case $m = 0$ and $n = 0$. Then $y = \lambda(h) + \alpha\mathbf{1}$ and $\varphi(y) = \alpha$. One can show that $x\lambda(h)x' = 0$ from Proposition 7.6 and the fact that the support of h is contained in $[s, +\infty) \times \mathbb{R}$. Thus we obtain $xyx' = \alpha xx' = \varphi(y)xx'$. □

Theorem 7.13. *Let $n \in \mathbb{N}$ and $0 = t_0 < t_1 < \cdots < t_n$. The $*$-subalgebras*

$$A_{[t_0,t_1)},\ A_{[t_1,t_2)},\ \ldots\ ,\ A_{[t_{n-1},t_n)}$$

are monotonically independent.

Proof. Almost the same as the proof of Theorem 7.9. □

Note that this theorem implies the monotone independence of the increments of (x_t), which was already proved in Theorem 7.9.

7.2. **A construction from Markov processes.** The second construction of monotone additive processes is based on classical Markov processes. Basic facts on Markov processes are reviewed briefly here. For further information on Markov processes, the reader is referred to e.g., [96, 134].

Recall from Lemma 5.9 that the composition of two probability kernels k, l on $\mathbb{R}$

$$(kl)(x, B) := \int_{\mathbb{R}} k(x, \mathrm{d}y)l(y, B), \qquad x \in \mathbb{R},\ B \in \mathcal{B}(\mathbb{R}),$$

is also a probability kernel.

Definition 7.14. A family $(k_{s,t})_{(s,t)\in\triangle}$ of probability kernels on $\mathbb{R}$ is called **transition kernels** if $k_{s,s}(x, \cdot) = \delta_x$ for all $x \in \mathbb{R}$, $s \geq 0$ and the following **Chapman–Kolmogorov equation** holds:

$$k_{s,t}k_{t,u} = k_{s,u}, \qquad 0 \leq s \leq t \leq u. \tag{7.23}$$

Definition 7.15. Let $(k_{s,t})_{(s,t)\in\triangle}$ be transition kernels on $\mathbb{R}$. A family of $\mathbb{R}$-valued measurable functions $(X_t)_{t\geq 0}$ on a measurable space $(\Omega, \mathcal{F})$, together with a family of sub-σ-fields $\mathcal{F}_I \subseteq \mathcal{F}$ indexed by the closed intervals I of $[0, +\infty)$ and a family of probability measures $\mathbb{P}^{(s,x)}$ on $(\Omega, \mathcal{F}_{[s,\infty)})$ ($s \geq 0$, $x \in \mathbb{R}$), is called a **Markov process having the transition kernels** $(k_{s,t})_{(s,t)\in\triangle}$ if the following conditions hold:

- $\mathcal{F}_I \subseteq \mathcal{F}_J$ whenever $I \subseteq J$;
- for every bounded Borel measurable function $f\colon \mathbb{R} \to \mathbb{R}$, $0 \le s \le t \le u$ and $x \in \mathbb{R}$,

$$\mathbb{E}^{(s,x)}[f(X_u)|\mathcal{F}_{[s,t]}] = \int_{\mathbb{R}} f(y) k_{t,u}(X_t, \mathrm{d}y), \qquad \mathbb{P}^{(s,x)}\text{-a.s.}; \tag{7.24}$$

- $\mathbb{P}^{(s,x)}(X_s = x) = 1$ for all $s \ge 0$ and $x \in \mathbb{R}$.

The following is a rather standard result.

Proposition 7.16. *Let $(k_{s,t})_{(s,t)\in\triangle}$ be transition kernels on $\mathbb{R}$. Then there exists a Markov process that has the transition kernels $(k_{s,t})_{(s,t)\in\triangle}$.*

Proof. A standard construction is called the coordinate process that we present below. Let $\Omega = \mathbb{R}^{[0,+\infty)}$ be the set of all functions $\omega\colon [0,+\infty) \to \mathbb{R}$. Let $\mathcal{C} \subseteq 2^\Omega$ be the set of the cylinder sets $\{\omega \in \Omega : \omega(t_1) \in A_1,\ \omega(t_2) \in A_2,\ \dots,\ \omega(t_n) \in A_n\}$, $0 \le t_1 < t_2 < \cdots < t_n$, $A_1, A_2, \dots, A_n \in \mathcal{B}(\mathbb{R})$, $n \in \mathbb{N}$. Let $\mathcal{F} \subseteq 2^\Omega$ be the σ-field generated by $\mathcal{C}$. The coordinate process $X_t\colon \Omega \to \mathbb{R}$ is defined by $X_t(\omega) := \omega(t)$. The sub-σ-field $\mathcal{F}_I$ is the σ-field generated by $\{X_t^{-1}(A) : t \in I,\ A \in \mathcal{B}(\mathbb{R})\}$.

We fix (s,x) and consider the family of probability measures $\mu^{(s,x)}_{t_1,t_2,\dots,t_n}$ on $\mathbb{R}^n$, indexed by $s \le t_1 < t_2 < \cdots < t_n$ $(n \in \mathbb{N})$, defined by the iterated integrals

$$\mu^{(s,x)}_{t_1,t_2,\dots,t_n}(A) := \int_{\mathbb{R}} k_{s,t_1}(x, \mathrm{d}x_1) \int_{\mathbb{R}} k_{t_1,t_2}(x_1, \mathrm{d}x_2) \cdots \int_{\mathbb{R}} k_{t_{n-1},t_n}(x_{n-1}, \mathrm{d}x_n) \chi_A(x_1, x_2, \dots, x_n).$$

By the Chapman–Kolmogorov equation, these probability measures satisfy the consistency

$$\begin{aligned}&\mu^{(s,x)}_{t_1,t_2,\dots,t_n}(A_1 \times \cdots \times A_{i-1} \times \mathbb{R} \times A_{i+1} \times \cdots \times A_n)\\ &\qquad = \mu^{(s,x)}_{t_1,\dots,t_{i-1},t_{i+1},\dots,t_n}(A_1 \times \cdots \times A_{i-1} \times A_{i+1} \times \cdots \times A_n), \qquad 1 \le i \le n,\end{aligned}$$

and so Kolmogorov's extension theorem (see, e.g. [59]) guarantees the existence of a probability measure $\mathbb{P}^{(s,x)}$ on $\mathcal{F}_{[s,\infty)}$ such that

$$\mathbb{P}^{(s,x)}(X_{t_1} \in A_1,\ X_{t_2} \in A_2,\ \dots,\ X_{t_n} \in A_n) = \mu^{(s,x)}_{t_1,t_2,\dots,t_n}(A_1 \times A_2 \times \cdots \times A_n)$$

for all $s \le t_1 < t_2 < \cdots < t_n$ and $A_1, A_2, \dots, A_n \in \mathcal{B}(\mathbb{R})$.

All the requirements for a Markov process are obvious except (7.24). To show this, by the well known characterization of the conditional expectation, it suffices to show

$$\mathbb{E}^{(s,x)}[f(X_u)\chi_F] = \mathbb{E}^{(s,x)}\left[\int_{\mathbb{R}} f(y) k_{t,u}(X_t, \mathrm{d}y)\chi_F\right], \qquad F \in \mathcal{F}_{[s,t]}. \tag{7.25}$$

It further suffices to show that the above holds for all F in the class

$$\mathcal{G} := \left\{\bigcap_{i=1}^n X_{r_i}^{-1}(A_i) : n \in \mathbb{N},\ s \le r_i \le t,\ A_i \in \mathcal{B}(\mathbb{R})\right\}.$$

Indeed, as soon as the set $\mathcal{F}' := \{F \in \mathcal{F}_{[s,t]} : (7.25) \text{ holds}\}$ contains $\mathcal{G}$, since $\mathcal{G}$ is a π-system and $\mathcal{F}'$ is a λ-system, by the π-λ theorem (Theorem 4.2), $\mathcal{F}'$ contains $\sigma(\mathcal{G}) = \mathcal{F}_{[s,t]}$.

To finish the proof, let $F = \bigcap_{i=1}^n X_{r_i}^{-1}(A_i) \in \mathcal{G}$ with $s \le r_1 < r_2 < \cdots < r_n \le t$. To avoid heavy notation we only consider $n = 2$; the general case is similar. We then have

$$\begin{aligned}&\mathbb{E}^{(s,x)}\left[\int_{\mathbb{R}} f(y) k_{t,u}(X_t, \mathrm{d}y)\chi_F\right]\\
&= \mathbb{E}^{(s,x)}\left[\int_{\mathbb{R}} f(y) k_{t,u}(X_t, \mathrm{d}y)\chi_{A_1}(X_{r_1})\chi_{A_2}(X_{r_2})\right]\\
&= \int_{\mathbb{R}^2} \mu^{(s,x)}_{r_1,r_2,t}(\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3) \int_{\mathbb{R}} k_{t,u}(x_3, \mathrm{d}y)\chi_{A_1}(x_1)\chi_{A_2}(x_2) f(y)\\
&= \int_{\mathbb{R}} k_{s,r_1}(x, \mathrm{d}x_1) \int_{\mathbb{R}} k_{r_1,r_2}(x_1, \mathrm{d}x_2) \int_{\mathbb{R}} k_{r_2,t}(x_2, \mathrm{d}x_3) \int_{\mathbb{R}} k_{t,u}(x_3, \mathrm{d}y)\chi_{A_1}(x_1)\chi_{A_2}(x_2) f(y) \qquad (7.26)\\
&= \int_{\mathbb{R}} k_{s,r_1}(x, \mathrm{d}x_1) \int_{\mathbb{R}} k_{r_1,r_2}(x_1, \mathrm{d}x_2) \int_{\mathbb{R}} k_{r_2,u}(x_2, \mathrm{d}y)\chi_{A_1}(x_1)\chi_{A_2}(x_2) f(y) \qquad (7.27)\\
&= \mathbb{E}^{(s,x)}[\chi_{A_1}(X_{r_1})\chi_{A_2}(X_{r_2}) f(X_u)] = \mathbb{E}^{(s,x)}[\chi_F f(X_u)],\end{aligned}$$

where the Chapman–Kolmogorov equation was used from (7.26) to (7.27). □

Remark 7.17. The above construction of a Markov process does not tell us how the sample path $t \mapsto X_t(\omega)$ behaves for each $\omega \in \Omega$. To have samples paths with some kind of continuity, one has to "modify" the above Markov process. For our purpose, sample paths do not matter and the above construction is enough.

We introduce a suitable Markov process for constructing a monotone additive process. Let $(\mu_{s,t})_{(s,t)\in\triangle}$ be a monotone convolution hemigroup. Let $k_{s,t}(x,\cdot) := (\delta_x \rhd \mu_{s,t})(\cdot)$. Recall from Proposition 5.11 that $k_{s,t}$ is a probability kernel and that for every probability measure μ we have

$$(\mu \rhd \mu_{t,u})(\cdot) = \int_{\mathbb{R}} \mu(\mathrm{d}y) k_{t,u}(y, \cdot).$$

Selecting $\mu = \delta_x \rhd \mu_{s,t} = k_{s,t}(x,\cdot)$ yields

$$(\delta_x \rhd \mu_{s,u})(\cdot) = \int_{\mathbb{R}} k_{s,t}(x,\mathrm{d}y)k_{t,u}(y,\cdot),$$

which reads $k_{s,u} = k_{s,t}k_{t,u}$, i.e., the Chapman–Kolmogorov equation. By Proposition 7.16 there exists a Markov process $(X_t)_{t\geq 0}$ that has the constructed transition kernels $(k_{s,t})$. We then set $H_\rhd := L^2(\Omega, \mathcal{F}, \mathbb{P}^{(0,0)})$ and work on the C^*-probability space $(\mathbb{B}(H_\rhd), \varphi)$, where $\varphi(a) := \langle \chi_\Omega, a\chi_\Omega\rangle$. The identity operator on $H_\rhd$ is denoted as $\mathbf{1}$. For notational simplicity we denote $\mathbb{P} := \mathbb{P}^{(0,0)}$, $\mathbb{E} := \mathbb{E}^{(0,0)}$ and $\mathcal{F}_t := \mathcal{F}_{[0,t]}$. Also for analytic transforms, we set the shorthand symbols $F_{s,t}(z) := F_{\mu_{s,t}}(z)$ and $G_{s,t}(z) := G_{\mu_{s,t}}(z)$.

For the sake of simplicity, we assume that each $\mu_{s,t}$ has compact support. Let $p_t \in \mathbb{B}(H_\rhd)$ be the conditional expectation $p_t Z := \mathbb{E}[Z|\mathcal{F}_t]$, $Z \in H_\rhd$. It is known that conditional expectations onto sub-σ-fields are orthogonal projections on the L^2 space, so each p_t is an orthogonal projection. The multiplication operator on $H_\rhd$ by the random variable X_t is denoted by m_t, i.e., $m_t(Z) := X_t Z$. Since $\mu_{0,t}$ has compact support, $X_t \in L^\infty$ and so the operator m_t is bounded.

We introduce a family of bounded self-adjoint operators on $H_\rhd$

$$y_t := p_t m_t, \qquad t \geq 0. \tag{7.28}$$

Note that $p_t m_t = m_t p_t$ because

$$p_t m_t(Z) = \mathbb{E}[X_t Z|\mathcal{F}_t] = X_t\mathbb{E}[Z|\mathcal{F}_t] = m_t p_t(Z), \qquad Z \in H_\rhd.$$

We will show that $(y_t)_{t\geq 0}$ is a monotone additive process. Again we analyze the resolvents of the increments $y_t - y_s$.

Lemma 7.18. *For $z \in \mathbb{C}^+$ and $0 \leq s \leq t$ it holds that*

$$p_s(z\mathbf{1} - m_t)^{-1}p_s = (F_{s,t}(z)\mathbf{1} - m_s)^{-1}p_s.$$

Proof. For $Z \in H_\rhd$, keeping in mind that $p_s(Z)$ is $\mathcal{F}_s$-measurable, we have

$$\begin{aligned}
p_s(z\mathbf{1} - m_t)^{-1}p_s(Z) &= \mathbb{E}\left[\frac{1}{z - X_t}p_s(Z)\middle|\mathcal{F}_s\right] = \mathbb{E}\left[\frac{1}{z-X_t}\middle|\mathcal{F}_s\right]p_s(Z)\\
&= \int_{\mathbb{R}} \frac{1}{z-y}k_{s,t}(X_s,\mathrm{d}y)p_s(Z) = G_{k_{s,t}(X_s,\cdot)}(z)p_s(Z)\\
&= \frac{1}{F_{s,t}(z) - X_s}p_s(Z) = (F_{s,t}(z)\mathbf{1} - m_s)^{-1}p_s(Z).
\end{aligned}$$

□

Proposition 7.19. *For $z \in \mathbb{C}^+$ and $0 \leq s \leq t$ it holds that*

$$p_s[z\mathbf{1} - (y_t - y_s)]^{-1}p_s = G_{s,t}(z)p_s. \tag{7.29}$$

In particular, the distribution of $y_t - y_s$ with respect to $\varphi(\cdot) = \langle\chi_\Omega, \cdot\,\chi_\Omega\rangle_{H_\rhd}$ equals $\mu_{s,t}$, and

$$p_s(y_t - y_s)^n p_s = \varphi[(y_t - y_s)^n]p_s, \qquad n \in \mathbb{N}. \tag{7.30}$$

Proof. By analytic continuation, it suffices to show the formula for $z \in \mathbb{C}^+$ with large $|z|$. We first observe

$$p_s[z\mathbf{1} - (y_t - y_s)]^{-1}p_s = p_s[z\mathbf{1} - (m_t - m_s p_s)]^{-1}p_s. \tag{7.31}$$

Indeed, by series expansion

$$p_s[z\mathbf{1} - (y_t - y_s)]^{-1}p_s = \sum_{n\geq 0}\frac{p_s(m_t p_t - m_s p_s)^n p_s}{z^{n+1}}.$$

In the expansion of $(m_t p_t - m_s p_s)^n p_s Z$ ($Z \in H_\rhd$), the operator p_t always acts on $\mathcal{F}_t$-measurable random variables, so that p_t can be replaced with the identity operator. This verifies (7.31).

We continue the calculation (7.31) as follows:

$$\begin{aligned}
p_s[z\mathbf{1} - (m_t - m_s p_s)]^{-1}p_s &= p_s[\mathbf{1} + (z\mathbf{1} - m_t)^{-1}m_s p_s]^{-1}(z\mathbf{1} - m_t)^{-1}p_s\\
&= \sum_{n\geq 0}(-1)^n p_s[(z\mathbf{1} - m_t)^{-1}m_s p_s]^n(z\mathbf{1} - m_t)^{-1}p_s\\
&= \sum_{n\geq 0}(-1)^n p_s m_s^n[p_s(z\mathbf{1} - m_t)^{-1}p_s]^{n+1} \qquad (7.32)\\
&= \sum_{n\geq 0}p_s(-m_s)^n[(F_{s,t}(z)\mathbf{1} - m_s)^{-1}p_s]^{n+1} \qquad (7.33)\\
&= \sum_{n\geq 0}(-m_s)^n(F_{s,t}(z)\mathbf{1} - m_s)^{-n-1}p_s\\
&= [F_{s,t}(z)\mathbf{1} - m_s - (-m_s)]^{-1}p_s\\
&= G_{s,t}(z)p_s.
\end{aligned}$$

Here, (7.32) holds because m_s commutes with both p_s and $(z\mathbf{1} - m_t)^{-1}$, and (7.33) is obtained from Lemma 7.18.

For the last statement, observe that

$$\begin{aligned}
G_{\mu_{y_t - y_s}}(z) &= \varphi([z\mathbf{1} - (y_t - y_s)]^{-1}) = \varphi(p_s[z\mathbf{1} - (y_t - y_s)]^{-1}p_s)\\
&= \varphi(G_{s,t}(z)p_s) = G_{s,t}(z),
\end{aligned}$$

so that $\mu_{y_t - y_s} = \mu_{s,t}$. The relation (7.30) follows by comparing the coefficients of z^{-n-1} in the series expansion of (7.29). □

Theorem 7.20. *The process $(y_t)_{t\geq 0}$ defined in* (7.28) *is a monotone additive process such that each increment $y_t - y_s$ has the given distribution $\mu_{s,t}$.*

Proof. Let us consider $0 = t_0 < t_1 < t_2 < t_3$, $m, n, p, q, r \in \mathbb{N}$ and calculate the example

$$\varphi[(y_{t_2} - y_{t_1})^m (y_{t_3} - y_{t_2})^n (y_{t_1} - y_{t_0})^p (y_{t_2} - y_{t_1})^q (y_{t_1} - y_{t_0})^r]. \tag{7.34}$$

Observe that $(y_{t_2} - y_{t_1})^m p_{t_2} = (y_{t_2} - y_{t_1})^m$ and $p_{t_2}(y_{t_1} - y_{t_0})^p = (y_{t_1} - y_{t_0})^p$ because of the tower property of conditional expectations $p_{t_2} p_{t_i} = p_{t_i}$ for $i = 0, 1, 2$. Hence, we are allowed to replace the factor $(y_{t_3} - y_{t_2})^n$ with $p_{t_2}(y_{t_3} - y_{t_2})^n p_{t_2}$, which is equal to $\varphi[(y_{t_3} - y_{t_2})^n] p_{t_2}$ by Proposition 7.19. This implies that (7.34) equals

$$\varphi[(y_{t_3} - y_{t_2})^n] \varphi[(y_{t_2} - y_{t_1})^m (y_{t_1} - y_{t_0})^p (y_{t_2} - y_{t_1})^q (y_{t_1} - y_{t_0})^r].$$

The general case can be shown analogously. □

Remark 7.21. The above construction of monotone additive processes is independent of the integral or integro-differential equation developed in Section 6.2. In the case of monotone convolution semigroup, i.e., $\mu_{s,t} = \mu_{0,t-s}$ for all $0 \leq s \leq t$, the Markov process $(X_t)_{t\geq 0}$ is a Feller process and its generator can be expressed in terms of the parameter (γ, σ) in (5.16); see [70] for further details.

7.3. **Notes.** The study of Fock spaces in noncommutative probability can be traced back to Boson and Fermion Fock spaces in quantum physics. Hudson and Parthasarathy developed a quantum version of Itô calculus on the Boson Fock space [91, 129]. In free probability, the corresponding space is the full (or free) Fock space [149], on which free stochastic calculus was initiated by Kümmerer and Speicher [100]. The Boson, Fermion and full Fock spaces are interpolated by the q-Fock space of Bożejko and Speicher [38]. Fock spaces have provided a canonical construction of independent random variables and continuous-time processes of independent increments. In particular, the q-Fock space and its relatives have offered remarkable von Neumann algebras that are still actively studied [99, 115]. Concerning monotone probability, there seems to be no notable von Neumann algebras building upon monotonically independent random variables so far. On the other hand, C^*-algebras related to the monotone Fock space have been investigated in the literature, see e.g. [53, 54, 55].

The construction of additive processes on the monotone Fock space followed Jekel [92], who developed Loewner theory and monotone convolution hemigroups in a more general operator-valued setting, where φ is an algebra-valued functional called a conditional expectation. Here we have presented the results in the simplified setting of $\mathbb{C}$-valued functional. Our results also contain advancements because we only assume the continuity (not absolute continuity) about the time parameter. The special case of monotone Brownian motion $x_t = a^*(\chi_{[0,t]}) + a(\chi_{[0,t]})$ first appeared in Lu [108] and Muraki [119]. Muraki considered a discrete monotone Fock space in [118, 121]. In free probability, a parallel construction of free additive processes on the full Fock space is known, see [125, Exercise 13.19] that assumes the stationarity of the distributions. The term "gauge operator" appears in [125] for an analogous operator on the full Fock space; note that the terms "gauge process" and "conservation process" are used in [91] and [129] respectively for the operator on the Boson Fock space corresponding to our $\lambda(\chi_{[0,t]} \otimes \mathrm{id}_{\mathbb{R}})$. The operator $\lambda(g)$ is called the multiplication operator in [92]. Concerning notation, it is common to denote $\ell(f)$ for the creation operator and $\ell^*(f)$ for the annihilation operator on the full Fock space [125, 149]. On the other hand, the symbol $a^*(f)$ or $a^\dagger(f)$ is commonly used for the creation operator on the Boson Fock space and $a(f)$ for the annihilation, which we have followed.

The construction of monotone additive processes from the Markov processes $(X_t)_{t\geq 0}$ is due to [70]. The book [138] by Schleissinger also discusses the construction. The original paper treated more general unbounded operator processes. Our construction is limited to the bounded case, which greatly simplifies the proofs and formulations; already the definition of monotone additive processes is more involved in the unbounded operator setting. The Markov process $(X_t)_{t\geq 0}$ was first considered by Biane [35] in connection to free additive processes and subordination functions.

There are other constructions of monotone additive processes as solutions to quantum stochastic differential equations, see [25, 67, 70]. Hamdi constructed a multiplicative monotone unitary Brownian motion as a solution to a quantum SDE [78].

Classical stochastic processes related to monotone independence are studied in the literature: a discrete-time analogue of the Markov process (X_t), i.e., Markov chains, can be similarly defined and is studied by Letac and Malouche [107], Wang and Wendler [154]; Biane mentions that the Markov process $(X_t)_{t\geq 0}$ associated with $F_t(z) = \sqrt{z^2 - 2t}$ (the reciprocal Cauchy transform of the arcsine law $\mathrm{A}(0,t)$) is the Azéma martingale [35]; Belton studied a semimartingale having the monotone Poisson distribution [26, 27].

8. Monotone independence in random matrix theory

So far we have studied monotone independence from the viewpoint of analogy to probability theory. Here we discuss different aspects of monotone independence in the context of large random matrices.

For a square matrix X_N of size N, recall from Example 1.6 (a) that the **empirical eigenvalue distribution** is the probability measure

$$\mu_{X_N} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(X_N)},$$

where $\lambda_i(X_N)$, $i = 1, 2, \ldots, N$, are the eigenvalues of X_N counting multiplicities. For many random matrix models X_N, the empirical eigenvalue distributions converge weakly to a nonrandom compactly supported probability measure μ on $\mathbb{C}$ as $N \to \infty$. In addition, the largest modulus of eigenvalues often converge to the largest modulus of the support of μ. For example, the **normalized Gaussian Unitary Ensemble (GUE)** is a random matrix $G_N = (g(i,j))_{i,j\in[N]}$ that satisfies the following conditions:

- $g(i,j) = \overline{g(j,i)}$ for all $i, j \in [N]$, i.e., G_N is Hermitian;
- the random variables $\{\Re g(i,j), \Im g(i,j), g(i,i) : i, j \in [N],\ i > j\}$ are independent;

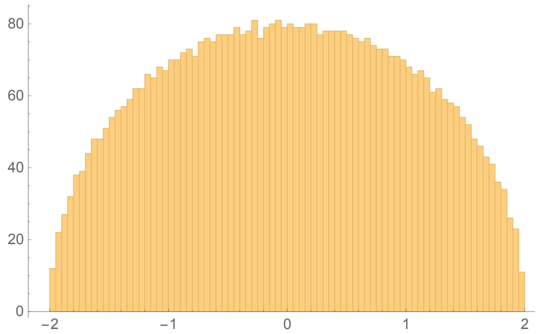

FIGURE 6. A histogram for the eigenvalues of G_N with $N = 5000$ and bin size 0.05. One can observe that the graph of the histogram looks like the probability density function of the semicircle distribution. In fact, one can show that when the bin size $\delta_N \to 0^+$ $(N \to \infty)$ is appropriately selected, then the graph of the normalized histogram (so that the area equals one) converges to $(2\pi)^{-1}\sqrt{4-x^2}$. A remarkable sufficient condition $N^{-2/3}\log N \ll \delta_N \ll 1$ is given in [61, Corollary 4.2].

- $\Re g(i,j)$ and $\Im X(i,j)$ are distributed as $\mathrm{N}(0,1/(2N))$ if $i \neq j$, and $g(i,i)$ is distributed as $\mathrm{N}(0,1/N)$.

It is known that the empirical eigenvalue distribution of G_N converges weakly to the semicircle distribution $\mathrm{S}(0,1)$ a.s., see e.g. [148, Theorem 2.4.2]. Moreover, the largest eigenvalue converges to 2 a.s., see e.g. [148, Section 2.3]. A simulation is shown in Figure 6.

In general, the weak convergence of the empirical eigenvalue distributions does not guarantee that the largest modulus of eigenvalues converges to the largest modulus of the support of the limiting measure. This is because, in the convergence of empirical eigenvalue distributions, a relatively small number of eigenvalues, e.g. of order $o(N)$, will disappear in the large N limit. For example, suppose that X_N, $N \in \mathbb{N}$, are Hermitian matrices and μ_{X_N} converges weakly to a probability measure μ. Then for any sequence of subsets $I_N \subseteq [N]$, $N \in \mathbb{N}$, with $\#I_N = o(N)$, and a bounded continuous function f on $\mathbb{R}$, we have

$$\int_{\mathbb{R}} f(x)\,\mu_{X_N}(\mathrm{d}x) = \frac{1}{N}\sum_{i\in I_N} f(\lambda_i(X_N)) + \frac{1}{N}\sum_{i\in[N]\setminus I_N} f(\lambda_i(X_N)).$$

The first sum goes to zero as $N \to \infty$ since

$$\left|\frac{1}{N}\sum_{i\in I_N} f(\lambda_i(X_N))\right| \le \|f\|_{L^\infty}\frac{\#I_N}{N} \to 0,$$

and therefore, the eigenvalues $\lambda_i(X_N)$, $i \in I_N$, do not contribute to the limit μ.

Eigenvalues located outside the support of the limiting measure in the large N limit are called **outliers**. As an application of monotone independence, we will analyze some random matrix models that have outliers in the large N limit.

8.1. **Weingarten calculus on the unitary group.** Let U be a **Haar unitary random matrix** of size N, i.e., it is a random variable taking values in the group U_N of unitary matrices of size N such that the distribution on U_N induced by U is the normalized Haar measure. We use several known results on expectations of moments of entries of U. Let $\mathfrak{S}_k$ denote the symmetric group on $[k]$ with unit denoted as e_k. For each $\sigma \in \mathfrak{S}_k$ and matrices $A_1, A_2, \dots, A_k \in \mathrm{M}_N(\mathbb{C})$, let $\mathrm{Tr}_\sigma(A_1, A_2, \dots, A_k)$ denote the product of traces according to the cycle decomposition of σ: if $\sigma = c_1c_2\dots c_\ell$ where $c_i = (k_i(1), k_i(2), \dots, k_i(p_i))$ are cyclic permutations, then

$$\mathrm{Tr}_\sigma(A_1, A_2, \dots, A_k) := \prod_{i=1}^{\ell} \mathrm{Tr}(A_{k_i(1)}A_{k_i(2)}\cdots A_{k_i(p_i)}).$$

The number ℓ is determined uniquely by σ and is denoted $\ell(\sigma)$. As an example, for $\sigma = (1,4,6)(2)(3,5)$ the above definition reads

$$\mathrm{Tr}_\sigma(A_1, A_2, \dots, A_6) = \mathrm{Tr}(A_1A_4A_6)\mathrm{Tr}(A_2)\mathrm{Tr}(A_3A_5).$$

There exists a function $\mathrm{Wg}\colon \left(\bigcup_{k\in\mathbb{N}}\mathfrak{S}_k\right)\times\mathbb{N} \to \mathbb{R}$, called the **Weingarten function**, such that

$$\begin{aligned}&\mathbb{E}[\mathrm{Tr}_\sigma(A_1UB_1U^*, A_3UB_2U^*, \dots, A_kUB_kU^*)] \\ &= \sum_{\substack{\sigma_1,\sigma_2,\sigma_3\in\mathfrak{S}_k\\ \sigma_1\sigma_2\sigma_3=\sigma}} \mathrm{Tr}_{\sigma_1}(A_1, A_2, \dots, A_k)\mathrm{Tr}_{\sigma_2}(B_1, B_2, \dots, B_k)\mathrm{Wg}(\sigma_3, N)\end{aligned} \tag{8.1}$$

for any $k, N \in \mathbb{N}$ and nonrandom square matrices $A_i, B_i \in \mathrm{M}_N(\mathbb{C})$, see [49].

Let C_p be the Catalan number

$$\mathrm{C}_p := \frac{(2p)!}{p!(p+1)!}, \qquad p \in \mathbb{N}.$$

For $\sigma \in \mathfrak{S}_k$ let $|\sigma|$ be the minimal number of transpositions such that σ can be expressed as the product of them. The relation $\ell(\sigma) = k - |\sigma|$ holds true. Let $\sigma = c_1c_2\cdots c_{\ell(\sigma)}$ be the cycle decomposition of $\sigma \in \mathfrak{S}_k$ and then let

$$\mathrm{M}(\sigma) := \prod_{i=1}^{\ell(\sigma)} (-1)^{|c_i|}\mathrm{C}_{|c_i|}. \tag{8.2}$$

The Weingarten function satisfies

$$\mathrm{Wg}(\sigma, N) = N^{-k-|\sigma|}\big[\mathrm{M}(\sigma) + O(N^{-2})\big], \qquad \sigma \in \mathfrak{S}_k, \tag{8.3}$$

see [49, Corollary 2.7].

8.2. **Asymptotic monotone independence of large random matrices.** Since Voiculescu's pioneering work on asymptotic free independence of large random matrices [151], the method of noncommutative probability has been applied to a wide range of theoretical and practical problems on random matrices. One of results in this direction is asymptotic monotone independence for large random matrices.

In this section, we use **noncommutative $*$-polynomials** $\boldsymbol{P}(x_i : i \in I)$, e.g., $\boldsymbol{P}(x_1, x_2) = -\mathbf{1} + x_2^* + x_1x_2 + 2x_2x_1^*$. Formally, these are defined as elements of the unital free associative algebra over $\mathbb{C}$ on a set $\{x_i : i \in I\} \sqcup \{x_i^* : i \in I\}$. One can substitute any family $(a_i)_{i\in I}$ in a unital $*$-algebra A into the noncommutative $*$-polynomial; in the above example, one has $\boldsymbol{P}(a_1, a_2) = -\mathbf{1}_A + a_2^* + a_1a_2 + 2a_2a_1^* \in A$.

A noncommutative $*$-polynomial whose coefficient of the unit is zero is called a **noncommutative $*$-polynomial without constant term**. For any subset $S = \{a_i : i \in I\}$ of a $*$-algebra A, the $*$-subalgebra $\langle S, S^*\rangle$ generated by S is exactly the set of all $\boldsymbol{P}(a_i : i \in I)$'s where $\boldsymbol{P}(x_i : i \in I)$ runs over the noncommutative $*$-polynomials without constant term.

Proposition 8.1. *Let $U(N)$ be an $N \times N$ Haar unitary random matrix and $(A(i,N) : i \in I)$, $(B(j,N) : j \in J)$ be families of $N \times N$ nonrandom matrices for $N \in \mathbb{N}$. Suppose that the limits*

$$\lim_{N\to\infty} \mathrm{Tr}(A(i_1,N)^{\varepsilon_1}A(i_2,N)^{\varepsilon_2}\cdots A(i_k,N)^{\varepsilon_k}) \in \mathbb{C}, \tag{8.4}$$

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{Tr}(B(j_1,N)^{\varepsilon_1}B(j_2,N)^{\varepsilon_2}\cdots B(j_k,N)^{\varepsilon_k}) \in \mathbb{C} \tag{8.5}$$

exist for any $k \in \mathbb{N}$, $i_1, i_2, \dots, i_k \in I$, $j_1, j_2, \dots, j_k \in J$ and $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k \in \{1, *\}$.

Let $\boldsymbol{P}_r(x_i : i \in I)$, $r \in [k]$, be noncommutative $*$-polynomials without constant terms and $\boldsymbol{Q}_s(y_j : j \in J)$, $s \in [k]$, be noncommutative $*$-polynomials. Then the matrices $A_r(N) := \boldsymbol{P}_r(A(i,N) : i \in I)$ and $B_s(N) := \boldsymbol{Q}_s(U(N)B(j,N)U(N)^* : j \in J)$ satisfy the estimates

$$\mathbb{E}[\mathrm{Tr}(A_1B_1A_2B_2\cdots A_kB_k)] = \mathrm{Tr}(A_1A_2\cdots A_k)\prod_{p=1}^{k}\left[\frac{1}{N}\mathrm{Tr}(B_p)\right] + O(N^{-1}), \tag{8.6}$$

$$\mathbb{E}\left[\Big|\mathrm{Tr}(A_1B_1A_2B_2\cdots A_kB_k) - \mathbb{E}[\mathrm{Tr}(A_1B_1A_2B_2\cdots A_kB_k)]\Big|^4\right] = O(N^{-2}) \tag{8.7}$$

with shorthand notation $A_r = A_r(N)$ and $B_s = B_s(N)$.

Proof. Let γ_k be the circular permutation $\gamma_k = (1, 2, \dots, k) \in \mathfrak{S}_k$ and $\mathfrak{S}_S$ be the permutation group on a set S.

Proof of (8.6). The left-hand side of the desired formula is exactly (8.1) for $\sigma = \gamma_k$. By the assumptions on the convergence of traces, the following estimates hold:

$$\mathrm{Tr}_{\sigma_1}(A_1, A_2, \dots, A_k) = O(1),$$
$$\mathrm{Tr}_{\sigma_2}(B_1, B_2, \dots, B_k) = O(N^{k-|\sigma_2|}).$$

Since $\mathrm{Wg}(\sigma_3, N) = O(N^{-k-|\sigma_3|})$, for a triple $\sigma_1, \sigma_2, \sigma_3$ such that $\sigma_1\sigma_2\sigma_3 = \gamma_k$, the contribution of the summand is $O(N^{-|\sigma_2|-|\sigma_3|})$. Therefore, the leading term of (8.1) is of order $O(1)$ and it appears when $|\sigma_2| = |\sigma_3| = 0$, i.e., only when $\sigma_2 = \sigma_3 = e_k$ and $\sigma_1 = \gamma_k$. Since $\mathrm{M}(\gamma_k) = 1$ and so $\mathrm{Wg}(\gamma_k, N) = N^{-k}(1 + O(N^{-2}))$, we obtain the desired formula (8.6).

Proof of (8.7)*—Step 1: reduction of the problem.* We prove a slightly stronger result: taking additional $*$-polynomials of matrices $A_r = \boldsymbol{P}_r(A(i,N) : i \in I)$ and $B_r = \boldsymbol{Q}_r(U(N)B(j,N)U(N)^* : j \in J)$, $k+1 \le r \le 4k$, and setting

$$X_i = \mathrm{Tr}\big(A_{(i-1)k+1}B_{(i-1)k+1}\cdots A_{(i-1)k+k}B_{(i-1)k+k}\big), \qquad i = 1, 2, 3, 4,$$
$$\mathring{X}_i = X_i - \mathbb{E}[X_i],$$

we prove

$$\mathbb{E}[\mathring{X}_1\mathring{X}_2\mathring{X}_3\mathring{X}_4] = O(N^{-2}). \tag{8.8}$$

This implies the desired (8.7) in the special case $X_1 = X_3 = \overline{X_2} = \overline{X_4}$. Note that the last condition $X_2 = \overline{X_1}$ can be satisfied by selecting $A_{k+1} := A_k^*$, $B_{k+1} := B_{k-1}^*$, $A_{k+2} := A_{k-1}^*$, $B_{k+2} := B_{k-2}^*$, $\dots$, $B_{2k-1} := B_1^*$, $A_{2k} := A_1^*$, $B_{2k} := B_k^*$, which yields

$$\overline{X_1} = \mathrm{Tr}((A_1B_1A_2B_2\cdots A_kB_k)^*) = \mathrm{Tr}_n(A_k^*B_{k-1}^*A_{k-1}^*\cdots A_2^*B_1^*A_1^*B_k^*) = X_2.$$

Let I_i be the interval $\{(i-1)k+1, (i-1)k+2, \dots, (i-1)k+k\}$, $\gamma_k^{(i)}$ the cyclic permutation $((i-1)k+1, (i-1)k+2, \dots, (i-1)k+k)$ of $\mathfrak{S}_{I_i}$, and $\gamma_k^{\cup 4} := \gamma_k^{(1)}\gamma_k^{(2)}\gamma_k^{(3)}\gamma_k^{(4)} \in \mathfrak{S}_{4k}$. Let us consider the expansion

$$\mathbb{E}[\mathring{X}_1\mathring{X}_2\mathring{X}_3\mathring{X}_4] = \sum_{J\subseteq[4]}\mathbb{E}_J, \qquad \text{where} \qquad \mathbb{E}_J := (-1)^{\#J}\mathbb{E}\left[\prod_{i\in J}X_i\right]\prod_{i\in[4]\setminus J}\mathbb{E}[X_i].$$

For each $J \subseteq [4]$, from the Weingarten formulas applied to $\mathbb{E}\left[\prod_{i\in J}X_i\right]$ and $\mathbb{E}[X_i]$ $(i \in [4]\setminus J)$, there exists a function $f_J(\sigma_1, \sigma_2, N)$ such that

$$\mathbb{E}_J = \sum_{\substack{\sigma_1,\sigma_2,\sigma_3\in\mathfrak{S}_{4k}\\ \sigma_1\sigma_2\sigma_3=\gamma_k^{\cup 4}}}\mathrm{Tr}_{\sigma_1}(A_1, A_2, \dots, A_{4k})\mathrm{Tr}_{\sigma_2}(B_1, B_2, \dots, B_{4k})f_J(\sigma_1, \sigma_2, N).$$

The function f_J is either a product of Weingarten functions with signs or zero. Take $J = \{1,3,4\}$ for example. The product of Weingarten formulas (8.1) for $\sigma = \gamma_k^{(2)}$ and for $\sigma = \gamma_k^{(1)}\gamma_k^{(3)}\gamma_k^{(4)}$ gives permutations of $\mathfrak{S}_{I_2} \times \mathfrak{S}_{I_1 \cup I_3 \cup I_4}$, and so, for $\sigma_1, \sigma_2, \sigma_3 \in \mathfrak{S}_{4k}$ with $\sigma_1\sigma_2\sigma_3 = \gamma_k^{\cup 4}$,

$$f_J(\sigma_1, \sigma_2, N) = \begin{cases} -\mathrm{Wg}(\sigma_3\restriction_{I_2}, N)\mathrm{Wg}(\sigma_3\restriction_{I_1\cup I_3\cup I_4}, N), & \text{if } \sigma_1, \sigma_2 \text{ preserve } I_2 \text{ and } I_1 \cup I_3 \cup I_4, \\ 0, & \text{otherwise.} \end{cases}$$

The function f_J has a similar expression for general J.

We have the expression

$$\mathbb{E}[\mathring{X}_1\mathring{X}_2\mathring{X}_3\mathring{X}_4] = \sum_{\substack{\sigma_1,\sigma_2,\sigma_3 \in \mathfrak{S}_{4k} \\ \sigma_1\sigma_2\sigma_3 = \gamma_k^{\cup 4}}} \mathrm{Tr}_{\sigma_1}(A_1, A_2, \dots, A_{4k})\mathrm{Tr}_{\sigma_2}(B_1, B_2, \dots, B_{4k}) f(\sigma_1, \sigma_2, N),$$

where

$$f(\sigma_1, \sigma_2, N) := \sum_{J \subseteq [4]} f_J(\sigma_1, \sigma_2, N). \tag{8.9}$$

From the assumption of convergence of traces, we have the estimates $\mathrm{Tr}_{\sigma_1}(A_1, A_2, \dots, A_{4k}) = O(1)$ and $\mathrm{Tr}_{\sigma_2}(B_1, B_2, \dots, B_{4k}) = O(N^{\ell(\sigma_2)}) = O(N^{4k-|\sigma_2|})$. Combined with (8.3) these estimates yield

$$\mathrm{Tr}_{\sigma_2}(B_1, B_2, \dots, B_{4k}) f(\sigma_1, \sigma_2, N) = O(N^{-|\sigma_2|-|\sigma_3|}).$$

We show the estimates

$$\mathrm{Tr}_{\sigma_2}(B_1, B_2, \dots, B_{4k}) f(\sigma_1, \sigma_2, N) = O(N^{-2}), \tag{8.10}$$

which suffice to finish the proof of (8.8), and hence of (8.7).

Proof of (8.7)—*Step 2: proof of* (8.10). Given $\sigma_1, \sigma_2, \sigma_3 \in \mathfrak{S}_{4k}$ with $\sigma_1\sigma_2\sigma_3 = \gamma_k^{\cup 4}$, we introduce an equivalence relation $\sim$ on $[4k]$: $i \sim j$ if there exists $\tau \in \mathrm{Grp}\langle\sigma_1, \sigma_2, \sigma_3\rangle$ such that $\tau(i) = j$. Since this group contains $\gamma_k^{\cup 4}$, every interval I_i must be a subset of some equivalence class. Then the permutations $\sigma_1, \sigma_2, \sigma_3$ associate the set partition $\pi(\sigma_1, \sigma_2, \sigma_3) = \{P_1, P_2, \dots, P_m\}$ of $[4]$ such that the subsets of $[4k]$

$$\bigcup_{i \in P_1} I_i, \quad \bigcup_{i \in P_2} I_i, \ \dots, \ \bigcup_{i \in P_m} I_i$$

are exactly the equivalence classes.

On the other hand, a subset $J \subseteq [4]$ also associates the set partition $\pi(J) = \{J, \{p\} : p \in [4] \setminus J\}$ of $[4]$. Since $f_J(\sigma_1, \sigma_2, N)$ vanishes if σ_1 or σ_2 do not preserve one of the subsets $\bigcup_{i \in J} I_i$ and I_p ($p \in [4] \setminus J$), the only f_J's satisfying $\pi(J) \geq \pi(\sigma_1, \sigma_2, \sigma_3)$ contribute to f in the sum (8.9) and the other f_J's are zero. We discuss several cases according to $\pi(\sigma_1, \sigma_2, \sigma_3)$.

Case 1: $\pi(\sigma_1, \sigma_2, \sigma_3) = \{[4]\}$, or equivalently, the group $\mathrm{Grp}\langle\sigma_1, \sigma_2, \sigma_3\rangle$ acts on $[4k]$ transitively. In this case, f_J vanishes unless $J = [4]$, and so $f = f_{[4]} = \mathrm{Wg}(\sigma_3, N)$. The case $|\sigma_2| + |\sigma_3| = 0$ is irrelevant because the condition $\sigma_1\sigma_2\sigma_3 = \gamma_k^{\cup 4}$ contradicts transitivity. The case $|\sigma_2| + |\sigma_3| = 1$ is also irrelevant because then one would be the identity and the other would be a transposition, and again the condition $\sigma_1\sigma_2\sigma_3 = \gamma_k^{\cup 4}$ would contradict the transitivity. Therefore, only the case $|\sigma_2| + |\sigma_3| \geq 2$ occurs, and so $\mathrm{Tr}_{\sigma_2}(B_1, B_2, \dots, B_{4k}) f(\sigma_1, \sigma_2, N) = O(N^{-2})$.

Case 2: $\pi(\sigma_1, \sigma_2, \sigma_3)$ is a pair partition, i.e., its each block has cardinality two. In this case again we have $f = f_{[4]} = \mathrm{Wg}(\sigma_3, N)$, and from a similar reasoning we must have $|\sigma_2| + |\sigma_3| \geq 2$ and hence $\mathrm{Tr}_{\sigma_2}(B_1, B_2, \dots, B_{4k}) f(\sigma_1, \sigma_2, N) = O(N^{-2})$.

In the other cases, a cancellation occurs between Wg functions.

Case 3: $\pi(\sigma_1, \sigma_2, \sigma_3)$ has two blocks with cardinality 1 and 3. For example, let us consider the case $\pi(\sigma_1, \sigma_2, \sigma_3) = \{\{1,3,4\}, \{2\}\}$. The equivalence classes are $I_1 \cup I_3 \cup I_4$ and I_2. The only indices J's for which f_J is non-zero are $J = \{1,3,4\}, \{1,2,3,4\}$. In these cases we have

$$\begin{aligned} f_{\{1,3,4\}} &= -\mathrm{Wg}(\sigma_3\restriction_{I_2}, N)\mathrm{Wg}(\sigma_3\restriction_{I_1\cup I_3\cup I_4}, N), \\ f_{\{1,2,3,4\}} &= \mathrm{Wg}(\sigma_3, N). \end{aligned}$$

By (8.3) and the multiplicativity (8.2) of M, we obtain

$$\begin{aligned} f &= f_{\{1,3,4\}} + f_{\{1,2,3,4\}} \\ &= N^{-4k-|\sigma_3|}\Big[\underbrace{-\mathrm{M}(\sigma_3\restriction_{I_2})\mathrm{M}(\sigma_3\restriction_{I_1\cup I_3\cup I_4}) + \mathrm{M}(\sigma_3)}_{=0} + O(N^{-2})\Big] \\ &= O(N^{-4k-|\sigma_3|-2}). \end{aligned}$$

Thus $\mathrm{Tr}_{\sigma_2}(B_1, B_2, \dots, B_{4k}) f(\sigma_1, \sigma_2, N) = O(N^{-|\sigma_2|-|\sigma_3|-2}) = O(N^{-2})$. The other cases of $\pi(\sigma_1, \sigma_2, \sigma_3)$ are similar.

Case 4: $\pi(\sigma_1,\sigma_2,\sigma_3)$ has three blocks. Let us consider the example $\pi(\sigma_1,\sigma_2,\sigma_3)=\{\{1\},\{2\},\{3,4\}\}$. The indices J's for which f_J is non-zero are $J=\{3,4\},\{1,3,4\},\{2,3,4\},\{1,2,3,4\}$. We see that

$$\begin{aligned}
f &= f_{\{3,4\}}+f_{\{1,3,4\}}+f_{\{2,3,4\}}+f_{\{1,2,3,4\}} \\
&= \mathrm{Wg}(\sigma_3\restriction_{I_1},N)\mathrm{Wg}(\sigma_3\restriction_{I_2},N)\mathrm{Wg}(\sigma_3\restriction_{I_3\cup I_4},N)-\mathrm{Wg}(\sigma_3\restriction_{I_2},N)\mathrm{Wg}(\sigma_3\restriction_{I_1\cup I_3\cup I_4},N) \\
&\quad -\mathrm{Wg}(\sigma_3\restriction_{I_1},N)\mathrm{Wg}(\sigma_3\restriction_{I_2\cup I_3\cup I_4},N)+\mathrm{Wg}(\sigma_3,N) \\
&= N^{-4k-|\sigma_3|}\Big[\mathrm{M}(\sigma_3\restriction_{I_1})\mathrm{M}(\sigma_3\restriction_{I_2})\mathrm{M}(\sigma_3\restriction_{I_3\cup I_4})-\mathrm{M}(\sigma_3\restriction_{I_2})\mathrm{M}(\sigma_3\restriction_{I_1\cup I_3\cup I_4}) \\
&\qquad\qquad -\mathrm{M}(\sigma_3\restriction_{I_1})\mathrm{M}(\sigma_3\restriction_{I_2\cup I_3\cup I_4})+\mathrm{M}(\sigma_3)+O(N^{-2})\Big] \\
&= O(N^{-4k-|\sigma_3|-2}).
\end{aligned}$$

The other $\pi(\sigma_1,\sigma_2,\sigma_3)$'s can be handled in the same way.

Case 5: $\pi(\sigma_1,\sigma_2,\sigma_3)=\{\{1\},\{2\},\{3\},\{4\}\}$, i.e., every interval I_i is preserved by $\sigma_1,\sigma_2,\sigma_3$. In this case f_J for all the 16 subsets $J\subseteq[4]$ contribute to f. By the multiplicativity (8.2), the dominant contribution to f is the sum of 16 terms

$$\pm N^{-4k-|\sigma_3|}\mathrm{M}(\sigma_3\restriction_{I_1})\mathrm{M}(\sigma_3\restriction_{I_2})\mathrm{M}(\sigma_3\restriction_{I_3})\mathrm{M}(\sigma_3\restriction_{I_4}).$$

Exactly half of them have the minus sign, so their sum cancel and we obtain $f=O(N^{-4k-|\sigma_3|-2})$.

The above arguments finish the proof of (8.10). □

Theorem 8.2. *Under the assumptions and notation of Proposition 8.1, we have*

$$\lim_{N\to\infty}\mathrm{Tr}(A_1B_1A_2B_2\cdots A_kB_k)=\Big[\lim_{N\to\infty}\mathrm{Tr}(A_1A_2\cdots A_k)\Big]\prod_{p=1}^{k}\Big[\lim_{N\to\infty}\frac{1}{N}\mathrm{Tr}(B_p)\Big]\qquad a.s.$$

Proof. The assumptions (8.4) and (8.5) imply that the limits

$$\lim_{N\to\infty}\mathrm{Tr}(A_1A_2\cdots A_k),\qquad \lim_{N\to\infty}\frac{1}{N}\mathrm{Tr}(B_p)$$

exist, and then the proven estimate (8.6) further implies

$$\lim_{N\to\infty}\mathbb{E}[\mathrm{Tr}(A_1B_1A_2B_2\cdots A_kB_k)]=\Big[\lim_{N\to\infty}\mathrm{Tr}(A_1A_2\cdots A_k)\Big]\prod_{p=1}^{k}\Big[\lim_{N\to\infty}\frac{1}{N}\mathrm{Tr}(B_p)\Big].\tag{8.11}$$

On the other hand, taking the sum $\sum_{N=1}^\infty$ of the estimate (8.7) yields

$$\sum_{N=1}^{\infty}\Big|\mathrm{Tr}(A_1B_1A_2B_2\cdots A_kB_k)-\mathbb{E}[\mathrm{Tr}(A_1B_1A_2B_2\cdots A_kB_k)]\Big|^4<+\infty\quad\text{a.s.},$$

so that the summand converges to zero a.s. in the limit $N\to\infty$. Combining this fact and (8.11) finishes the proof. □

Remark 8.3. To show the almost sure convergence, we have computed the L^4-norm in Proposition 8.1. Of course, the L^2-norm is easier to calculate, but it only gives the estimate $O(N^{-1})$, which is insufficient to deduce the almost sure convergence. One can confirm this through a simple example: for a rank-one projection $P=\mathrm{diag}(1,0^{N-1})$, a deterministic matrix $B\in\mathrm{M}_N(\mathbb{C})$ and a Haar unitary U of size N, we can see

$$\begin{aligned}
&\mathbb{E}\Big[\big|\mathrm{Tr}(PUBU^*)-\mathbb{E}[\mathrm{Tr}(PUBU^*)]\big|^2\Big] \\
&=\mathbb{E}\Big[\big|\mathrm{Tr}(PUBU^*)\big|^2\Big]-\big|\mathbb{E}[\mathrm{Tr}(PUBU^*)]\big|^2 \\
&=\mathbb{E}[\mathrm{Tr}(PUBU^*)\mathrm{Tr}(PUB^*U^*)]-\big|\mathbb{E}[\mathrm{Tr}(PUBU^*)]\big|^2 \\
&=\mathrm{Tr}(P)^2|\mathrm{Tr}(B)|^2\mathrm{Wg}(e_2,N)+\mathrm{Tr}(P)^2\mathrm{Tr}(BB^*)\mathrm{Wg}((1,2),N) \\
&\quad+\mathrm{Tr}(P^2)|\mathrm{Tr}(B)|^2\mathrm{Wg}((1,2),N)+\mathrm{Tr}(P^2)\mathrm{Tr}(BB^*)\mathrm{Wg}(e_2,N) \\
&\quad-\mathrm{Tr}(P)^2|\mathrm{Tr}(B)|^2\mathrm{Wg}(e_1,N)^2.
\end{aligned}$$

Applying formula (8.3) to the above expressions, we can see that the first and fifth terms contain a cancellation and yield $O(N^{-2})$ and the second term is $O(N^{-2})$; however, the third and fourth terms do not cancel and contribute $O(N^{-1})$.

Let us consider the partial trace

$$\varphi_\ell(X):=\frac{1}{\ell}\sum_{i=1}^{\ell}X_{i,i},\qquad X=(X_{i,j})_{i,j\in[N]},\ 1\le\ell\le N.\tag{8.12}$$

Corollary 8.4 (Asymptotic monotone independence)**.** *Let $\ell\in\mathbb{N}$ be fixed, $\tilde{A}(i,N)$ ($i\in I$, $N\in\mathbb{N}$) be deterministic $\ell\times\ell$ matrices such that $\lim_{N\to\infty}\tilde{A}(i,N)$ exists in $\mathrm{M}_\ell(\mathbb{C})$. Let*

$$A(i,N):=\begin{pmatrix}\tilde{A}(i,N) & O\\ O & O\end{pmatrix}\in\mathrm{M}_N(\mathbb{C}),\qquad N\ge\ell.$$

Let $(B(j,N))_{j\in J}$ be a family of $N\times N$ matrices that satisfy assumption (8.5)*. Let $U(N)$ be a Haar unitary random matrix of size N. Let $k\ge 2$ and $\boldsymbol{P}_r(x_i:i\in I)$, $r\in[k-1]$, be noncommutative $*$-polynomials without constant terms, and $\boldsymbol{Q}_s(y_j:j\in J)$,*

$s \in [k]$, be noncommutative $*$-polynomials. Then the matrices $A_r := \boldsymbol{P}_r(A(i,N) : i \in I)$ and $B_s := \boldsymbol{Q}_s(U(N)B(j,N)U(N)^* : j \in J)$ satisfy

$$\lim_{N\to\infty} \varphi_\ell(B_p) = \lim_{N\to\infty} \frac{1}{N}\mathrm{Tr}(B_p) \quad a.s., \qquad p \in [k], \tag{8.13}$$

$$\lim_{N\to\infty} \varphi_\ell(B_1A_1B_2A_2\cdots A_{k-1}B_k) = \lim_{N\to\infty} \varphi_\ell(A_1A_2\cdots A_{k-1}) \prod_{p=1}^k \left[\lim_{N\to\infty} \varphi_\ell(B_p)\right] \quad a.s. \tag{8.14}$$

Proof. To show (8.13), we apply Theorem 8.2 to the rank-ℓ projection $P_\ell(N) = \mathrm{diag}(1^\ell, 0^{N-\ell})$ regarded as a family of single matrix, and the family $(B(j,N) : j \in J)$. Then Theorem 8.2 for $k = 1$ and $A_1 = P_\ell(N)$ is exactly the desired (8.13).

To show (8.14), we again apply Theorem 8.2 now to the families $(P_\ell(N), A(i,N) : i \in I)$ and $(B(j,N) : j \in J)$, the former of which satisfies assumption (8.4). Then Theorem 8.2 by selecting $A_1 := P_\ell(N)$, together with the proven formula (8.13), yields the desired (8.14). □

Remark 8.5. For $\ell = 1$, the factorization (8.14) holds without taking the limit $N \to \infty$; see Example 1.18.

Remark 8.6. Results of this section can be extended to the case when $(B(j,N) : j \in J)$ is a family of random matrices independent of the Haar unitary $U(N)$ with the additional assumption that for any $k \in \mathbb{N}$, $j_1, j_2, \dots, j_k \in J$ and $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k \in \{1, *\}$,

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{Tr}(B(j_1)^{\varepsilon_1}\cdots B(j_k)^{\varepsilon_k}) = \lim_{N\to\infty} \mathbb{E}\left[\frac{1}{N}\mathrm{Tr}(B(j_1)^{\varepsilon_1}\cdots B(j_k)^{\varepsilon_k})\right] \in \mathbb{C} \quad \text{a.s.} \tag{8.15}$$

with shorthand symbol $B(j) := B(j,N)$. Then we modify Proposition 8.1 as follows.

- We only take the expectation in (8.6) for the Haar unitary part:

$$\mathbb{E}_U[\mathrm{Tr}(A_1B_1\cdots A_kB_k)] = \mathrm{Tr}(A_1\cdots A_k)\prod_{p=1}^k\left[\frac{1}{N}\mathrm{Tr}(B_p)\right] + O(N^{-1})f(B_1,\dots,B_k,N). \tag{8.16}$$

  Here the term $f(B_1,\dots,B_k,N)$ is a polynomial in N^{-1} and the normalized traces of B_i's that are bounded a.s. by assumption (8.15).

- Instead of (8.7) we can show

$$\mathbb{E}\Big[\big|\mathrm{Tr}(A_1B_1A_2B_2\cdots A_kB_k) - \mathbb{E}_U[\mathrm{Tr}(A_1B_1A_2B_2\cdots A_kB_k)]\big|^4\Big] = O(N^{-2}). \tag{8.17}$$

  For this first we show (8.7) only by taking the expectation concerning the Haar unitary U and fixing the $B(j,N)$'s. Then the term $O(N^{-2})$ in (8.7) would include random variables of the form $\frac{1}{N}\mathrm{Tr}(B(j_1)^{\varepsilon_1}\cdots B(j_k)^{\varepsilon_k})$. Further taking the expectation with respect to $B(j,N)$'s and using the condition (8.15) we can deduce (8.17). The modified formulas (8.16) and (8.17) yield the same conclusions Theorem 8.2 and Corollary 8.4.

If $(B(j,N) : j \in J)$ is an independent family of normalized GUEs, then the above requirements are satisfied because a GUE has the same law as UDU^* where D is a random diagonal matrix and U is a Haar unitary matrix such that D, U are independent. Condition (8.15) is satisfied because the convergence of the expected traces is a consequence of Voiculescu's asymptotic freeness [151], and the almost sure convergence is known in Hiai and Petz [89, Corollary 4.3.6].

8.3. **Outliers of additive and multiplicative perturbations.** We consider finite-rank perturbations of random matrices. In this section we always assume that the eigenvalues $\lambda_i(X)$, $i \in [N]$, of a Hermitian matrix $X \in \mathrm{M}_N(\mathbb{C})$ are arranged in the way

$$\lambda_1(X) \ge \lambda_2(X) \ge \cdots \ge \lambda_N(X).$$

Let μ_X be the empirical eigenvalue distribution of X and ν_X^ℓ the analytic distribution of X with respect to the partial trace φ_ℓ defined in (8.12). As being an analytic distribution, ν_X^ℓ is supported on $\mathrm{Sp}(X)$ that is the set of the eigenvalues of X. By means of the spectral measure E_X of X, ν_X^ℓ can be explicitly expressed as $\nu_X^\ell = \varphi_\ell \circ E_X$ along the same idea as Proposition 1.9.

We will use Weyl's inequalities. These can be proved from the min-max theorem and the reader is referred to [34, Chapter III.2] for the proofs.

Lemma 8.7. *Let X, Y be $N \times N$ Hermitian matrices.*

(i) *$\lambda_{i+j-1}(X+Y) \le \lambda_i(X) + \lambda_j(Y)$ holds for all $i, j \in [N]$ with $i + j - 1 \le N$.*

(ii) *If $X \le Y$ then $\lambda_i(X) \le \lambda_i(Y)$ for all $i \in [N]$, where $X \le Y$ means that $Y - X$ is a positive semi-definite matrix.*

(iii) *$|\lambda_i(X) - \lambda_i(Y)| \le \|X - Y\|$ for all $i \in [N]$, where $\|\cdot\|$ is the operator norm.*

Theorem 8.8. *Let $\ell \in \mathbb{N}$ and $m \in \mathbb{N}_0$ be fixed, $\theta_1 \ge \theta_2 \ge \cdots \ge \theta_\ell > 0 > \theta_{\ell+1} \ge \cdots \ge \theta_{\ell+m}$ and*

$$P = P(N) := \mathrm{diag}(\theta_1, \theta_2, \dots, \theta_{\ell+m}, 0, 0, \dots, 0) \in \mathrm{M}_N(\mathbb{C}), \qquad N \ge \ell + m.$$

Let $U = U(N)$ be a Haar unitary random matrix of size N defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $B = B(N)$ be an $N \times N$ Hermitian deterministic matrix. We consider the random matrix[¶]

$$X_N = X_N(\theta_1, \theta_2, \dots, \theta_\ell) := UBU^* + P.$$

[¶]We will consider $\theta_1, \theta_2, \dots, \theta_\ell$ as variables in the proof, while the negative parameters $\theta_{\ell+1}, \dots, \theta_{\ell+m}$ are always fixed.

Suppose that there exists a probability measure μ on $\mathbb{R}$ such that $\beta := \max\operatorname{supp}(\mu) < +\infty$, μ has finite moments of all orders that satisfy Carleman's condition (A.2)*, and*

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{Tr}(B^k) = \int_{\mathbb{R}} x^k\,\mu(\mathrm{d}x), \qquad k \in \mathbb{N}, \tag{8.18}$$

$$\lim_{N\to\infty} \lambda_1(B) = \beta. \tag{8.19}$$

Let $\gamma := F_\mu(\beta + 0)$ that lies in $[0, +\infty)$. For each $i \in [\ell]$ the following hold.

(i) *The empirical eigenvalue distribution of X_N converges weakly to μ a.s.*

(ii) *If $\theta_i > \gamma$, then the equation $F_\mu(x) = \theta_i$ has a unique solution $x = x_i \in (\beta, +\infty)$ and $\lim_{N\to\infty} \lambda_i(X_N) = x_i$ a.s.*

(iii) *If $\theta_i \le \gamma$ then $\lim_{N\to\infty} \lambda_i(X_N) = \beta$ a.s.*

Remark 8.9. The result also holds when B is a random matrix independent of U and satisfying conditions (8.15), (8.18) and (8.19) almost surely.

Proof. Before going into the details, it would be helpful for the reader to have the key idea. Let us use the simplified notation $\nu_Z = \nu_Z^{\ell+m}$. By Corollary 8.4, the matrices P and UBU^* are asymptotically monotonically independent with respect to the state $\varphi_{\ell+m}$. Then we can identify the limit distribution of ν_{X_N} as the monotone convolution of ν_P and μ. If this distribution has an atom in $(\beta, +\infty)$, then the matrix X_N must have an eigenvalue near the point, which becomes an outlier. This point is exactly the solution x to the equation $F_\mu(x) = \theta_i$. However, this argument only shows the existence of an outlier and does not tell us the number or multiplicities of them. Fortunately, Weyl's inequalities provide sufficiently sharp estimates of the number of outliers. The details are as follows.

Weak convergence of μ_B and μ_{X_N}. Assumption (8.18) and the determinacy of the moment sequence of μ imply that the empirical eigenvalue distribution μ_B converges weakly to μ. Concerning μ_{X_N}, we expand X_N^k into monomials in UBU^* and P. Each monomial except $(UBU^*)^k$ has at least one factor P, so that by Theorem 8.2 its evaluation by Tr converges almost surely to a finite value. Therefore, the evaluation by the normalized trace converges to zero. The above arguments yield

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{Tr}(X_N^k) = \lim_{N\to\infty} \frac{1}{N}\mathrm{Tr}((UBU^*)^k) = \lim_{N\to\infty} \frac{1}{N}\mathrm{Tr}(B^k) \quad \text{a.s.,} \quad k \in \mathbb{N}. \tag{8.20}$$

This implies the weak convergence $\mu_{X_N} \to \mu$ a.s.

General estimates on eigenvalues. As a preparation for proving (ii) and (iii) we derive some facts from Weyl's inequalities. First, the weak convergence $\mu_B \to \mu$ and (8.19) imply

$$\lim_{N\to\infty} \lambda_i(B) = \beta, \qquad i \in \mathbb{N}, \tag{8.21}$$

because otherwise μ would be supported on a smaller interval $(-\infty, \beta - \varepsilon)$.

Next, by Lemma 8.7 (i) we have

$$\lambda_{i+j-1}(X_N) \le \lambda_i(B) + \lambda_j(P), \qquad i + j - 1 \le N.$$

Since P has ℓ positive eigenvalues, we conclude $\lambda_{i+\ell}(X_N) \le \lambda_i(B)$ for all $1 \le i \le N - \ell$. Combining this and (8.21) yields

$$\limsup_{N\to\infty} \lambda_{i+\ell}(X_N) \le \beta, \qquad i \in \mathbb{N}. \tag{8.22}$$

Since the weak convergence limit of μ_{X_N} is also μ, we must have

$$\liminf_{N\to\infty} \lambda_i(X_N) \ge \beta \quad \text{a.s.,} \qquad i \in \mathbb{N}; \tag{8.23}$$

otherwise the measure μ would be supported on $(-\infty, \beta - \varepsilon)$ for some $\varepsilon > 0$, which would be a contradiction. The previous two inequalities imply

$$\lim_{N\to\infty} \lambda_{i+\ell}(X_N) = \beta \quad \text{a.s.,} \qquad i \in \mathbb{N}. \tag{8.24}$$

Asymptotic monotone independence. Corollary 8.4 can be applied to the families of single matrices $\{P(N)\}$ and $\{B(N)\}$, which yields with notation $\tilde{B} := UBU^*$

$$\lim_{N\to\infty} \varphi_{\ell+m}(\tilde{B}^p) = \lim_{N\to\infty} \frac{1}{N}\mathrm{Tr}(\tilde{B}^p) \quad \text{a.s.,} \qquad p \in \mathbb{N}_0, \tag{8.25}$$

$$\lim_{N\to\infty} \varphi_{\ell+m}(\tilde{B}^{p_0} P^{q_1} \tilde{B}^{p_1} \cdots P^{q_k} \tilde{B}^{p_k}) = \lim_{N\to\infty} \varphi_{\ell+m}(P^{q_1+\cdots+q_k}) \prod_{i=0}^{k} \left[\lim_{N\to\infty} \varphi_{\ell+m}(\tilde{B}^{p_i})\right] \quad \text{a.s.} \tag{8.26}$$

for all $k \in \mathbb{N}$, $q_i \in \mathbb{N}$ and $p_i \in \mathbb{N}_0$. A technical issue to note here is that the almost sure convergence above holds on an event $\Omega_{\boldsymbol{\theta}_+}$ of probability one depending on $\boldsymbol{\theta}_+ = (\theta_1, \theta_2, \ldots, \theta_\ell)$. Because we will change the parameter $\boldsymbol{\theta}_+$ later, let us consider the countable set

$$S := \{(\theta_1, \theta_2, \ldots, \theta_\ell) \in \mathbb{Q}^\ell : \theta_1 > \theta_2 > \cdots > \theta_\ell > 0\}$$

and an event $\Omega' \in \mathcal{F}$ with probability one defined by

$$\Omega' := \{\omega \in \Omega : \text{(8.20), (8.22), (8.23), (8.25) and (8.26) hold}$$
$$\text{for all } i, k \in \mathbb{N},\ p, p_i \in \mathbb{N}_0,\ q_i \in \mathbb{N} \text{ and } \boldsymbol{\theta}_+ \in S\}.$$

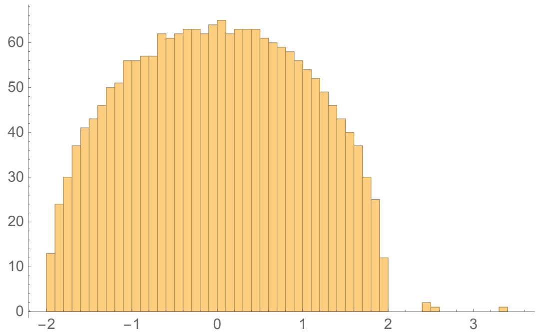

FIGURE 7. A histogram for the eigenvalues of $G + P$, where G is the normalized GUE of matrix size 2000, $P = \mathrm{diag}(3, 2, 2, 2, 1, 0, 0, \dots, 0)$ and the bin size is selected to be $1/10$.

For any sample $\omega \in \Omega'$ it follows from (8.26) and Proposition 5.13 that

$$\lim_{N\to\infty} \int_{\mathbb{R}} x^k \, \nu_{X_N}(\mathrm{d}x) = \lim_{N\to\infty} \varphi_{\ell+m}((P + \tilde{B})^k) = \int_{\mathbb{R}} x^k (\nu_P \triangleright \mu)(\mathrm{d}x), \qquad k \in \mathbb{N},$$

where

$$\nu_P = \frac{1}{\ell + m}(\delta_{\theta_1} + \delta_{\theta_2} + \cdots + \delta_{\theta_{\ell+m}}).$$

By Proposition A.5, $\nu_P \triangleright \mu$ has a determinate moment sequence, and hence by Proposition A.7, ν_{X_N} converges weakly to $\nu_P \triangleright \mu$. According to Theorem 5.1, the limit distribution $\nu_P \triangleright \mu$ has the Cauchy transform

$$G_{\nu_P}(F_\mu(z)) = \frac{1}{\ell + m}\left(\frac{1}{F_\mu(z) - \theta_1} + \frac{1}{F_\mu(z) - \theta_2} + \cdots + \frac{1}{F_\mu(z) - \theta_{\ell+m}}\right).$$

Since μ is supported on $(-\infty, \beta]$, the reciprocal Cauchy transform F_μ has an analytic continuation to $\mathbb{C} \setminus (-\infty, \beta]$, strictly increasing and taking positive values on $(\beta, +\infty)$, and $\lim_{x\to+\infty} F_\mu(x) = +\infty$; see Proposition 4.42. This implies $\gamma = F_\mu(\beta+0) \in [0, +\infty)$ and, if $\theta_i > \gamma$, the existence of a unique solution x_i to $F_\mu(x) = \theta_i$ as stated in (ii).

Completing the proof. Now we are ready to finish the proofs of (ii) and (iii).

Case 1: $\boldsymbol{\theta}_+ \in S$, $\theta_1 > \cdots > \theta_\ell > \gamma$. Since for each $i \in [\ell]$ the equation $F_\mu(x) = \theta_i$ has a solution $x = x_i > \beta$ with $x_1 > x_2 > \cdots > x_\ell > \beta$, the function $G_{\nu_P}(F_\mu(z))$ has a pole at x_i, so that the monotone convolution $\nu_P \triangleright \mu$ has an atom at each x_i. We take $\varepsilon > 0$ so that $x_\ell > \beta + \varepsilon$. This implies that for sufficiently large N, the matrix X_N has at least one eigenvalue close to x_i for each i, so that altogether at least ℓ eigenvalues on the interval $(\beta + \varepsilon, +\infty)$. On the other hand, (8.24) implies that X_N has at most ℓ eigenvalues greater than $\beta + \varepsilon$, so that has exactly ℓ eigenvalues on $(\beta + \varepsilon, +\infty)$. This shows $\lim_{N\to\infty} \lambda_i(X_N) = x_i$ for all $i \in [\ell]$ and $\omega \in \Omega'$.

Case 2: $\boldsymbol{\theta}_+ \in S$, $\theta_1 > \cdots > \theta_{\ell-1} > \gamma \geq \theta_\ell$. We again fix a sample $\omega \in \Omega'$. We consider θ_ℓ as a variable. In Case 1, the function $(\gamma, \theta_{\ell-1}) \cap \mathbb{Q} \ni \theta_\ell \mapsto x_\ell(\theta_\ell) \in (\beta, +\infty)$ is continuous and $\lim_{\theta_\ell \downarrow \gamma, \theta_\ell \in \mathbb{Q}} x_\ell(\theta_\ell) = \beta$. By Lemma 8.7 (ii), the function $\mathbb{Q} \ni \theta_\ell \mapsto \lambda_\ell(X_N)$ is non-decreasing, so that for each $\theta_\ell \in (0, \gamma]$ we obtain $\limsup_{N\to\infty} \lambda_\ell(X_N) \leq \beta$. Combined with (8.23) this yields

$$\lim_{N\to\infty} \lambda_\ell(X_N) = \beta.$$

Hence, for each $\varepsilon > 0$, X_N has at most $\ell - 1$ eigenvalues on $(\beta + \varepsilon, +\infty)$ for all large N. On the other hand, $\nu_P \triangleright \mu$ has an atom at $x_i > \beta + \varepsilon$ for $i \in [\ell - 1]$. Therefore, for large N, X_N must have exactly $\ell - 1$ eigenvalues on $(\beta + \varepsilon, +\infty)$. The weak convergence $\nu_{X_N} \to \nu_P \triangleright \mu$ implies the convergence $\lambda_i(X_N) \to x_i$ for each $i \in [\ell - 1]$, finishing the proof of Case 2. Repeating similar arguments yields the statement in the case $\boldsymbol{\theta}_+ \in S$, $\theta_i > \gamma \geq \theta_{i+1}$.

Case 3: $\boldsymbol{\theta}_+ \in \mathbb{R}^\ell$, $\theta_1 \geq \cdots \geq \theta_{\ell-1} > 0$ (the most general case). We take a sequence $\boldsymbol{\theta}_+^{(n)} \in S$ that converges to $\boldsymbol{\theta}_+$. By Lemma 8.7 (iii), we have the uniform estimate $|\lambda_i(X_N(\boldsymbol{\theta}_+)) - \lambda_i(X_N(\boldsymbol{\theta}_+^{(n)}))| \leq \max_{i\in[\ell]} |\theta_i - \theta_i^{(n)}|$, which finishes the proof. □

Example 8.10. Let μ be the standard semicircle distribution $\mathrm{S}(0, 1)$. From (4.40) we have

$$F_\mu(z) = \frac{z + \sqrt{z^2 - 4}}{2}.$$

We see that $F_\mu(2 + 0) = 1$ and for each $\theta > 1$ the solution to $F_\mu(x) = \theta$ is given by $x = \theta + 1/\theta > 2$. Therefore,

$$\lim_{N\to\infty} \lambda_i(UBU^* + P) = \begin{cases} \theta_i + \dfrac{1}{\theta_i}, & \theta_i > 1, \\ 2, & 0 < \theta_i \leq 1. \end{cases}$$

See Figure 7 for a simulation.

A similar result holds for the multiplicative perturbation. Let us denote the identity matrix of size N as $\mathbf{1}_N$. We use the following additional lemmas.

Lemma 8.11. *Let A, B be $N \times N$ matrices. Then, counting multiplicities, AB and BA have the same eigenvalues.*

Proof. Suppose first that A is invertible. Then, for all $\lambda \in \mathbb{C}$,

$$\det(\lambda \mathbf{1}_N - AB) = \det A \det(\lambda A^{-1} - B) = \det(\lambda A^{-1} - B)\det A = \det(\lambda \mathbf{1}_N - BA).$$

If A is not invertible, then the matrix $A_n := A + (1/n)\mathbf{1}_N$ is invertible for all sufficiently large $n \in \mathbb{N}$, and so $\det(\lambda \mathbf{1}_N - A_n B) = \det(\lambda \mathbf{1}_N - BA_n)$ holds. Passing to the limit $n \to \infty$ yields the desired formula $\det(\lambda \mathbf{1}_N - AB) = \det(\lambda \mathbf{1}_N - BA)$. □

Lemma 8.12. *Let X be an $N \times N$ Hermitian matrix and A be an $N \times N$ matrix. Then, counting multiplicities, the number of positive eigenvalues of A^*XA is not greater than that of X.*

Proof. By the min-max theorem, we have

$$\lambda_i(A^*XA) = \max_{\substack{W \subseteq \mathbb{C}^N \\ \dim(W) = i}} \min_{\substack{x \in W \\ \|x\|=1}} \langle x, A^*XAx\rangle.$$

We first assume that A is invertible. Then for each subspace W of dimensions i and a unit vector $x \in W$ we have

$$\langle x, A^*XAx\rangle = \langle Ax, XAx\rangle = \|Ax\|^2 \langle y, Xy\rangle,$$

where $y := Ax/\|Ax\|$ is a unit vector in AW. Note that $Ax \neq 0$ as $\|x\| = 1$ and A is invertible. Since $\|Ax\| \le \|A\|$, we have

$$\langle x, A^*XAx\rangle \le \max\{\|A\|^2\langle y, Xy\rangle, 0\}.$$

As the subspace AW moves over all the subspaces of $\mathbb{C}^N$ of dimensions i, we have $\max\limits_{W} \min\limits_{x\in W} = \max\limits_{W} \min\limits_{y \in W}$ and hence

$$\lambda_i(A^*XA) \le \max\{\|A\|^2 \lambda_i(X), 0\}.$$

This yields the desired claim.

If A is not invertible, then we take a sequence A_n of invertible matrices that converges to A. By Lemma 8.7 (iii), $\lambda_i(A_n^*XA_n)$ converges to $\lambda_i(A^*XA)$ for each i, so that the conclusion follows. □

Theorem 8.13. *Let ℓ, m, P and U be as defined in Theorem 8.8, where we further assume that the negative diagonal entries of P, i.e., $\theta_{\ell+i}$, $i \in [m]$, are not smaller than -1. Let $B = B(N)$ be an $N \times N$ positive semi-definite deterministic matrix and*

$$Y_N := (\mathbf{1}_N + P)^{\frac{1}{2}} UBU^* (\mathbf{1}_N + P)^{\frac{1}{2}}.$$

We assume that the empirical eigenvalue distribution μ_B converges weakly to a probability measure $\mu \neq \delta_0$ supported on an interval $[0, \beta]$ with $0 < \beta < +\infty$, and that $\lambda_1(B)$ converges to β as $N \to \infty$. The function η_μ has analytic continuation to $\mathbb{C} \setminus [1/\beta, +\infty)$, still denoted as η_μ. The limit $\delta := \eta_\mu(1/\beta - 0) \in (0, 1]$ exists and the following holds for each $i \in [\ell]$.

(i) *The empirical eigenvalue distribution of Y_N converges weakly to μ a.s.*

(ii) *If $1 + \theta_i > 1/\delta$, then the equation $\eta_\mu(1/y) = 1/(1+\theta_i)$ has a unique solution $y = y_i \in (\beta, +\infty)$ and $\lim_{N\to\infty} \lambda_i(Y_N) = y_i$ a.s.*

(iii) *If $1 + \theta_i \le 1/\delta$ then $\lim_{N\to\infty} \lambda_i(Y_N) = \beta$ a.s.*

Proof. The proof is quite analogous to that of Theorem 8.8. We omit the details and only mention the main differences. Let us use the same notation $\tilde{B} := UBU^*$ as before.

- Under our assumptions, we have the moment convergence

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{Tr}(B^k) = \int_{[0,\beta]} x^k \,\mu(\mathrm{d}x), \qquad k \in \mathbb{N}.$$

- According to Lemma 8.11, the matrix $\tilde{Y}_N = \tilde{Y}_N(\boldsymbol{\theta}_+) := \sqrt{\tilde{B}}(\mathbf{1}_N + P)\sqrt{\tilde{B}}$ has the same eigenvalues as Y_N. We then use the decomposition $\tilde{Y}_N = \tilde{B} + \sqrt{\tilde{B}}P\sqrt{\tilde{B}}$. According to Lemma 8.12, the number of positive eigenvalues of $\sqrt{\tilde{B}}P\sqrt{\tilde{B}}$ is at most ℓ. Therefore, as in the proof of Theorem 8.8, we can deduce $\lambda_{i+\ell}(Y_N) = \lambda_{i+\ell}(\tilde{Y}_N) \to \beta$ a.s. for each $i \in \mathbb{N}$.

- The almost sure limit of ν_{Y_N} is the multiplicative monotone convolution $\nu_{\mathbf{1}_N+P} \circlearrowright \mu$, where

$$\nu_{\mathbf{1}_N+P} = \frac{1}{\ell+m}\sum_{i=1}^{\ell+m} \delta_{1+\theta_i}.$$

  The relation $\psi_{\nu_{\mathbf{1}_N+P}\circlearrowright\mu} = \psi_{\nu_{\mathbf{1}_N+P}}(\eta_\mu(z))$ reads

$$\psi_{\nu_{\mathbf{1}_N+P}\circlearrowright\mu}(z) = \frac{1}{\ell+m}\sum_{i=1}^{\ell+m} \frac{(1+\theta_i)\eta_\mu(z)}{1-(1+\theta_i)\eta_\mu(z)}.$$

  The function ψ_μ has an analytic continuation to $\mathbb{C} \setminus [1/\beta, +\infty)$, still denoted as ψ_μ, such that $\psi_\mu(\overline{z}) = \overline{\psi_\mu(z)}$. We can check that $\psi_\mu(0) = 0$ and $\psi_\mu' > 0$ on $(0, 1/\beta)$, and so $0 < \eta_\mu = \psi_\mu/(1+\psi_\mu) < 1$ and $\eta_\mu' = \psi_\mu'/(1+\psi_\mu)^2 > 0$ on $(0, 1/\beta)$. Therefore, $\delta := \eta_\mu(1/\beta - 0)$ exists in $(0, 1]$. The equation $\eta_\mu(1/y) = 1/(1+\theta_i)$ has a solution $y = y_i \in (\beta, +\infty)$ if and only if $1 + \theta_i > 1/\delta$, in which case $\psi_{\nu_{\mathbf{1}_N+P}\circlearrowright\mu}$ has a pole at $1/y_i$, and so the measure $\nu_{\mathbf{1}_N+P} \circlearrowright \mu$ has an atom at y_i.

- Using the matrix $\tilde{Y}_N$ above, we can deduce the monotonicity of the map $\theta_i \mapsto \lambda_i(\tilde{Y}_N(\boldsymbol{\theta}_+))$ and the estimate $|\lambda_i(\tilde{Y}_N(\boldsymbol{\theta}_+)) - \lambda_i(\tilde{Y}_N(\boldsymbol{\theta}_+^{(n)}))| \le \beta \max_{i\in[\ell]} |\theta_i - \theta_i^{(n)}|$. □

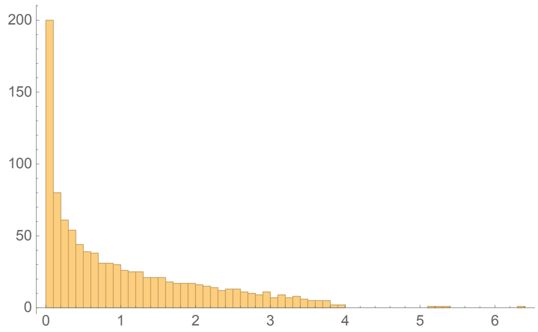

FIGURE 8. A histogram for the eigenvalues of $(\mathbf{1}_N + P)^{\frac{1}{2}} G^* G (\mathbf{1}_N + P)^{\frac{1}{2}}$, where G is the normalized GUE of size $N = 1000$ and $P = \operatorname{diag}(4, 3, 3, 3, 1, 0, 0, \dots, 0)$ and the bin size is selected to be $1/10$. There are three eigenvalues near the theoretical limit $16/3 = 5.333\dots$ and one eigenvalue near 6.25.

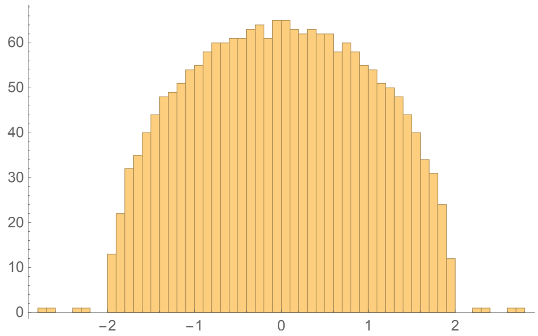

FIGURE 9. A histogram for the eigenvalues of $(\mathbf{1}_N + P)^{\frac{1}{2}} G (\mathbf{1}_N + P)^{\frac{1}{2}}$, where G is the normalized GUE of size $N = 2000$ and $P = \operatorname{diag}(5, 5, 3, 3, 1/2, 0, 0, \dots, 0)$ and the bin size is selected to be $1/10$. This simulation suggests that the number of eigenvalues in $(2+\varepsilon, +\infty)$ in the large N limit is exactly the number of θ_i's larger than one.

Remark 8.14. We can also consider the more general case when B is just Hermitian. For simplicity, assume that μ_B converges to μ supported on a compact interval $[\alpha, \beta]$ with $\alpha < 0 < \beta$. In this case the above method to estimate the number of outliers does not work because $\tilde{Y}_N = \sqrt{\tilde{B}}(\mathbf{1}_N + P)\sqrt{\tilde{B}}$ cannot be defined. However, we can still prove that if $\eta_\mu(1/y) = 1/(1+\theta_i)$ has a solution $y = y_i \in (\beta, +\infty)$ then for any $\varepsilon > 0$ there exists $N_0 \in \mathbb{N}$ such that $\min\{|\lambda_j(Y_N) - y_i| : j \in [N]\} < \varepsilon$ a.s. for all $N \geq N_0$. This is because, as soon as the measure $\nu_{\mathbf{1}_N + P} \circlearrowright \mu$ has an atom at y_i, there exists at least one eigenvalue of Y_N close to y_i. Note that η_μ still has an analytic continuation to $(0, 1/\beta)$ but it might not be increasing on $(0, 1/\beta)$ anymore.

Example 8.15. Suppose that μ is the Marchenko–Pastur distribution

$$\mu(\mathrm{d}x) = \frac{1}{2\pi}\sqrt{\frac{4-x}{x}}\,\chi_{(0,4)}(x)\,\mathrm{d}x.$$

The Cauchy transform of μ is known in Example 4.51 as $G_\mu(z) = (z - \sqrt{z^2 - 4z})/(2z)$, which implies

$$F_\mu(z) = \frac{z + \sqrt{z^2 - 4z}}{2} \qquad \text{and} \qquad \eta_\mu(z) = \frac{1 - \sqrt{1 - 4z}}{2}.$$

We see that $\eta_\mu(1/4 - 0) = 1/2$ and hence for each $\theta > 1$ the solution to $\eta_\mu(1/y) = 1/(1+\theta)$ is given by $y = (\theta+1)^2/\theta > 4$. Therefore, for each $i \in \mathbb{N}$,

$$\lim_{N\to\infty} \lambda_i((\mathbf{1}_N + P)^{\frac{1}{2}} U B U^* (\mathbf{1}_N + P)^{\frac{1}{2}}) = \begin{cases} \theta_i + \dfrac{1}{\theta_i} + 2, & \theta_i > 1, \\ 4, & 0 < \theta_i \leq 1. \end{cases}$$

See Figure 8 for a simulation where UBU^* is selected to be a Wishart matrix.

Example 8.16. Suppose that μ is the standard semicircle distribution $\mathrm{S}(0,1)$. The η-transform is given by

$$\eta_\mu(z) = \frac{1 - \sqrt{1 - 4z^2}}{2}.$$

We see that $\eta_\mu(1/2 - 0) = 1/2$ and hence for each $\theta > 1$ the solution to $\eta_\mu(1/y) = 1/(1+\theta)$ is given by $y = \sqrt{\theta} + 1/\sqrt{\theta} > 2$. If $\theta_i > 1$, from Remark 8.14, we know that there is an eigenvalue of $(\mathbf{1}_N + P)^{\frac{1}{2}} U B U^* (\mathbf{1}_N + P)^{\frac{1}{2}}$ converging to $\sqrt{\theta_i} + 1/\sqrt{\theta_i}$. See Figure 9 for a simulation where UBU^* is selected to be the normalized GUE.

8.4. **Notes.** The change of the location of the largest eigenvalue of random matrices depending on perturbation is often called the BBP phase transition, named after the work of Baik, Ben Arous and Péché [17]. Forrester gives an excellent survey on rank-one perturbations including earlier works [65].

The application of monotone independence to outliers is due to Cébron, Dahlqvist and Gabriel [41]. Our Corollary 8.4 is pointed out in [41] and Theorem 8.8 is an extension of a result stated in [41, Section 1.4]. Theorems 8.8 and 8.13 are close to the results of Benaych-Georges and Nadakuditi [30, Theorems 2.1 and 2.6]. The difference is that our result for additive perturbation model allows the limit distribution to have unbounded support, while if the matrices $B(j,N)$ are random then we need a stronger assumption than [30] even if the limit distribution is compactly supported; see Remark 8.6. The reader is referred to e.g. [22, 130] for further results on outliers.

Formula (8.6) is a slight extension of a result of Shlyakhtenko [141]; note that a stronger statement is given in [46, Theorem 4.1] but it contains an error: the result would be correct if the matrices $A_j(n), B_j(n)$ are deterministic just as in our result (8.6). The almost sure version (Theorem 8.2) is proved by Collins, Hasebe and Sakuma [46]. The factorization formula (8.6) or Theorem 8.2 is abstracted to the notion of "cyclic monotone independence" in [46] and is further developed e.g. in [8, 12, 41, 47, 71]. Another random matrix model that satisfies asymptotic monotone independence is constructed and studied by Lenczewski [104, 105] and by Banna, Mingo and Tseng [18, 114].

Appendix A. Moments and weak convergence

In noncommutative probability, moments of random variables are fundamental concepts. We collect here supplementary results on moments.

A characterization of when a sequence of real numbers is the moment sequence of a finite Borel measure (Hamburger's moment problem) is known as follows. The reader is referred to [3, Theorem 2.1.1] or [140, Theorem 3.8] for the proof.

Theorem A.1. *Let $(\alpha_n)_{n\geq 0}$ be a sequence of real numbers. The following are equivalent.*

(1) *There exists a finite Borel measure τ on $\mathbb{R}$ with finite moments of all orders such that*

$$\alpha_n = \int_{\mathbb{R}} x^n \, \tau(\mathrm{d}x), \qquad n \in \mathbb{N}_0.$$

(2) *The sequence $(\alpha_n)_{n\geq 0}$ is **positive semi-definite**, i.e., for any $n \in \mathbb{N}_0$ and $c_0, c_1, \dots, c_n \in \mathbb{R}$ it holds that*

$$\sum_{i,j=0}^{n} c_i c_j \alpha_{i+j} \geq 0.$$

In the theorem above, the measure τ is in general not unique. This leads to the following definition.

Definition A.2. Let τ be a finite Borel measure on $\mathbb{R}$ having finite moments of all orders. We say that τ has a **determinate moment sequence** if no other finite Borel measures have the same moment sequence.

Example A.3. Let $-1 \leq \varepsilon \leq 1$. The probability measure

$$\mu_\varepsilon(\mathrm{d}x) = \frac{1}{\sqrt{\pi}} x^{-1} e^{-(\log x)^2} [1 - \varepsilon \sin(2\pi \log x)] \, \chi_{(0,+\infty)}(x) \, \mathrm{d}x$$

has an indeterminate moment sequence because the moments are independent of ε. The measure μ_0 is the distribution of the random variable e^X where X has the distribution $\mathrm{N}(0,1/2)$ and is called a lognormal distribution. See [3, p. 88] and [140, pp. 88–90] for more examples of indeterminate moment sequences.

We show a simple criterion for the determinacy of the moment problem.

Proposition A.4. *Let μ be a probability measure on $\mathbb{R}$ with compact support. Then the moment sequence of μ is determinate.*

Proof. Suppose that μ is supported on the compact interval $[-R,R]$. Suppose that ν is a probability measure on $\mathbb{R}$ having the same moment sequence.

Step 1: ν is also supported on $[-R,R]$. For this, let us observe the obvious bound

$$\mathrm{m}_{2k}(\mu) = \int_{[-R,R]} x^{2k} \, \mu(\mathrm{d}x) \leq R^{2k}.$$

Suppose to the contrary that ν is not supported on $[-R,R]$. Then there would exist $R_2 > R$ such that $\nu(|x| > R_2) > 0$. Then

$$R^{2k} \geq \mathrm{m}_{2k}(\nu) \geq \int_{|x|>R_2} x^{2k} \nu(\mathrm{d}x) \geq R_2^{2k} \nu(|x| > R_2),$$

which would yield the contradiction

$$0 < \nu(|x| > R_2) \leq \left(\frac{R}{R_2}\right)^{2k} \to 0, \qquad k \to \infty.$$

Step 2: $\mu = \nu$. We have now

$$\int_{[-R,R]} f(x) \, \mu(\mathrm{d}x) = \int_{[-R,R]} f(x) \, \nu(\mathrm{d}x) \tag{A.1}$$

for any polynomial f. By Weierstrass's theorem, the above holds for any continuous function f on $[-R,R]$. Then a standard technique in probability theory shows $\mu = \nu$; for example, we first approximate every indicator function χ_I over a closed interval $I \subseteq [-R,R]$ by an increasing sequence of nonnegative continuous functions, and then by the monotone convergence theorem, (A.1) implies $\mu(I) = \nu(I)$. By the closed interval version of Proposition 4.4, we conclude $\mu = \nu$. ☐

A more general criterion is **Carleman's condition**

$$\sum_{n\geq 1} \mathrm{m}_{2n}(\mu)^{-\frac{1}{2n}} = +\infty. \tag{A.2}$$

If a probability measure μ satisfies Carleman's condition then the moment sequence of μ is determinate; see [3, p. 85] or [140, Theorem 4.3]. It is easy to see that if μ has compact support then μ satisfies Carleman's condition. The normal distribution $\mathrm{N}(m, \sigma^2)$ and the exponential distribution $\lambda e^{-x/\lambda}\chi_{(0,+\infty)}(x)\,\mathrm{d}x$, $\lambda > 0$, satisfy Carleman's condition.

Proposition A.5. *Let μ, ν be probability measures on $\mathbb{R}$. If μ has compact support and ν satisfies Carleman's condition, then $\mu \rhd \nu$ also satisfies Carleman's condition.*

Proof. From the moment formula (5.8) for monotone convolution and inequality (4.20), we have

$$\begin{aligned}
\mathrm{m}_{2n}(\mu \rhd \nu) &\leq \sum_{\ell=0}^{2n} \sum_{\substack{k_0,k_1,\dots,k_\ell\geq 0,\\ k_0+k_1+\cdots+k_\ell=2n-\ell}} |\mathrm{m}_\ell(\mu)||\mathrm{m}_{k_0}(\nu)||\mathrm{m}_{k_1}(\nu)|\cdots|\mathrm{m}_{k_\ell}(\nu)| \\
&\leq \sum_{\ell=0}^{2n} \sum_{\substack{k_0,k_1,\dots,k_\ell\geq 0,\\ k_0+k_1+\cdots+k_\ell=2n-\ell}} \mathrm{m}_{2n}(\mu)^{\frac{\ell}{2n}}\mathrm{m}_{2n}(\nu)^{\frac{k_0+k_1+\cdots+k_\ell}{2n}} \\
&= \sum_{\ell=0}^{2n} \binom{2n}{\ell} \mathrm{m}_{2n}(\mu)^{\frac{\ell}{2n}}\mathrm{m}_{2n}(\nu)^{\frac{2n-\ell}{2n}} \\
&= \left[\mathrm{m}_{2n}(\mu)^{\frac{1}{2n}} + \mathrm{m}_{2n}(\nu)^{\frac{1}{2n}}\right]^{2n}.
\end{aligned}$$

Let $a_n := \mathrm{m}_{2n}(\mu)^{\frac{1}{2n}}$ and $b_n := \mathrm{m}_{2n}(\nu)^{\frac{1}{2n}}$. Suppose that μ is supported on $[-R, R]$. Then $a_n \leq R$. Note that due to (4.20) the sequence $(b_n)_{n\geq 1}$ is nondecreasing. If $b_1 = 0$ then $\nu = \delta_0$ and the statement is obvious. If $b_1 > 0$ is then we can find $c > 0$ such that

$$\mathrm{m}_{2n}(\mu \rhd \nu)^{-\frac{1}{2n}} \geq \frac{1}{a_n + b_n} \geq \frac{1}{R + b_n} \geq \frac{c}{b_n}, \qquad n \in \mathbb{N}.$$

Since $\sum_{n\geq 1} 1/b_n = +\infty$ we have $\sum_{n\geq 1} \mathrm{m}_{2n}(\mu \rhd \nu)^{-\frac{1}{2n}} = +\infty$. ☐

Next we discuss the relation between convergence of moments and weak convergence.

Lemma A.6. *Let $\mathscr{P}$ be a family of probability measures on $\mathbb{R}$. Let $f\colon \mathbb{R} \to [0, +\infty)$ be a measurable function such that $\lim_{|x|\to\infty} f(x) = +\infty$ and*

$$\sup_{\mu\in\mathscr{P}} \int_{\mathbb{R}} f(x)\,\mu(\mathrm{d}x) < +\infty. \tag{A.3}$$

Then $\mathscr{P}$ is tight.

Proof. Let $C \geq 0$ be the finite value in (A.3). For any $\varepsilon > 0$ there exists $R > 0$ such that $\inf_{|x|>R} f(x) \geq \frac{C+1}{\varepsilon}$. For all $\mu \in \mathscr{P}$ we have

$$\begin{aligned}
\mu(\mathbb{R} \setminus [-R, R]) &= \int_{|x|>R} \mu(\mathrm{d}x) \leq \int_{|x|>R} \frac{\varepsilon}{C+1} f(x)\,\mu(\mathrm{d}x) \\
&\leq \frac{\varepsilon}{C+1}\int_{\mathbb{R}} f(x)\,\mu(\mathrm{d}x) \leq \frac{\varepsilon C}{C+1} \leq \varepsilon.
\end{aligned}$$ ☐

Proposition A.7. *Let μ, μ_n $(n \in \mathbb{N})$ be probability measures on $\mathbb{R}$. Suppose that μ, μ_n all have finite moments of all orders, and*

$$\lim_{n\to\infty} \mathrm{m}_k(\mu_n) = \mathrm{m}_k(\mu), \qquad k \in \mathbb{N}.$$

If the moment sequence of μ is determinate, then μ_n converges weakly to μ.

Proof. Since $\mathrm{m}_2(\mu_n)$ converges to $\mathrm{m}_2(\mu)$, it is a bounded sequence, and hence the assumption of Lemma A.6 is satisfied for $\mathscr{P} = (\mu_n)_{n\in\mathbb{N}}$ and $f(x) = x^2$. Consequently, the sequence $(\mu_n)_{n\geq 1}$ is tight. Let $(\mu_{n(j)})_{j\geq 1}$ be a subsequence that converges weakly to a probability measure μ'.

Step 1: μ' has finite moments of all orders. For this we take a sequence of continuous functions $f_N\colon \mathbb{R} \to [0, 1]$, $N = 1, 2, 3, \dots$, such that

- f_N is supported on $[-2N, 2N]$,
- $f_N = 1$ on $[-N, N]$,
- $f_N(x) \uparrow 1$ as $N \to \infty$ at every $x \in \mathbb{R}$.

In the obvious inequality

$$\int_{\mathbb{R}} x^{2k} f_N(x)\mu_{n(j)}(\mathrm{d}x) \leq \int_{\mathbb{R}} x^{2k}\mu_{n(j)}(\mathrm{d}x) = \mathrm{m}_{2k}(\mu_{n(j)})$$

passing to the limit $j \to \infty$ yields

$$\int_{\mathbb{R}} x^{2k} f_N(x)\mu'(\mathrm{d}x) \leq \mathrm{m}_{2k}(\mu).$$

Further passing to the limit $N \to \infty$, together with the monotone convergence theorem, shows

$$\int_{\mathbb{R}} x^{2k} \mu'(\mathrm{d}x) \le \mathrm{m}_{2k}(\mu) < +\infty.$$

Thus μ' has finite moments of all orders.

Step 2: $\mathrm{m}_k(\mu') = \mathrm{m}_k(\mu)$ for all $k \in \mathbb{N}$. To show this we first observe that

$$\int_{|x|\ge N} |x|^k \mu_n(\mathrm{d}x) \le \int_{|x|\ge N} \left(\frac{|x|}{N}\right)^k |x|^k \mu_n(\mathrm{d}x) \le \frac{1}{N^k}\mathrm{m}_{2k}(\mu_n) \le C_k N^{-k},$$

where $C_k := \sup_{n\in\mathbb{N}} \mathrm{m}_{2k}(\mu_n) < +\infty$. With this inequality we obtain

$$\begin{aligned} |\mathrm{m}_k(\mu') - \mathrm{m}_k(\mu_{n(j)})| &\le \underbrace{\left|\int_{\mathbb{R}} x^k f_N(x)\mu'(\mathrm{d}x) - \int_{\mathbb{R}} x^k f_N(x)\mu_{n(j)}(\mathrm{d}x)\right|}_{=:\,\varepsilon(N,j)} \\ &\quad + \left|\int_{\mathbb{R}} x^k(1-f_N(x))\mu'(\mathrm{d}x)\right| + \left|\int_{\mathbb{R}} x^k(1-f_N(x))\mu_{n(j)}(\mathrm{d}x)\right| \\ &\le \varepsilon(N,j) + \int_{|x|\ge N} |x|^k \mu'(\mathrm{d}x) + C_k N^{-k}. \end{aligned}$$

As $\lim_{j\to\infty} \varepsilon(N,j) = 0$, taking N large enough and then j large enough, we obtain

$$\mathrm{m}_k(\mu) = \lim_{j\to\infty} \mathrm{m}_k(\mu_{n(j)}) = \mathrm{m}_k(\mu').$$

Step 3. Since the moment sequence of μ is determinate, we have $\mu = \mu'$. The whole above argument shows that any subsequence of $(\mu_n)_{n\ge1}$ has a further subsequence that converges weakly to μ. By Lemma 4.8, we conclude that the original sequence $(\mu_n)_{n\ge1}$ converges weakly to μ. □

A.1. **Notes.** Proposition A.5 is a new result. It is also natural to ask whether monotone convolution preserves Carleman's condition or determinacy of moment sequences, but the answer is unknown. Proposition A.7 is well known and can be found in the literature, e.g. in [44, Theorem 4.5.5].

Appendix B. Inverse map of Cauchy transform

We study the inverse map of (reciprocal) Cauchy transform to complete the proof of Proposition 5.13.

Lemma B.1. *Let $E \subset \mathbb{C}$ be a convex subset and $f\colon E \to \mathbb{C}$ be the restriction of a holomorphic function defined on an open set containing E such that $\Im[e^{i\alpha} f'(z)] > 0$ holds on E for some constant $\alpha \in [0, 2\pi)$. Then f is injective on E.*

Proof. By considering $g(z) := e^{i\alpha} f(z)$, we may assume from the beginning that $\alpha = 0$. For any $z_0, z_1 \in D$, the following identity holds:

$$\begin{aligned} f(z_1) - f(z_0) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} f((1-t)z_0 + tz_1)\,\mathrm{d}t \\ &= (z_1 - z_0)\int_0^1 f'((1-t)z_0 + tz_1)\,\mathrm{d}t. \end{aligned}$$

Since the number $C(z_0, z_1) = \inf\{\Im[f'((1-t)z_0 + tz_1)] : t \in [0,1]\}$ is positive, the inequality

$$|f(z_1) - f(z_0)| \ge C(z_0, z_1)|z_1 - z_0|$$

implies the injectivity of f. □

For a concise statement, we consider the domain

$$\nabla_{\gamma,\delta} := \nabla_\gamma \cap \{z : \Im(z) > \delta\} = \{z \in \mathbb{C}^+ : \Im(z) > \max\{\gamma|\Re(z)|, \delta\}\}, \qquad \gamma, \delta > 0.$$

Proposition B.2. *Let μ be a probability measures on $\mathbb{R}$. For every $0 < \varepsilon < \gamma < 1$ there exists $\delta_0 = \delta_0(\gamma, \varepsilon) > 0$ such that, for all $\delta \ge \delta_0$, the function F_μ is injective in $\nabla_{\gamma,\delta}$ and $\nabla_{\gamma+\varepsilon,(1+\varepsilon)\delta} \subseteq F_\mu(\nabla_{\gamma,\delta}) \subseteq \nabla_{\gamma-\varepsilon,(1-\varepsilon)\delta}$.*

Proof. Injectivity. We first establish

$$|F'_\mu(z) - 1| = o(1), \qquad z \to \infty,\ z \in \nabla_\gamma. \tag{B.1}$$

Using Proposition 4.37 (3) we get

$$F'_\mu(z) = -F_\mu(z)^2 G'_\mu(z) = (1 + o(1)) \int_{\mathbb{R}} \frac{z^2}{(z-x)^2}\,\mu(\mathrm{d}x)$$

as $z \to \infty$ within ∇_γ. Therefore, it suffices to show that

$$\limsup_{\substack{z\to\infty \\ z\in\nabla_\gamma}} \left|\int_{\mathbb{R}} \frac{z^2}{(z-x)^2}\,\mu(\mathrm{d}x) - 1\right| = 0. \tag{B.2}$$

To see this, we use the inequalities $|z^2/(z-x)^2| \le 1+\gamma^{-2}$ and $|z| \le \sqrt{1+\gamma^{-2}}\Im z$ for $z \in \nabla_\gamma$ to proceed as, for each $R>0$,

$$\begin{aligned}\int_{\mathbb{R}} \left|\frac{z^2}{(z-x)^2}-1\right| \mu(\mathrm{d}x) &\le \int_{[-R,R]} \left|\frac{z^2}{(z-x)^2}-1\right| \mu(\mathrm{d}x) + (2+\gamma^{-2})\mu(\mathbb{R}\setminus[-R,R]) \\ &\le \int_{[-R,R]} \frac{x^2+2|zx|}{(\Im z)^2}\,\mu(\mathrm{d}x) + (2+\gamma^{-2})\mu(\mathbb{R}\setminus[-R,R]) \\ &\le \frac{R^2+2R\sqrt{1+\gamma^{-2}}\Im z}{(\Im z)^2} + (2+\gamma^{-2})\mu(\mathbb{R}\setminus[-R,R]),\end{aligned}$$

and then

$$\limsup_{\substack{z\to\infty \\ z\in\nabla_\gamma}} \int_{\mathbb{R}} \left|\frac{z^2}{(z-x)^2}-1\right| \mu(\mathrm{d}x) \le (2+\gamma^{-2})\mu(\mathbb{R}\setminus[-R,R]).$$

Letting R tend to infinity here yields (B.2), and hence (B.1).

By (B.1) we can find $\delta>0$ so large that

$$\Re[F_\mu'(z)] \ge \frac{1}{2}, \qquad z \in \overline{\nabla_{\gamma,\delta}},$$

and by Lemma B.1, F_μ is injective on $\overline{\nabla_{\gamma,\delta}}$.

The inclusion $\nabla_{\gamma+\varepsilon,(1+\varepsilon)\delta} \subseteq F_\mu(\nabla_{\gamma,\delta})$. For $z\in\partial\nabla_{\gamma,\delta}$ with $\gamma|\Re z| = \Im z$, one can see that

$$d(z, \nabla_{\gamma+\varepsilon,(1+\varepsilon)\delta}) \ge \frac{\varepsilon|z|}{\sqrt{(1+\gamma^2)(1+(\gamma+\varepsilon)^2)}} =: c_1|z|.$$

Here d denotes the Euclidean distance. For $z\in\partial\nabla_{\gamma,\delta}$ with $\Im z=\delta$, we have

$$d(z, \nabla_{\gamma+\varepsilon,(1+\varepsilon)\delta}) \ge \varepsilon\delta \ge \frac{\varepsilon\gamma|z|}{\sqrt{1+\gamma^2}} =: c_2|z|.$$

Therefore, if one takes $\delta_0>0$ so that $\sup_{z\in\nabla_{\gamma,\delta_0}}|F_\mu(z)-z|/|z| < \min\{c_1,c_2\}$, then the simple closed curve $\{\,F_\mu(z) : z\in\partial\nabla_{\gamma,\delta}\cup\{\infty\}\,\} \subseteq \mathbb{C}\cup\{\infty\}$ surrounds each point of $\nabla_{\gamma+\varepsilon,(1+\varepsilon)\delta}$ exactly once as soon as $\delta\ge\delta_0$, and hence by the argument principle we obtain $F_\mu(\nabla_{\gamma,\delta}) \supseteq \nabla_{\gamma+\varepsilon,(1+\varepsilon)\delta}$ as desired. The other inclusion can be proved analogously. □

The previous proposition allows us to define an injective function $F_\mu^{-1}\colon \nabla_{\gamma+\varepsilon,(1+\varepsilon)\delta} \to \mathbb{C}^+$ which satisfies $F_\mu\circ F_\mu^{-1} = \mathrm{id}$ on $\nabla_{\gamma+\varepsilon,(1+\varepsilon)\delta}$. Fixing any $0<\varepsilon<\gamma<1$ and large $\delta>0$, we set $\gamma' := \gamma+\varepsilon$ and $\delta' := (1+\varepsilon)\delta$ for notational simplicity.

Proposition B.3. *Let μ be a probability measures on $\mathbb{R}$ and $n\in\mathbb{N}$. Let $F_\mu^{-1}\colon \nabla_{\gamma',\delta'} \to \mathbb{C}^+$ be defined as above. Then the following are equivalent.*

(1) $\displaystyle\int_{\mathbb{R}} x^{2n}\,\mu(\mathrm{d}x) < +\infty.$

(2) *There exist $c_1, c_2, \dots, c_{2n}\in\mathbb{R}$ such that*

$$F_\mu^{-1}(z) = z + c_1 + \frac{c_2}{z} + \cdots + \frac{c_{2n}}{z^{2n-1}} + o(|z|^{-(2n-1)}), \qquad z\to\infty,\ z\in\nabla_{\gamma',\delta'}. \tag{B.3}$$

Proof. (1) $\Longrightarrow$ (2). We only consider the case $n=2$, which should be enough to see how to handle the general n. Recall that, by definition, $F_\mu^{-1}(\nabla_{\gamma',\delta'})$ is contained in some $\nabla_{\gamma,\delta}$. By Proposition 4.46, for some real constants b_1,b_2,b_3,b_4 we have

$$F_\mu(w) = w - b_1 - \frac{b_2}{w} - \frac{b_3}{w^2} - \frac{b_4}{w^3} + o(|w|^{-3}), \qquad w\to\infty,\ w\in\nabla_{\gamma,\delta}. \tag{B.4}$$

We set the notation $w := F_\mu^{-1}(z)$ for each $z\in\nabla_{\gamma',\delta'}$. By the construction of the inverse function and the fact $F_\mu(w) = w+o(w)$ we can see that $w\to\infty$ whenever $z\to\infty$. Putting w into (B.4) we get

$$z = F_\mu^{-1}(z) + O(1),$$

where $O(1)$ is a function bounded on $\nabla_{\gamma',\delta'}$. In particular we get

$$w = F_\mu^{-1}(z) = z + o(z). \tag{B.5}$$

Next we substitute (B.5) into (B.4) to obtain

$$z = F_\mu^{-1}(z) - b_1 - \frac{b_2}{z+o(z)} + o(z^{-1}),$$

which amounts to

$$w = F_\mu^{-1}(z) = z + b_1 + \frac{b_2}{z} + o(z^{-1}). \tag{B.6}$$

Finally, we substitute (B.6) into (B.4) to obtain

$$z = F_\mu^{-1}(z) - b_1 - \frac{b_2}{z+b_1+b_2/z+o(z^{-1})} - \frac{b_3}{(z+b_1+o(1))^2} - \frac{b_4}{(z+o(z))^3} + o(|z|^3).$$

Using the geometric series expansion $1/(1-\zeta) = 1+\zeta+\zeta^2+\cdots$ and recollecting terms, we obtain

$$F_\mu^{-1}(z) = z + b_1 + \frac{b_2}{z} + \frac{b_3-b_1b_2}{z^2} + \frac{b_4-2b_1b_3-b_2^2+b_1^2b_2}{z^3} + o(z^3)$$

as desired.

(2) $\Longrightarrow$ (1) is very similar to the proof of (1) $\Longrightarrow$ (2). In this case we use the formula $F_\mu^{-1}(F_\mu(w)) = w$ instead, which holds on a subdomain $\nabla_{\tilde\gamma,\tilde\delta} \subseteq \nabla_{\gamma,\delta}$ selected so that $F_\mu(\nabla_{\tilde\gamma,\tilde\delta}) \subseteq \nabla_{\gamma',\delta'}$, e.g. $\tilde\gamma = \gamma' + \varepsilon$ and $\tilde\delta = \delta'/(1-\varepsilon)$. The remaining argument is exactly the same. □

Completing the proof of Proposition 5.13. It remains to prove $\int_{\mathbb{R}} t^{2n}\,\mu(\mathrm{d}t) < +\infty$, assuming that $\int_{\mathbb{R}} t^{2n}\,\lambda(\mathrm{d}t) < +\infty$ and $\int_{\mathbb{R}} t^{2n}\,\nu(\mathrm{d}t) < +\infty$, where $\lambda := \mu \rhd \nu$. Take a domain $\nabla_{\gamma',\delta'}$ on which F_ν^{-1} can be defined, and set $w = F_\nu^{-1}(\mathrm{i}y)$ for each $y > \delta'$. Then $F_\lambda(w) = F_\mu(F_\nu(w))$ yields $F_\mu(\mathrm{i}y) = F_\lambda(F_\nu^{-1}(\mathrm{i}y))$. By Propositions 4.46 and B.3, we can expand $F_\lambda(\zeta)$ and $F_\nu^{-1}(\mathrm{i}y)$ into truncated Laurent series. Following the lines for computing (5.11), we obtain a truncated Laurent series of $F_\mu(\mathrm{i}y) = F_\lambda(F_\nu^{-1}(\mathrm{i}y))$, from which we can conclude that $\int_{\mathbb{R}} t^{2n}\,\mu(\mathrm{d}t) < +\infty$. □

B.1. **Notes.** Lemma B.1 was proved by Noshiro and Warschawski [126, 155]. Proposition B.2 was proved by Bercovici and Voiculescu [32]. Proposition B.3 is due to Benaych-Georges [29, Theorem 1.3]. The coefficients $c_1, c_2, c_3, \dots$ in Proposition B.3 are called the free cumulants of μ. There are combinatorial formulas relating $(c_n)_{n\ge1}$, the moment sequence $(\mathrm{m}_n(\mu))_{n\ge1}$ and the Boolean cumulants $(b_n)_{n\ge1}$ and monotone cumulants $(\kappa_n(\mu))_{n\ge1}$; see [13] and references therein.

Appendix C. One-parameter semigroups of holomorphic self-maps

This section collects results that have been used in the proof of Theorem 5.15.

Definition C.1. Let D be an open subset of $\mathbb{C}$. A **one-parameter semigroup** of holomorphic self-maps of D is a family $(f_t)_{t\ge0}$ consisting of holomorphic functions $f_t\colon D \to D$ such that

(i) $t \mapsto f_t$ is continuous with respect to locally uniform convergence, i.e., for any sequence $(t_n)_{n\ge1} \subseteq [0,+\infty)$ converging to t, the function f_{t_n} converges to f_t locally uniformly on D,

(ii) $f_{s+t} = f_s \circ f_t$ for all $s,t \ge 0$,

(iii) $f_0 = \mathrm{id}_D$.

If D is a simply connected domain of $\mathbb{C}$ and is not equal to $\mathbb{C}$, then D is conformally equivalent to $\mathbb{C}^+$ by the Riemann mapping theorem. Then, according to Proposition 4.29, condition (i) is equivalent to the weaker condition that the map $t \mapsto f_t(z) \in D$ is continuous for every $z \in D$.

It turns out that any one-parameter semigroup is differentiable with respect to t.

Theorem C.2. *Let D be an open subset of $\mathbb{C}$ and $(f_t)_{t\ge0}$ be a one-parameter semigroup of holomorphic self-maps of D. Then the limit*

$$g := \lim_{t\to0^+} \frac{f_t - \mathrm{id}_D}{t} \tag{C.1}$$

exists locally uniformly, g is holomorphic on D, and $(\mathrm{d}/\mathrm{d}t)f_t(z) = g(f_t(z))$ holds for $(t,z) \in [0,+\infty) \times D$, where $\mathrm{d}/\mathrm{d}t$ at $t = 0$ is to be interpreted as the right-derivative. The function g is called the ***infinitesimal generator*** *of $(f_t)_{\ge0}$.*

Proof. Let $B \subseteq D$ be a compact ball of positive radius. By conditions (i) and (iii) of Definition C.1, there exist $0 < \alpha < 1$ and a compact ball $B' \subseteq D$ with the same center as B and a larger radius such that

$$\{f_t(z) : (t,z) \in [0,\alpha] \times B\} \subseteq B'.$$

Let us consider the identity

$$f_{2t}(z) - 2f_t(z) + z = \int_z^{f_t(z)} \frac{\mathrm{d}}{\mathrm{d}w}(f_t(w) - w)\,\mathrm{d}w, \qquad (t,z) \in [0,\alpha] \times B, \tag{C.2}$$

where the integration is performed over the line segment connecting z and $f_t(z)$, which is contained in B'. By Cauchy's integral formula,

$$\left|\frac{\mathrm{d}}{\mathrm{d}w}(f_t(w) - w)\right| \le C_1 \sup_{\zeta\in\partial B''} |f_t(\zeta) - \zeta|, \qquad (t,w) \in [0,\alpha] \times B', \tag{C.3}$$

where B'' is a compact ball in D with the same center as B' and a larger radius, and $C_1 > 0$ is a constant depending on the distance of B' and $\partial B''$. Selecting sufficiently small $\beta \in (0,\alpha]$, we can bound the right side of (C.3) by $1/10$ for all $0 \le t \le \beta$, and hence

$$|f_{2t}(z) - 2f_t(z) + z| \le \frac{1}{10}|f_t(z) - z|, \qquad (t,z) \in [0,\beta] \times B.$$

By the triangular inequality, this implies $(19/10)|f_t(z) - z| \le |f_{2t}(z) - z|$ and so $|f_t(z) - z| \le 2^{-2/3}|f_{2t}(z) - z|$. This further yields

$$|f_t(z) - z| \le C_2 t^{2/3}, \qquad (t,z) \in [0,1] \times B \tag{C.4}$$

for some constant $C_2 > 0$. Indeed, for $t \in (\beta,1]$ this is obvious by selecting C_2 large enough. For each $t \in (0,\beta]$ and $z \in B$ we pick $n \in \mathbb{N}$ such that $t \in (\beta 2^{-n}, \beta 2^{-n+1}]$ to get

$$|f_t(z) - z| \le 2^{-2/3}|f_{2t}(z) - z| \le \cdots \le 2^{-2n/3}|f_{2^n t}(z) - z| \le C_2 t^{2/3},$$

where $C_2 := \beta^{-2/3} \sup\{|f_s(\zeta) - \zeta| : \zeta \in B,\ s \in [\beta, 2\beta]\}$.

Inequality (C.4) also holds on $[0,1] \times B''$ with a constant C_3 possibly larger than C_2, as the ball B was arbitrary. Combining this fact and (C.2) and (C.3) yields

$$|f_{2t}(z) - 2f_t(z) + z| \le C_1 C_3 t^{2/3} |f_t(z) - z| \le C_1 C_2 C_3 t^{4/3}, \qquad (t,z) \in [0,\alpha] \times B.$$

Selecting $t = t_n := 2^{-n}$ and setting $n_0(\alpha) := \lfloor \log_2(1/\alpha) \rfloor$, we obtain

$$\left| \frac{f_{t_{n-1}}(z) - z}{t_{n-1}} - \frac{f_{t_n}(z) - z}{t_n} \right| \le \frac{C_1 C_2 C_3}{2} t_n^{1/3}, \qquad n \ge n_0(\alpha),\ z \in B.$$

Combined with Weierstrass' M-test, this estimate implies that

$$\frac{f_{t_n}(z) - z}{t_n} = f_1(z) - z + \sum_{k=1}^{n} \left(\frac{f_{t_k}(z) - z}{t_k} - \frac{f_{t_{k-1}}(z) - z}{t_{k-1}} \right)$$

converges uniformly to a function g on B. Since B was arbitrary, we conclude that $[f_{t_n}(z) - z]/t_n$ converges to a function g locally uniformly on D. By Weierstrass' theorem, g is holomorphic. In addition, for each fixed $z \in D$ and $t > 0$ the convergence

$$\frac{f_{t_n+s}(z) - f_s(z)}{t_n} = \frac{f_{t_n}(f_s(z)) - f_s(z)}{t_n} \to g(f_s(z)), \qquad n \to \infty$$

holds uniformly on $s \in [0, t]$ since $\{f_s(z) : s \in [0, t]\}$ is a compact subset of D. Combining this and Riemann sums yields, as $n \to \infty$,

$$\begin{aligned} \int_0^t g(f_s(z))\, \mathrm{d}s &= \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} \frac{1}{2^n} g(f_{k2^{-n}}(z)) + o(1) \\ &= \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} [f_{(k+1)2^{-n}}(z) - f_{k2^{-n}}(z)] + o(1) \\ &= f_{\lfloor 2^n t \rfloor 2^{-n}}(z) - z + o(1) \to f_t(z) - z. \end{aligned}$$

Therefore, we have obtained

$$f_t(z) = z + \int_0^t g(f_s(z))\, \mathrm{d}s, \qquad t \ge 0,\ z \in D,$$

which readily yields the desired statements. Indeed, $(\mathrm{d}/\mathrm{d}t) f_t(z) = g(f_t(z))$ for $t > 0$ is obvious. As for (C.1), for any compact subset $K \subseteq D$ and $t > 0$, we have

$$\begin{aligned} \sup_{z \in K} \left| \frac{f_t(z) - z}{t} - g(z) \right| &= \sup_{z \in K} \left| \frac{1}{t} \int_0^t [g(f_s(z)) - g(z)]\, \mathrm{d}s \right| \\ &\le \sup_{z \in K} \frac{1}{t} \int_0^t |g(f_s(z)) - g(z)|\, \mathrm{d}s \\ &\le \sup_{\substack{z \in K \\ s \in [0,t]}} |g(f_s(z)) - g(z)|, \end{aligned}$$

which converges to 0 as $t \to 0^+$ due to condition (i). □

The next problem is to characterize the possible functions g that appear as the infinitesimal generator of a one-parameter semigroup. We only prove the sufficiency part in the case of $D = \mathbb{C}^+$, see also Notes C.1.

Theorem C.3. *Let $\mathscr{G}_1(\mathbb{C}^+)$ be the set of Nevanlinna functions and $\mathscr{G}_2(\mathbb{C}^+)$ the set of the functions of the form $G(z) = N(z)(z + \xi)(z + \overline{\xi})$, where N is a Nevanlinna function and $\xi \in \mathbb{C}^+ \cup \mathbb{R}$. Then any function $G \in \mathscr{G}_1(\mathbb{C}^+) \cup \mathscr{G}_2(\mathbb{C}^+)$ is the infinitesimal generator of a one-parameter semigroup of holomorphic self-maps of $\mathbb{C}^+$, i.e., the equation*

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t} F_t(z) = G(F_t(z)), & z \in \mathbb{C}^+,\ t \ge 0, \\ F_0(z) = z, & z \in \mathbb{C}^+ \end{cases} \tag{C.5}$$

has a (unique) solution $(F_t)_{t \ge 0}$ that forms a one-parameter semigroup of holomorphic self-maps of $\mathbb{C}^+$.

Proof. The local existence of a unique solution follows from the standard ODE argument, i.e., for each $z \in \mathbb{C}^+$ there exists $\varepsilon_z > 0$ and a function $(-\varepsilon_z, \varepsilon_z) \ni t \mapsto F_t(z) \in \mathbb{C}^+$ such that (C.5) holds for $t \in (-\varepsilon_z, \varepsilon_z)$. We show the global existence.

Case 1: $G \in \mathscr{G}_2(\mathbb{C}^+)$ and $\xi \in \mathbb{C}^+$. Let $L_\xi \colon \mathbb{C}^+ \to \mathbb{D}$ be the conformal bijection

$$L_\xi(z) := \frac{z + \overline{\xi}}{z + \xi}. \tag{C.6}$$

Setting $f_t := L_\xi \circ F_t \circ L_\xi^{-1}$ transforms (C.5) into

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t} f_t(w) = -f_t(w) h(f_t(w)), & t \ge 0,\ w \in \mathbb{D}, \\ f_0(w) = w, & w \in \mathbb{D}, \end{cases}$$

where $h(w) := -2\mathrm{i}\Im(\xi) N(L_\xi^{-1}(w))$. As h has nonnegative real part, we see that

$$\frac{\mathrm{d}}{\mathrm{d}t} |f_t(w)|^2 = -2|f_t(w)|^2 \Re[h(f_t(w))] \le 0,$$

and hence the map $t \mapsto |f_t(w)|$ is non-increasing as long as the solution exists. This ensures that the solution does not blow up in finite time, so that a solution $f_t(w)$ exists in $\mathbb{D}$ for all $t \ge 0$. For each $t \ge 0$ the map f_t is holomorphic because of its iterative construction. Finally, $F_t := L_\xi^{-1} \circ f_t \circ L_\xi$ gives the desired solution.

Case 2: $G \in \mathscr{G}_2(\mathbb{C}^+)$ *and* $\xi \in \mathbb{R}$. Let us take a sequence $\xi^n \in \mathbb{C}^+$ that converges to $\xi \in \mathbb{R}$. According to case 1, for each $n \in \mathbb{N}$ there is a one-parameter semigroup $(F_t^n)_{t\geq 0}$ corresponding to the infinitesimal generator $G^n(z) := N(z)(z+\xi^n)(z+\overline{\xi^n})$. Let $B := \{z \in \mathbb{C} : |z - \mathrm{i}| \leq 1/2\}$. There is a constant $M > 0$ such that $|G^n(z)| \leq M$ for all $z \in B$ and $n \in \mathbb{N}$. We see that

$$|F_t^n(\mathrm{i}) - \mathrm{i}| \leq \int_0^t |G^n(F_s^n(\mathrm{i}))|\,\mathrm{d}s \leq Mt$$

as long as $F_t^n(\mathrm{i}) \in B$. A standard ODE argument indeed shows that $F_t^n(\mathrm{i}) \in B$ for all $n \in \mathbb{N}$ and $0 \leq t \leq \alpha$, where $\alpha := 1/(2M)$. This further implies that for any compact subset $K \subseteq \mathbb{C}^+$, there is a compact subset $K' \subseteq \mathbb{C}^+$ such that

$$\{F_t^n(z) : n \in \mathbb{N},\ t \in [0,\alpha],\ z \in K\} \subseteq K'. \tag{C.7}$$

Indeed, let $v_t^n := L_{-\overline{F_t^n(\mathrm{i})}} \circ F_t^n \circ (L_\mathrm{i})^{-1}$, where L_w is the map defined in (C.6). This is a holomorphic self-map of $\mathbb{D}$ fixing 0. By the Schwarz lemma, the inequality $|v_t^n(z)| \leq |z|$ holds on $\mathbb{D}$. From this inequality and the fact that $F_t^n(\mathrm{i}) \in B$, the desired claim (C.7) can be deduced. Alternatively, one could also use the fact that applying a holomorphic self-map decreases the *hyperbolic distance* of two points, see e.g. [40, Theorem 1.3.7]. Yet alternatively, one could use the Nevanlinna formula (4.9) for F_t^n, together with the fact that $F_t^n(\mathrm{i}) \in B$, to estimate $|F_t^n(z)|$ uniformly from above and $\Im[F_t^n(z)]$ from below, though it would require some additional work.

There are constants $C_1, C_2, C_3 > 0$ such that

$$|G^n(\zeta)| \leq C_1, \qquad \zeta \in K',\ n \in \mathbb{N}, \tag{C.8}$$

$$|G^n(\zeta_1) - G^n(\zeta_2)| \leq C_2|\zeta_1 - \zeta_2|, \qquad \zeta_1,\zeta_2 \in K',\ n \in \mathbb{N}, \tag{C.9}$$

$$|F_t^n(\zeta_1) - F_t^n(\zeta_2)| \leq C_3|\zeta_1 - \zeta_2|, \qquad \zeta_1,\zeta_2 \in K,\ n \in \mathbb{N},\ t \in [0,\alpha], \tag{C.10}$$

where (C.9) and (C.10) hold due to Cauchy's integral formula and the uniform boundedness of G^n and F_t^n. Combining (C.7)–(C.10) and the integrated form of $(\mathrm{d}/\mathrm{d}t)F_t^n(z) = G^n(F_t^n(z))$, for all $0 \leq s \leq t \leq \alpha$ and $z, w \in K$, we have

$$\begin{aligned}
|F_t^n(z) - F_s^n(w)| &\leq |F_t^n(z) - F_t^n(w)| + |F_t^n(w) - F_s^n(w)| \\
&\leq |z-w| + \int_0^t |G^n(F_r^n(z)) - G^n(F_r^n(w))|\,\mathrm{d}r + \int_s^t |G^n(F_r^n(w))|\,\mathrm{d}r \\
&\leq |z-w| + C_2\int_0^t |F_r^n(z) - F_r^n(w)|\,\mathrm{d}r + C_1|t-s| \\
&\leq C_1|t-s| + (1+\alpha C_2 C_3)|z-w|.
\end{aligned}$$

Therefore, the functions $(t,z) \mapsto F_t^n(z)$, $n \in \mathbb{N}$, are uniformly bounded and equicontinuous on $[0,\alpha] \times K$. By Arzéla-Ascoli's theorem, there is a subsequence $(F_t^{n(k)}(z))_{k\geq 1}$ that converges locally uniformly to a function $F_t(z)$ on $[0,\alpha]\times\mathbb{C}^+$. By Weierstrass' theorem, each F_t is holomorphic on $\mathbb{C}^+$. By the maximum principle for the harmonic function $z \mapsto \Im[F_t(z)]$ and the fact that $F_t(\mathrm{i}) \in B$, we have $F_t(\mathbb{C}^+) \subseteq \mathbb{C}^+$. In the integrated form of $(\mathrm{d}/\mathrm{d}t)F_t^n(z) = G^n(F_t^n(z))$, passing to the limit along the subsequence $n = n(k)$ yields

$$F_t(z) = z + \int_0^t G(F_s(z))\,\mathrm{d}s, \qquad 0 \leq t \leq \alpha,\ z \in \mathbb{C}^+. \tag{C.11}$$

Since $\alpha = 1/(2M) > 0$ is independent of z, we can define $F_t(z) := F_{t-\alpha}(F_\alpha(z))$ for $\alpha \leq t \leq 2\alpha$ and $z \in \mathbb{C}^+$. Straightforward reasoning shows that $(F_t)_{0\leq t\leq 2\alpha}$ is continuous in t and satisfies (C.11) for $0 \leq t \leq 2\alpha$. Repeating this procedure, we obtain a global solution $F_t(z)$ defined on $[0,+\infty) \times \mathbb{C}^+$ that satisfies (C.11) for all $t \geq 0$. A standard ODE argument shows that the family $(F_t)_{t\geq 0}$ satisfies $F_{s+t} = F_s \circ F_t$, $s,t \geq 0$.

Case 3: $G \in \mathscr{G}_1(\mathbb{C}^+)$. The function $\tilde{G}(z) := -G(1/z)z^2$ belongs to $\mathscr{G}_2(\mathbb{C}^+)$, where $G(1/z) := \overline{G(1/\overline{z})}$. From case 2 there exists a one-parameter semigroup $(\tilde{F}_t)_{t\geq 0}$ associated with $\tilde{G}$. Then we can see that $F_t(z) := 1/\tilde{F}_t(1/z)$ forms a one-parameter semigroup associated with G. Alternatively, one could approximate G by $G^n(z) := G(z)(z+n\mathrm{i})(z-n\mathrm{i})/n^2$, $n \in \mathbb{N}$; then the arguments in case 2 would work without changes. □

C.1. **Notes.** Theorems C.2 and C.3 are due to Berkson and Porta [33, Theorems 1.1 and 2.6]. In the original paper, it is also shown that $\mathscr{G}_1(\mathbb{C}^+) \cap \mathscr{G}_2(\mathbb{C}^+) = \{0\}$ and the set of the infinitesimal generators of one-parameter semigroups is exactly $\mathscr{G}_1(\mathbb{C}^+) \cup \mathscr{G}_2(\mathbb{C}^+)$. Another proof is given in [40, Theorems 10.1.4 and 10.1.10] in the setting of the unit disk.

Appendix D. Cauchy's functional equation

We give a short proof of the well known fact that a measurable solution to Cauchy's functional equation is linear. The following proof is due to Alexiewicz and Orlicz [5].

Proposition D.1. *Let* $f \colon [0,+\infty) \to \mathbb{R}$ *be Borel measurable such that* $f(x+y) = f(x) + f(y)$ *holds for all* $x, y \geq 0$. *Then* $f(x) = f(1)x$ *for all* $x \geq 0$.

Proof. Since $f(0) = f(0) + f(0)$, we have $f(0) = 0$. We fix $x > 0$ and define

$$g(t) := f(t) - \frac{f(x)}{x}t, \qquad h(t) := \frac{1}{1+|g(t)|}, \qquad t \geq 0.$$

We see that $g(s+t) = g(s) + g(t)$ for all $s, t \geq 0$, and $g(x) = 0$. This implies $g(t+x) = g(t)$, so that g and hence h has period x. Then we get

$$\begin{aligned} 0 &= \int_0^x h(t)\,\mathrm{d}t - \int_0^x h(2t)\,\mathrm{d}t \\ &= \int_0^x \frac{1}{1+|g(t)|}\,\mathrm{d}t - \int_0^x \frac{1}{1+2|g(t)|}\,\mathrm{d}t \\ &= \int_0^x \frac{|g(t)|}{(1+|g(t)|)(1+2|g(t)|)}\,\mathrm{d}t, \end{aligned}$$

and so $g(t) = 0$ a.e. on $[0, x]$. For $x = 1$ this means $f(t) = f(1)t$ a.e. on $[0, 1]$. Therefore, for any $x > 0$ there exists $t_x \in (0, x] \cap (0, 1]$ such that $f(t_x) = f(1)t_x$ and $g(t_x) = 0$. This implies $f(x) = [f(t_x)/t_x]x = f(1)x$. □

References

[1] Luigi Accardi, Anis Ben Ghorbal, and Nobuaki Obata. Monotone independence, comb graphs and Bose-Einstein condensation. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 7(3):419–435, 2004.

[2] Luigi Accardi, Romuald Lenczewski, and Rafał Sałapata. Decompositions of the free product of graphs. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 10(3):303–334, 2007.

[3] N. I. Akhiezer. *The classical moment problem and some related questions in analysis*. Hafner Publishing Co., New York, 1965. Translated by N. Kemmer.

[4] N. I. Akhiezer and I. M. Glazman. *Theory of linear operators in Hilbert space*. Dover Publications, Inc., New York, 1993. Translated from the Russian and with a preface by Merlynd Nestell, Reprint of the 1961 and 1963 translations, Two volumes bound as one.

[5] A. Alexiewicz and W. Orlicz. Remarque sur l'équation fonctionelle $f(x + y) = f(x) + f(y)$. *Fund. Math.*, 33:314–315, 1945.

[6] Michael Anshelevich and Octavio Arizmendi. The exponential map in non-commutative probability. *Int. Math. Res. Not. IMRN*, 2017(17):5302–5342, 2017.

[7] Michael Anshelevich and John D. Williams. Limit theorems for monotonic convolution and the Chernoff product formula. *Int. Math. Res. Not. IMRN*, 2014(11):2990–3021, 2014.

[8] Octavio Arizmendi and Adrián Celestino. Polynomial with cyclic monotone elements with applications to random matrices with discrete spectrum. *Random Matrices Theory Appl.*, 10(2):Paper No. 2150020, 19, 2021.

[9] Octavio Arizmendi and Adrian Celestino. Monotone cumulant-moment formula and Schröder trees. *SIGMA Symmetry Integrability Geom. Methods Appl.*, 18:Paper No. 073, 22, 2022.

[10] Octavio Arizmendi and Takahiro Hasebe. Classical scale mixtures of Boolean stable laws. *Trans. Amer. Math. Soc.*, 368(7):4873–4905, 2016.

[11] Octavio Arizmendi, Takahiro Hasebe, and Yu Kitagawa. Free multiplicative convolution with an arbitrary measure on the real line. ArXiv preprint, https://arxiv.org/pdf/2503.14992, 2025.

[12] Octavio Arizmendi, Takahiro Hasebe, and Franz Lehner. Cyclic independence: Boolean and monotone. *Algebr. Comb.*, 6(6):1697–1734, 2023.

[13] Octavio Arizmendi, Takahiro Hasebe, Franz Lehner, and Carlos Vargas. Relations between cumulants in noncommutative probability. *Adv. Math.*, 282:56–92, 2015.

[14] Octavio Arizmendi, Saul Rogelio Mendoza, and Josue Vazquez-Becerra. BMT independence. *J. Funct. Anal.*, 288(2):Paper No. 110712, 40, 2025.

[15] Octavio Arizmendi, Mauricio Salazar, and Jiun-Chau Wang. Berry-Esseen type estimate and return sequence for parabolic iteration in the upper half-plane. *Int. Math. Res. Not. IMRN*, 2021(23):18037–18056, 2021.

[16] Albert Baernstein II, David Drasin, Peter Duren, and Albert Marden, editors. *The Bieberbach conjecture*, volume 21 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1986.

[17] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.

[18] Marwa Banna and Pei-Lun Tseng. The distribution of polynomials in monotone-independent elements. *Random Matrices Theory Appl.*, 13(4):Paper No. 2450016, 21, 2024.

[19] Robert O. Bauer. Chordal Loewner families and univalent Cauchy transforms. *J. Math. Anal. Appl.*, 302(2):484–501, 2005.

[20] S. T. Belinschi and H. Bercovici. Partially defined semigroups relative to multiplicative free convolution. *Int. Math. Res. Not.*, 2005(2):65–101, 2005.

[21] S. T. Belinschi and H. Bercovici. A new approach to subordination results in free probability. *J. Anal. Math.*, 101:357–365, 2007.

[22] Serban T. Belinschi, Hari Bercovici, Mireille Capitaine, and Maxime Février. Outliers in the spectrum of large deformed unitarily invariant models. *Ann. Probab.*, 45(6A):3571–3625, 2017.

[23] Serban Teodor Belinschi. *Complex analysis methods in noncommutative probability*. ProQuest LLC, Ann Arbor, MI, 2005. Thesis (Ph.D.)–Indiana University.

[24] Serban Teodor Belinschi. The Lebesgue decomposition of the free additive convolution of two probability distributions. *Probab. Theory Related Fields*, 142(1-2):125–150, 2008.

[25] Alexander C. R. Belton. A note on vacuum-adapted semimartingales and monotone independence. In *Quantum probability and infinite dimensional analysis*, volume 18 of *QP–PQ: Quantum Probab. White Noise Anal.*, pages 105–114. World Sci. Publ., Hackensack, NJ, 2005.

[26] Alexander C. R. Belton. The monotone Poisson process. In *Quantum probability*, volume 73 of *Banach Center Publ.*, pages 99–115. Polish Acad. Sci. Inst. Math., Warsaw, 2006.

[27] Alexander C. R. Belton. On the path structure of a semimartingale arising from monotone probability theory. *Ann. Inst. Henri Poincaré Probab. Stat.*, 44(2):258–279, 2008.

[28] Anis Ben Ghorbal and Michael Schürmann. Non-commutative notions of stochastic independence. *Math. Proc. Cambridge Philos. Soc.*, 133(3):531–561, 2002.

[29] Florent Benaych-Georges. Taylor expansions of R-transforms: application to supports and moments. *Indiana Univ. Math. J.*, 55(2):465–481, 2006.

[30] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.*, 227(1):494–521, 2011.

[31] Hari Bercovici. Multiplicative monotonic convolution. *Illinois J. Math.*, 49(3):929–951, 2005.

[32] Hari Bercovici and Dan Voiculescu. Free convolution of measures with unbounded support. *Indiana Univ. Math. J.*, 42(3):733–773, 1993.

[33] Earl Berkson and Horacio Porta. Semigroups of analytic functions and composition operators. *Michigan Math. J.*, 25(1):101–115, 1978.

[34] Rajendra Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[35] Philippe Biane. Processes with free increments. *Math. Z.*, 227(1):143–174, 1998.

[36] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[37] Alexei Borodin, Andrei Okounkov, and Grigori Olshanski. Asymptotics of Plancherel measures for symmetric groups. *J. Amer. Math. Soc.*, 13(3):481–515, 2000.

[38] Marek Bożejko and Roland Speicher. An example of a generalized Brownian motion. *Comm. Math. Phys.*, 137(3):519–531, 1991.

[39] Filippo Bracci, Manuel D. Contreras, and Santiago Díaz-Madrigal. Evolution families and the Loewner equation I: the unit disc. *J. Reine Angew. Math.*, 672:1–37, 2012.

[40] Filippo Bracci, Manuel D. Contreras, and Santiago Díaz-Madrigal. *Continuous semigroups of holomorphic self-maps of the unit disc*. Springer Monographs in Mathematics. Springer, Cham, 2020.

[41] Guillaume Cébron, Antoine Dahlqvist, and Franck Gabriel. Freeness of type B and conditional freeness for random matrices. *Indiana Univ. Math. J.*, 73(3):1207–1252, 2024.

[42] Guillaume Cébron, Antoine Dahlqvist, and Camille Male. Traffic distributions and independence II: universal constructions for traffic spaces. *Doc. Math.*, 29(1):39–114, 2024.

[43] Guillaume Cébron and Nicolas Gilliers. Asymptotic cyclic-conditional freeness of random matrices. *Random Matrices Theory Appl.*, 13(1):Paper No. 2350014, 48, 2024.

[44] Kai Lai Chung. *A course in probability theory*. Academic Press, Inc., San Diego, CA, third edition, 2001.

[45] E. F. Collingwood and A. J. Lohwater. *The theory of cluster sets*, volume No. 56 of *Cambridge Tracts in Mathematics and Mathematical Physics*. Cambridge University Press, Cambridge, 1966.

[46] Benoit Collins, Takahiro Hasebe, and Noriyoshi Sakuma. Free probability for purely discrete eigenvalues of random matrices. *J. Math. Soc. Japan*, 70(3):1111–1150, 2018.

[47] Benoît Collins, Felix Leid, and Noriyoshi Sakuma. Matrix models for cyclic monotone and monotone independences. *Electron. Commun. Probab.*, 29:Paper No. 58, 14, 2024.

[48] Benoît Collins and Ion Nechita. Random matrix techniques in quantum information theory. *J. Math. Phys.*, 57(1):015215, 34, 2016.

[49] Benoît Collins and Piotr Śniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Comm. Math. Phys.*, 264(3):773–795, 2006.

[50] Manuel D. Contreras, Santiago Díaz-Madrigal, and Pavel Gumenyuk. Loewner chains in the unit disk. *Rev. Mat. Iberoam.*, 26(3):975–1012, 2010.

[51] John B. Conway. *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990.

[52] Romain Couillet and Mérouane Debbah. *Random matrix methods for wireless communications*. Cambridge University Press, Cambridge, 2011.

[53] Vitonofrio Crismale, Simone Del Vecchio, and Stefano Rossi. On the monotone C^*-algebra. *Linear Algebra Appl.*, 659:33–41, 2023.

[54] Vitonofrio Crismale, Simone Del Vecchio, Stefano Rossi, and Janusz Wysoczański. Weakly-monotone C^*-algebras as Exel-Laca algebras. *Ann. Mat. Pura Appl. (4)*, 204(3):1075–1094, 2025.

[55] Vitonofrio Crismale, Francesco Fidaleo, and Yun Gang Lu. Ergodic theorems in quantum probability: an application to monotone stochastic processes. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 17(1):113–141, 2017.

[56] Vitonofrio Crismale, Maria Elena Griseta, and Janusz Wysoczański. Weakly monotone Fock space and monotone convolution of the Wigner law. *J. Theoret. Probab.*, 33(1):268–294, 2020.

[57] Vitonofrio Crismale, Maria Elena Griseta, and Janusz Wysoczański. Distributions for nonsymmetric monotone and weakly monotone position operators. *Complex Anal. Oper. Theory*, 15(6):Paper No. 101, 26, 2021.

[58] M. De. Giosa and Y. G. Lu. The free creation and annihilation operators as the central limit of the quantum Bernoulli process. *Random Oper. Stochastic Equations*, 5(3):227–236, 1997.

[59] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[60] Kurusch Ebrahimi-Fard and Frédéric Patras. Monotone, free, and boolean cumulants: a shuffle algebra approach. *Adv. Math.*, 328:112–132, 2018.

[61] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.*, 37(3):815–852, 2009.

[62] Lawrence C. Evans and Ronald F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[63] Michele Fava, Jorge Kurchan, and Silvia Pappalardi. Designs via free probability. *Phys. Rev. X*, 15:011031, Feb 2025.

[64] P. J. Forrester. *Log-gases and random matrices*, volume 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010.

[65] Peter J. Forrester. Rank 1 perturbations in random matrix theory—a review of exact results. *Random Matrices Theory Appl.*, 12(4):Paper No. 2330001, 48, 2023.

[66] Uwe Franz. Monotone independence is associative. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 4(3):401–407, 2001.

[67] Uwe Franz. Unification of Boolean, monotone, anti-monotone, and tensor independence and Lévy processes. *Math. Z.*, 243(4):779–816, 2003.

[68] Uwe Franz. Multiplicative monotone convolutions. In *Quantum probability*, volume 73 of *Banach Center Publ.*, pages 153–166. Polish Acad. Sci. Inst. Math., Warsaw, 2006.

[69] Uwe Franz. Monotone and Boolean convolutions for non-compactly supported probability measures. *Indiana Univ. Math. J.*, 58(3):1151–1185, 2009.

[70] Uwe Franz, Takahiro Hasebe, and Sebastian Schleißinger. Monotone increment processes, classical Markov processes, and Loewner chains. *Dissertationes Math.*, 552:119, 2020.

[71] Katsunori Fujie and Takahiro Hasebe. Free probability of type B prime. *Trans. Amer. Math. Soc.*, 378(8):5551–5577, 2025.

[72] Jorge Garza-Vargas and Archit Kulkarni. Spectra of infinite graphs via freeness with amalgamation. *Canad. J. Math.*, 75(5):1633–1684, 2023.

[73] Malte Gerhold. Schoenberg correspondence for multifaced independence. *Journal of Functional Analysis*, 290(1):111212, 2026.

[74] Malte Gerhold, Takahiro Hasebe, and Michaël Ulrich. Towards a classification of multi-faced independence: a representation-theoretic approach. *J. Funct. Anal.*, 285(3):Paper No. 109907, 66, 2023.

[75] Malte Gerhold and Stephanie Lachs. Classification and GNS-construction for general universal products. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 18(1):1550004, 29 pages, 2015.

[76] B. V. Gnedenko and A. N. Kolmogorov. *Limit distributions for sums of independent random variables*. Addison-Wesley Publishing Co., Inc., Cambridge, MA, 1954. Translated and annotated by K. L. Chung. With an Appendix by J. L. Doob.

[77] V. V. Goryainov and I. Ba. Semigroup of conformal mappings of the upper half-plane into itself with hydrodynamic normalization at infinity. *Ukraïn. Mat. Zh.*, 44(10):1320–1329, 1992.

[78] Tarek Hamdi. Monotone and boolean unitary Brownian motions. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 18(2):1550012, 19, 2015.

[79] Takahiro Hasebe. On monotone convolution and monotone infinite divisibility. A revision of master's thesis at Kyoto University, 2009. ArXiv preprint, https://arxiv.org/pdf/1002.3430.

[80] Takahiro Hasebe. Monotone convolution semigroups. *Studia Math.*, 200(2):175–199, 2010.

[81] Takahiro Hasebe. A three-state independence in non-commutative probability. ArXiv preprint, https://arxiv.org/pdf/1009.1505, 2010, revised in 2022.

[82] Takahiro Hasebe. Conditionally monotone independence I: Independence, additive convolutions and related convolutions. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 14(3):465–516, 2011.

[83] Takahiro Hasebe and Ikkei Hotta. Additive processes on the unit circle and Loewner chains. *Int. Math. Res. Not. IMRN*, 2022(22):17797–17848, 2022.

[84] Takahiro Hasebe, Ikkei Hotta, and Takuya Murayama. Additive processes on the real line and Loewner chains. ArXiv preprint https://arxiv.org/pdf/2412.18742, 2024.

[85] Takahiro Hasebe and Franz Lehner. Cumulants, spreadability and the Campbell-Baker-Hausdorff series. *Doc. Math.*, 28(3):515–601, 2023.

[86] Takahiro Hasebe and Hayato Saigo. Joint cumulants for natural independence. *Electron. Commun. Probab.*, 16:491–506, 2011.

[87] Takahiro Hasebe and Hayato Saigo. The monotone cumulants. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(4):1160–1170, 2011.

[88] Takahiro Hasebe and Noriyoshi Sakuma. Unimodality of Boolean and monotone stable distributions. *Demonstr. Math.*, 48(3):424–439, 2015.

[89] Fumio Hiai and Dénes Petz. *The semicircle law, free random variables and entropy*, volume 77 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2000.

[90] Akihito Hora and Nobuaki Obata. *Quantum probability and spectral analysis of graphs*. Theoretical and Mathematical Physics. Springer, Berlin, 2007. With a foreword by Luigi Accardi.

[91] R. L. Hudson and K. R. Parthasarathy. Quantum Ito's formula and stochastic evolutions. *Comm. Math. Phys.*, 93(3):301–323, 1984.

[92] David Jekel. Operator-valued chordal Loewner chains and non-commutative probability. *J. Funct. Anal.*, 278(10):108452, 100, 2020.

[93] David Jekel and Weihua Liu. An operad of non-commutative independences defined by trees. *Dissertationes Math.*, 553:100, 2020.

[94] Hong Chang Ji. Regularity properties of free multiplicative convolution on the positive line. *Int. Math. Res. Not. IMRN*, 2021(6):4522–4563, 2021.

[95] Kurt Johansson. Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. of Math. (2)*, 153(1):259–296, 2001.

[96] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[97] Achim Klenke. *Probability theory*. Universitext. Springer, London, second edition, 2014. Translation from the German edition.

[98] Wolfram Koepf. Bieberbach's conjecture, the de Branges and Weinstein functions and the Askey-Gasper inequality. *Ramanujan J.*, 13(1-3):103–129, 2007.

[99] Manish Kumar and Simeng Wang. Fullness of q-Araki-Woods factors. *J. Lond. Math. Soc. (2)*, 110(4):Paper No. e12989, 25, 2024.

[100] Burkhard Kümmerer and Roland Speicher. Stochastic integration on the Cuntz algebra O_∞. *J. Funct. Anal.*, 103(2):372–408, 1992.

[101] Stephanie Lachs. *A new family of universal products and aspects of a non-positive quantum probability theory*. PhD thesis, EMAU Greifswald, 2015. http://ub-ed.ub.uni-greifswald.de/opus/volltexte/2015/2242/.

[102] Gregory F. Lawler. *Conformally invariant processes in the plane*, volume 114 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2005.

[103] Romuald Lenczewski. Operators related to subordination for free multiplicative convolutions. *Indiana Univ. Math. J.*, 57(3):1055–1103, 2008.

[104] Romuald Lenczewski. Asymptotic properties of random matrices and pseudomatrices. *Adv. Math.*, 228(4):2403–2440, 2011.

[105] Romuald Lenczewski. Limit distributions of random matrices. *Adv. Math.*, 263:253–320, 2014.

[106] Romuald Lenczewski. Conditionally monotone independence and the associated products of graphs. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 22(4):1950023, 24, 2019.

[107] Gérard Letac and Dhafer Malouche. The Markov chain associated to a Pick function. *Probab. Theory Related Fields*, 118(4):439–454, 2000.

[108] Y. G. Lu. An interacting free Fock space and the arcsine law. *Probab. Math. Statist.*, 17(1):149–166, 1997.

[109] Hans Maassen. Addition of freely independent random variables. *J. Funct. Anal.*, 106(2):409–438, 1992.

[110] Francesco Maggi. *Sets of finite perimeter and geometric variational problems*, volume 135 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2012. An introduction to geometric measure theory.

[111] Sarah Manzel and Michael Schürmann. Non-commutative stochastic independence and cumulants. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 20(2):1750010, 38, 2017.

[112] Madan Lal Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.

[113] James A. Mingo and Roland 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.

[114] James A. Mingo and Pei-Lun Tseng. Infinitesimal operators and the distribution of anticommutators and commutators. *J. Funct. Anal.*, 287(9):Paper No. 110591, 35, 2024.

[115] Akihiro Miyagawa and Roland Speicher. A dual and conjugate system for q-Gaussians for all q. *Adv. Math.*, 413:Paper No. 108834, 36, 2023.

[116] Wojciech Młotkowski. Fuss-Catalan numbers in noncommutative probability. *Doc. Math.*, 15:939–955, 2010.

[117] Wojciech Młotkowski. Probability distributions with rational free R-transform. ArXiv preprint, https://arxiv.org/pdf/2111.10150, 2021.

[118] Naofumi Muraki. A new example of noncommutative "de Moivre-Laplace theorem". In *Probability theory and mathematical statistics (Tokyo, 1995)*, pages 353–362. World Sci. Publ., River Edge, NJ, 1996.

[119] Naofumi Muraki. Noncommutative Brownian motion in monotone Fock space. *Comm. Math. Phys.*, 183(3):557–570, 1997.

[120] Naofumi Muraki. Monotonic convolution and monotonic Lévy-Hinčin formula. Preprint, 2000.

[121] Naofumi Muraki. Monotonic independence, monotonic central limit theorem and monotonic law of small numbers. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 4(1):39–58, 2001.

[122] Naofumi Muraki. The five independences as quasi-universal products. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 5(1):113–134, 2002.

[123] Naofumi Muraki. The five independences as natural products. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 6(3):337–371, 2003.

[124] Naofumi Muraki. A simple proof of the classification theorem for positive natural products. *Probab. Math. Statist.*, 33(2):315–326, 2013.

[125] Alexandru Nica and Roland Speicher. *Lectures on the combinatorics of free probability*, volume 335 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2006.

[126] Kiyoshi Noshiro. On the theory of schlicht functions. *J. Fac. Sci. Hokkaido Imp. Univ. Ser. I Math.*, 2(3):129–155, 1934.

[127] Nobuaki Obata. *Spectral analysis of growing graphs*, volume 20 of *SpringerBriefs in Mathematical Physics*. Springer, Singapore, 2017. A quantum probability point of view.

[128] Andrei Okounkov. Random matrices and random permutations. *Internat. Math. Res. Notices*, 2000(20):1043–1095, 2000.

[129] K. R. Parthasarathy. *An introduction to quantum stochastic calculus*, volume 85 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1992.

[130] S. Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields*, 134(1):127–173, 2006.

[131] Jeffrey Pennington, Samuel Schoenholz, and Surya Ganguli. The emergence of spectral universality in deep networks. In Amos Storkey and Fernando Perez-Cruz, editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1924–1932. PMLR, 09–11 Apr 2018.

[132] Mihai Popa. A combinatorial approach to monotonic independence over a C^*-algebra. *Pacific J. Math.*, 237(2):299–325, 2008.

[133] Reinhold Remmert. *Classical topics in complex function theory*, volume 172 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998. Translated from the German by Leslie Kay.

[134] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[135] Hayato Saigo. A simple proof for monotone CLT. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 13(2):339–343, 2010.

[136] Ken-iti Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, revised edition, 2013. Translated from the 1990 Japanese original.

[137] Sebastian Schleißinger. Loewner's differential equation and spidernets. *Complex Anal. Oper. Theory*, 13(8):3899–3921, 2019.

[138] Sebastian Schleissinger. *Univalent functions in quantum probability theory*, volume 294 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2025.

[139] Konrad Schmüdgen. *Unbounded self-adjoint operators on Hilbert space*, volume 265 of *Graduate Texts in Mathematics*. Springer, Dordrecht, 2012.

[140] Konrad Schmüdgen. *The moment problem*, volume 277 of *Graduate Texts in Mathematics*. Springer, Cham, 2017.

[141] D. Shlyakhtenko. Free probability of type-B and asymptotics of finite-rank perturbations of random matrices. *Indiana Univ. Math. J.*, 67(2):971–991, 2018.

[142] Barry Simon. Spectral analysis of rank one perturbations and applications. In *Mathematical quantum theory. II. Schrödinger operators (Vancouver, BC, 1993)*, volume 8 of *CRM Proc. Lecture Notes*, pages 109–149. Amer. Math. Soc., Providence, RI, 1995.

[143] Barry Simon and Tom Wolff. Singular continuous spectrum under rank one perturbations and localization for random Hamiltonians. *Comm. Pure Appl. Math.*, 39(1):75–90, 1986.

[144] Paul Skoufranis. Independences and partial R-transforms in bi-free probability. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(3):1437–1473, 2016.

[145] Roland Speicher. Multiplicative functions on the lattice of non-crossing partitions and free convolution. *Mathematische Annalen*, 298(1):611–628, 1994.

[146] Roland Speicher. On universal products. In *Free probability theory (Waterloo, ON, 1995)*, volume 12 of *Fields Inst. Commun.*, pages 257–266. Amer. Math. Soc., Providence, RI, 1997.

[147] Roland Speicher and Reza Woroudi. Boolean convolution. In *Free probability theory (Waterloo, ON, 1995)*, volume 12 of *Fields Inst. Commun.*, pages 267–279. Amer. Math. Soc., Providence, RI, 1997.

[148] Terence Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[149] D. V. Voiculescu, K. J. Dykema, and A. Nica. *Free random variables*, volume 1 of *CRM Monograph Series*. American Mathematical Society, Providence, RI, 1992. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.

[150] Dan Voiculescu. Symmetries of some reduced free product C^*-algebras. In *Operator algebras and their connections with topology and ergodic theory (Buşteni, 1983)*, volume 1132 of *Lecture Notes in Math.*, pages 556–588. Springer, Berlin, 1985.

[151] Dan Voiculescu. Limit laws for random matrices and free products. *Invent. Math.*, 104(1):201–220, 1991.

[152] Jiun-Chau Wang. Strict limit types for monotone convolution. *J. Funct. Anal.*, 262(1):35–58, 2012.

[153] Jiun-Chau Wang. The central limit theorem for monotone convolution with applications to free Lévy processes and infinite ergodic theory. *Indiana Univ. Math. J.*, 63(2):303–327, 2014.

[154] Jiun-Chau Wang and Enzo Wendler. Law of large numbers for monotone convolution. *Probab. Math. Statist.*, 33(2):225–231, 2013.

[155] Stefan E. Warschawski. On the higher derivatives at the boundary in conformal mapping. *Trans. Amer. Math. Soc.*, 38(2):310–340, 1935.

[156] Janusz Wysoczański. bm-independence and central limit theorems associated with symmetric cones. In *Noncommutative harmonic analysis with applications to probability*, volume 78 of *Banach Center Publ.*, pages 315–320. Polish Acad. Sci. Inst. Math., Warsaw, 2007.

[157] Janusz Wysoczański. Monotonic independence associated with partially ordered sets. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 10(1):17–41, 2007.

[158] Janusz Wysoczański. bm-independence and bm-central limit theorems associated with symmetric cones. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 13(3):461–488, 2010.