UNIVERSITY OF CALIFORNIA

Los Angeles

Evolution equations in non-commutative probability

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Mathematics

by

David Andrew Jekel

2020

ABSTRACT OF THE DISSERTATION

Evolution equations in non-commutative probability

by

David Andrew Jekel
Doctor of Philosophy in Mathematics
University of California, Los Angeles, 2020
Professor Dimitri Shlyakhtenko, Chair

The thesis presents two applications of evolution equations for non-commutative variables to the theory of non-commutative probability and von Neumann algebras.

In the first part, non-commutative processes $(X_t)_{t \in [0,T]}$ with boolean, free, monotone, or anti-monotone independent increments, under certain continuity and boundedness assumptions, are classified in terms of certain evolution equations for their F-transforms $F_{X_t}(z) = (E[(z - X_t)^{-1}])^{-1}$. This classification is done in the setting of operator-valued non-commutative probability, in which the expectation takes values in a C*-algebra $\mathcal{B}$ rather than $\mathbb{C}$. Thus, the F-transform is a function of an operator variable z from (matrices over) $\mathcal{B}$, and it is understood through the theory of fully matricial or non-commutative functions, an operator-valued analogue of complex analysis. The classification of these processes generalizes previous work on the Lévy-Hinčin formula for processes with independent *and stationary* increments, and it leads to Bercovici-Pata-type bijections between the processes with independent increments for the four different types of independence. We also describe a canonical model for each process with independent increments using operators on a Fock space. In fact, the interaction between operator models and analytic function theory is a major theme of the first part, and leads to a new "coupling" technique to prove estimates

for the non-commutative central limit theorem and for Loewner chains.

In the second part, we strengthen the probabilistic, information-theoretic, and transport-theoretic connections between asymptotic random matrix theory and tracial W*-algebras through the study of functions and differential equations for several non-commuting self-adjoint variables. We consider a random variable $X^{(n)}$ in $M_n(\mathbb{C})_{\text{sa}}^d$ given by a probability distribution

$$d\mu^{(n)}(x) = \frac{1}{\int e^{-n^2 V^{(n)}}} e^{-n^2 V^{(n)}(x)}\, dx,$$

where $V^{(n)} : M_n(\mathbb{C})_{sa}^d \to \mathbb{R}$ is uniformly convex and semi-concave. We assume that $(\nabla V^{(n)})_{n \in \mathbb{N}}$ is asymptotically approximable by trace polynomials, which means that $\nabla V^{(n)}$ behaves asymptotically like some element f from a certain space of "functions of d self-adjoint variables from a tracial W*-algebra."

Then we show first that $X^{(n)}$ almost surely converges in non-commutative law to some d-tuple X from a tracial W*-algebra $(\mathcal{M}, \tau)$, meaning that $(1/n)\operatorname{Tr}(p(X^{(n)})) \to \tau(p(X))$ almost surely for every non-commutative polynomial p (which is comparable to earlier known results). The strategy to prove convergence of the expectation $E[(1/n)\operatorname{Tr}(p(X^{(n)}))]$ is to show that the heat semigroup associated to the measure $\mu^{(n)}$ preserves asymptotic approximability by trace polynomials. The same method leads to a new conditional version of this result, which shows that if $k < d$ and if $(f^{(n)})$ is asymptotically approximable by polynomials, then so is the function $g^{(n)}$ given by $g^{(n)}(X_1^{(n)}, \dots, X_k^{(n)}) = E[f^{(n)}(X^{(n)}) | X_1^{(n)}, \dots, X_k^{(n)}]$. Understanding the large-n behavior of such conditional expectations is a key step in showing our second main result that the classical entropy of $X^{(n)}$, after renormalization, converges to Voiculescu's non-microstates free entropy $\chi^*(X)$ (and an analogous result for conditional entropy given $X_1^{(n)}$, …, $X_k^{(n)}$). In particular, we obtain a new proof of the result from a preprint of Dabrowski [Dab17] that $\chi^*(X)$ agrees with the microstates free entropy $\chi(X)$ for any X that arises from such random matrix models.

The final main result studies the large-n behavior of certain functions $F^{(n)}$ that transport the measure $\mu^{(n)}$ to the distribution $\sigma_1^{(n)}$ of a standard Gaussian self-adjoint d-tuple $Z^{(n)}$. The transport map $F^{(n)}$ is obtained by the same construction as in Otto and Villani's famous proof of the Talagrand inequality based on heat semigroups and transport equations [OV00]. Using successive conditioning, we can obtain a transport function $F^{(n)}$ that is lower-triangular in the sense that

$$F^{(n)}(x_1, \dots, x_d) = (F_1^{(n)}(x_1), F_2^{(n)}(x_1, x_2), \dots, F_d^{(n)}(x_1, \dots, x_d)),$$

where $x = (x_1, \dots, x_d) \in M_n(\mathbb{C})^d$. We show that $F^{(n)}$ is asymptotically approximable by trace polynomials as $n \to \infty$, and consequently, in the large-n limit, we obtain an isomorphism $\mathrm{W}^*(X_1, \dots, X_d) \to \mathrm{W}^*(Z_1, \dots, Z_d)$ that maps $\mathrm{W}^*(X_1, \dots, X_k)$ to $\mathrm{W}^*(Z_1, \dots, Z_k)$ for every $k = 1, \dots, d$. As an application, we show that this statement holds when X itself is given by $X_j = Z_j + \delta p_j(Z)$ where Z is a free semicircular d-tuple, p_j is a self-adjoint non-commutative polynomial, and δ is sufficiently small, depending on p_1, …, p_d.

The dissertation of David Andrew Jekel is approved.

Terence Tao

Sorin Popa

Mario Bonk

Dimitri Shlyakhtenko, Committee Chair

University of California, Los Angeles

2020

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank all my teachers and mentors in mathematics for bringing me to this point, in particular: James Morrow for introducing me to analysis as well as discrete inverse problems, Tatiana Toro for supervising my undergraduate thesis on PDE, Steven Mitchell for teaching me algebra, Don Marshall for his unique course on complex analysis, Terence Tao and Georg Menz for teaching me probability, Mario Bonk for teaching me about Loewner chains, Sorin Popa and Dimitri Shlyakhtenko for teaching and mentoring me in operator algebras.

I want to thank the following people for inspiring research discussions related to this work: Guillaume Cébron, Yoann Dabrowski, Ilijas Farah, Isaac Goldbring, Alice Guionnet, Todd Kemp, Ben Hayes, Brent Nelson, Sorin Popa, Dimitri Shlyakhtenko, Roland Speicher, Yoshimichi Ueda, Dan Voiculescu, Jurij Volčič. In particular, the "triangular transport" results in the second part were inspired by questions discussed with Ben Hayes and Yoann Dabrowski at the workshop on model theory and operator algebras at Banff International Research Station in 2018. I thank Ilijas Farah and Isaac Goldbring for explaining the model-theoretic point of view on operator algebras, as well as the referee of [Jek19] for making a more precise connection with my non-commutative function spaces (see §13.7).

The first part of this thesis draws heavily on the paper [Jek20], and one section §9.2 is closely related to [JL19, §8.2] in my joint work with Weihua Liu. The second part of the thesis is based on [Jek19] and (to a lesser extent) [Jek18]. I would like to thank the referees of those papers for their comments, some of which have affected how I wrote the thesis. My perspective has also been influenced by my joint work with Weihua Liu [JL19] and with Ben Hayes, Brent Nelson, and Thomas Sinclair [HJN19] (though the main results of those papers are separate from those of the thesis), so I would like to thank my coauthors and the referees of those papers as well.

I am indebted to everyone who took time to read and offer editorial comments to parts of the thesis, or of the papers that it draws upon. Besides the thesis committee, this includes Collin Cranston, Yoann Dabrowski, Ben Hayes, Dan Hoff, Yoshimichi Ueda, and Ping Zhong.

Funding for this research was provided in part by the following institutions and grants: The UCLA Graduate Division Dean's Scholarship, the UCLA Graduate Division Dissertation Year Fellowship, National Science Foundation grants DMS-1344970, DMS-1500035, DMS-1762360. Furthermore, I thank the Mathematische Forschungsinstitut Oberwolfach for the Simons Visiting Professor grant in March 2020 which enabled me to visit École Normale Supérieure de Lyon and Universität des Saarlandes to discuss results related to this dissertation as well as future work.

VITA

2013–2014	Teaching assistant for Math 134/5/6 (honors calculus), University of Washington, Seattle.
2014–2015	Teaching assistant for Math 334/5/6 (honors advanced calculus), University of Washington, Seattle.
2015	B.S. (Mathematics) and B.A. (Classics), University of Washington, Seattle.
2015–2019	Teaching assistant and research assistant, Department of Mathematics, University of California, Los Angeles.
2019–2020	Dissertation Year Fellowship, Graduate Division, University of California, Los Angeles.

PUBLICATIONS

David Jekel, Avi Levy, Will Dana, Austin Stromme, and Collin Litterell. "Algebraic properties of generalized graph Laplacians: Resistor networks, critical groups, and homological algebra." *SIAM Journal of Discrete Mathematics* **32**(2):1040-1110, 2018.

David Jekel. "Operator-valued chordal Loewner chains and non-commutative probability." *Journal of Functional Analysis* **278**(10):108452, 2020.

Part I

Complex evolution equations

CHAPTER 1

Introduction I

1.1 Non-commutative probability theory

Non-commutative probability theory is a branch of mathematics that studies random variables which do not commute under multiplication. In classical probability theory, random variables are one of the fundamental objects of study. If (Ω, P) is a probability space, then the space $L^\infty(\Omega, P)$ of bounded random variables is an algebra, and the expectation $E : L^\infty(\Omega, P)$ is a linear functional on this algebra. In non-commutative probability theory, $L^\infty(\Omega, P)$ is replaced by a possibly non-commutative algebra $\mathcal{A}$, and the expectation is a linear function $\phi : \mathcal{A} \to \mathbb{C}$. Often, $\mathcal{A}$ is an algebra of operators on a Hilbert space (a C*- or W*-algebra), and E is assumed to satisfy certain positivity conditions.

In classical probability, the notion of independence is so fundamental that many consider the study of probability spaces without independence to be merely measure theory. Therefore, we regard *non-commutative* probability theory as having begun when Voiculescu defined free independence in [Voi85, Voi86], even though the notion of operator algebras as non-commutative analogues of measure spaces had already been around for decades.

If $\mathcal{A}_1$, ..., $\mathcal{A}_N$ are unital C*-algebras and $\phi_j : \mathcal{A}_j \to \mathbb{C}$ is a state (the linear functional giving the "expectation"), then one can construct the *reduced free product* $(\mathcal{A}, \phi)$ of $(\mathcal{A}_1, \phi_1)$, ..., $(\mathcal{A}_N, \phi_N)$; it similar to the free product of groups. Voiculescu realized that the subalgebras $\mathcal{A}_1$, ..., $\mathcal{A}_N$ satisfy a certain condition in "joint moments" with respect to the state ϕ, called *free independence*. Furthermore, the addition of freely independent random variables satisfies a central limit theorem directly analogous to the classical one.

From that point, free probability has expanded in several related directions:

(1) A large number of results and constructions from classical probability theory were found to have analogues in free probability, with strong similarities but also interesting differences. Several other types of independence, namely, boolean and monotone independence, have parallel theories as well.

(2) Classically independent $n \times n$ random matrices (under certain conditions) become freely independent in the large-n limit. Thus, free probability has been an invaluable tool for asymptotic random matrix theory.

(3) Free probability theory has proved new structural results about operator algebras generated by freely independent random variables.

We shall explain the results and history in more detail later.

1.2 Broad structure of the thesis

This thesis comprises two parts, the first focusing on item (1) above and the second focusing on (2) and (3). The main results in the first part are largely based on [Jek20], and the the second part is based on [Jek18] and [Jek19]. The results in the two parts are largely separate and they can be read independently. However, there are broad similarities between the topics and the techniques in both parts.

- The first part explores the parallel between free, boolean, and monotone probability theories. The second part explores the connection between free probability and classical probability theory, describing how $n \times n$ random matrix tuples behave like non-commutative random variables in the large-n limit.

- The first part studies operator-valued expectations, which are a natural non-commutative analogue of conditional expectations. The second part relates classical conditional expectations for $n \times n$ random matrix models to operator-algebraic conditional expectations for the non-commutative random variables in the large-n limit.

- The first part studies "functions of non-commutative complex variables" in the sense of Taylor [Tay72, Tay73], Voiculescu [Voi00, Voi04, Voi10], and Kaliuzhnyi-Verbovetskyi and Vinnikov [KV14]. The second parts studies "functions of several non-commuting real variables" for tracial W*-algebras using a new framework introduced in [Jek18, Jek19, HJN19].

- In the first part, a non-commutative function F over a C*-algebra $\mathcal{B}$ is a sequence of functions $(F^{(n)})_{n\in\mathbb{N}}$, where $F^{(n)} : M_n(\mathcal{B}) \to M_n(\mathcal{B})$. In the second part, we study a sequence of functions $F^{(n)} : M_n(\mathbb{C})_{\text{sa}}^d \to M_n(\mathbb{C})$ for each $n \in \mathbb{N}$ and describe its large-n asymptotic behavior using a non-commutative function (where $d \in \mathbb{N}$ is fixed).

- In both parts, we study the evolution over time of non-commutative functions associated to certain families of non-commutative random variables $(X_t)_{t\geq 0}$.

1.3 Non-commutative convolutions and Lévy-Hinčin formulas

A large amount of work in free probability has been devoted to developing free versions of results and constructions in classical probability. In particular, the *free convolution* of two measures μ and ν on $\mathbb{R}$ is defined as the distribution of $X + Y$ where $X \sim \mu$ and $Y \sim \nu$ are freely independent; the boolean and monotone convolution are defined in the same way. For monotone independence, the order of the variables matters, and the independence relation obtained by reversing the order is called *anti-monotone* independence, so that the monotone convolution of μ and ν is the anti-monotone convolution of ν and μ. There are also multiplicative free, boolean, and monotone convolutions, although we will not discuss them in this thesis.

In the study of classical convolution, a central role is played by the Fourier transform, which changes convolution into multiplication. For free, boolean, and (anti-)monotone independence, that role is played by certain complex-analytic functions related to the Cauchy-Stieltjes transform $G_\mu(z) = \int_{\mathbb{R}} (z-t)^{-1}\, d\mu(t)$ and the F-transform $F_\mu(z) = 1/G_\mu(z)$. These transforms for instance provide one standard proof of the free, Boolean, and monotone central limit theorems.

A primary motivation for our main results in the first part is the Lévy-Hinčin formula, which classifies convolution semigroups. If $(X_t)_{t \geq 0}$ is a process with classically independent and stationary increments and μ_t is the distribution of X_t, then $\mu_s * \mu_t = \mu_{s+t}$ for $s, t \geq 0$, that is, $(\mu_t)_{t\geq 0}$ is a *convolution semigroup*. Such convolution semigroups are classified by the *Lévy-Hinčin formula* which expresses the Fourier transform of μ_t in terms of some other measure π on $\mathbb{R}$.

The Lévy-Hinčin formula for additive free convolution was studied in [Voi86, Maa92, BV92, BV93]. For general measures (not necessarily of bounded support), it says that $(\mu_t)_{t\geq 0}$ is a free convolution semigroup if and only if there exists $a \in \mathbb{R}$ and a finite measure ρ such that

$$F_{\mu_t}^{-1}(z) - z = t\left(a + \int_{\mathbb{R}} \frac{1+xz}{z-x}\, d\rho(x)\right). \tag{1.1}$$

In the case where μ_t has finite variance, this can be equivalently expressed as

$$F_{\mu_t}^{-1}(z) - z = t(b + G_\sigma(z)) \tag{1.2}$$

for some $b \in \mathbb{R}$ and some finite measure σ on $\mathbb{R}$.

It follows from this characterization that there is a bijection between the convolution semigroups for free convolution and classical convolution since both can be parametrized by a real number a and a finite measure ρ. Amazingly, this bijection arises from a correspondence between free and classical limit theorems. Bercovici and Pata showed in [BP99] that given sequences of measures μ_k and natural numbers $n_k \to \infty$, the classical convolution powers $\mu_k^{*n_k}$ converge to some probability measure μ if and only if the free convolution powers $\mu_k^{\boxplus n_k}$ converge to some probability measure ν. (This framework includes the sums of independent random variables considered for instance in both the central limit theorem and the Poisson limit theorem.) In this case, the measures μ and ν embed into a classical convolution semigroup $(\mu_t)_{t\geq 0}$ and a free convolution semigroup $(\nu_t)_{t\geq 0}$ as $\mu = \mu_1$ and $\nu = \nu_1$ respectively, and $(\mu_t)_{t\geq 0}$ and $(\nu_t)_{t\geq 0}$ correspond under the bijection described by the Lévy-Hinčin formulas. Thus, this correspondence between infinitely divisible measures is known as the *Bercovici-Pata bijection*.

In the same paper, Bercovici and Pata proved the analogous result linking boolean convolution semigroups and limit theorems with the free and classical cases. Here the boolean

Lévy-Hinčin formula [SW97, PV13] is given by

$$z - F_{\mu_t}(z) = t\left(a + \int_{\mathbb{R}} \frac{1+xz}{z-x}\,d\rho(x)\right), \tag{1.3}$$

and *every* probability measure on $\mathbb{R}$ turns out to embed into a boolean convolution semigroup.

The Lévy-Hinčin formula and Bercovici-Pata bijection for monotone independence were studied in [Mur00, Has10a, Has10b, AW14, AW16]. If $(\mu_t)_{t\geq 0}$ is a convolution semigroup, then $(F_{\mu_t})_{t\geq 0}$ forms a semigroup under composition and there are some $a \in \mathbb{R}$ and finite measure ρ on $\mathbb{R}$ with

$$\frac{d}{dt}\bigg|_{t=0} F_{\mu_t}(z) = a + \int_{\mathbb{R}} \frac{1+xz}{z-x}\,d\rho(x).$$

Let $A(z) = a + \int_{\mathbb{R}} (1+tz)(t-z)\,d\rho(t)$. Some straightforward computations yield that for a convolution semigroup, we have

$$\partial_t F_{\mu_t}(z) = \begin{cases} -F_{\mu_t}'(z) A(F_{\mu_t}(z)), & \text{free case,} \\ -A(z), & \text{boolean case,} \\ -F_{\mu_t}'(z) A(z), & \text{monotone case,} \\ -A(F_{\mu_t}(z)), & \text{monotone case.} \end{cases}$$

(The last two equations both hold in the monotone case, and also in the anti-monotone case since every anti-monotone convolution semigroup is also a monotone convolution semigroup and vice versa.) Written in these terms, the parallel among the four equations is quite apparent, and is much closer than the parallels between classical independence and the non-commutative independences.

Our goal is to study the situation where $(\mu_t)_{t\geq 0}$ arises from a process with independent but not necessarily stationary increments. Then in the above equations, $A(z)$ is replaced by a time-dependent $A_t(z)$. In this setting, the third equation describes the monotone case and the fourth equation the anti-monotone case (while for the case of convolution semigroups there is no distinction between the monotone and anti-monotone cases).

We should also mention here the connection between the above evolution equations and the chordal Loewner equation from complex analysis [Bau04, Sch17, FHS18]. In particular, Schleißinger [Sch17] noticed that the monotone and anti-monotone equations were exactly the forward and backward Loewner equations in the upper half-plane (see [Bau05]).

1.4 Operator-valued non-commutative probability

We will carry out our analysis of the above equations in the setting of operator-valued non-commutative probability, a version of non-commutative probability in which the expectation is not scalar-valued, but rather takes values in a C^*-algebra $\mathcal{B}$. One of the main motivations was that if $\mathcal{B}$ is a W*-subalgebra of a tracial W*-algebra $(\mathcal{A}, \tau)$, then there is a unique trace-preserving conditional expectation $\mathcal{A} \to \mathcal{B}$, which can be thought of as an $\mathcal{B}$-valued expectation and has many of the same properties as the scalar-valued expectation τ (see §10.2). Furthermore, conditional independence can be thought of simply as an $\mathcal{B}$-valued version of independence.

Thus, in the operator-valued theory we take the additional complexity of conditioning and remove it at the cost of enlarging the algebra of scalars. Many other types of complexity can be absorbed into the algebra $\mathcal{B}$ in this way.

(1) The law of a tuple $(X_1, \dots, X_n)$ over $\mathcal{B}$ can be represented as the $M_n(\mathcal{B})$-valued law of the diagonal matrix $X_1 \oplus \cdots \oplus X_n$.

(2) A non-commutative polynomial over $\mathcal{B}$ can be represented as the top left corner of a monomial over $M_n(\mathcal{B})$ in the variable $\begin{pmatrix} X & 1 \\ 1 & X \end{pmatrix}$.

(3) The resolvent of a polynomial in $X_1, \dots, X_n$ can be represented as the corner of a matrix-valued resolvent $(z - \widehat{X})^{-1}$ where z is scalar matrix and $\widehat{X}$ is a matrix with entries that are affine in $X_1, \dots, X_n$.

For further discussion on these matrix amplification tricks, refer to [HMS15, Liu18].

Motivated by these examples, mathematicians began to develop operator-valued non-commutative probability along the same lines as scalar-valued non-commutative probability. Thus, for instance, free, boolean, and monotone independence all generalize to the operator-valued setting, as well as the central limit theorems and Lévy-Hinčin formulas.

A crucial difference is that in the $\mathcal{B}$-valued setting, the notions of positivity for laws and analyticity for the various transforms associated to a law need to take into account matrix amplification. This means, roughly speaking, that anything we write down should make sense in $M_n(\mathcal{B})$ just as well as it does in $\mathcal{B}$.

Unfortunately, in the operator-valued setting, it is difficult to make sense of measures with unbounded support, and thus the Lévy-Hinčin formulas carry an additional boundedness assumption compared to the optimal results in the scalar-valued setting. We have not found any way to circumvent this issue, and thus will stick to the case of bounded support. Actually, we will provide estimates for the support radius of various distributions throughout the thesis.

1.5 Overview of the first part

The main new results of the first part are contained in §7 and §8. However, before that, we shall give extensive exposition of the technical background in operator-valued non-commutative probability theory. The aim of these chapters is to provide a reasonably short, accessible, and self-contained development of a lot of material that has not been explained in one place before.

§2 reviews C*-correspondences, which are representations of a C*-algebra $\mathcal{A}$ on a "Hilbert space" with a $\mathcal{B}$-valued inner product for some other C*-algebra $\mathcal{B}$. We define non-commutative laws and explain the GNS construction which allows any $\mathcal{B}$-valued law non-commutative law to be realized by operators on some C*-correspondence.

§3 explains the theory of fully matricial, or non-commutative, functions from [Tay72, Tay73, Voi00, Voi04, Voi10, KV14]. We give a succinct development emphasizes the parallels with single variable complex analysis. The main point of this chapter is the exposition, since the subject already has a good systematic treatment in [KV14] which is more general and emphasizes the algebraic aspects.

§4 explains the theory of the $\mathcal{B}$-valued Cauchy-Stieltjes transform of a non-commutative law as in [Voi00, Voi04, Voi10]. We give basic explicit estimates on the Cauchy-Stieltjes transform that will be important for our results on evolution equations. The chapter includes an exposition of Williams' analytic characterization of the Cauchy-Stieltjes transform [Wil17].

§5 explains $\mathcal{B}$-valued boolean, free, monotone, and anti-monotone independence. Besides setting the stage for our later results, we aim to draw out the parallels between the four types of independence as much as possible. We include the characterizations of convolution in terms of the Cauchy-Stieltjes transform with purely algebraic-analytic (rather than combinatorial) arguments.

§6 explains estimates that compare the "support radius" of non-commutative laws μ_1 and μ_2 with that of their convolution. These results are well known in the scalar-valued setting, but slightly more subtle in the operator-valued setting and do not seem to have been explained systematically before.

§7 explains our first main result about subordination families. A subordination family is a collection of laws $(\mu_t)_{t \in [0,T]}$ which arises from a process with independent (but not necessarily stationary) increments. Assuming that the mean and variance are Lipschitz in t and the support is uniformly bounded, we show that the reciprocal Cauchy-Stieltjes transforms $(F_{\mu_t})_{t \geq 0}$ of such a family satisfy a certain differential equation, which relates them to another family of $\mathcal{B}$-valued generalized laws σ_t.

The equation is slightly different for each type of independence. For instance, in the case of monotone independence, this is the operator-valued version of the chordal Loewner equation [Bau04, Sch17, Jek20]. The four equations are presented in direct parallel, and even much of the proof is shared for the four types of independence.

Although the results of §7 are generalizations of known results about convolution semigroups, the treatment of $\mathcal{B}$-valued subordination families is technically much more subtle because we cannot differentiate F_{μ_t} with respect to t in a pointwise sense. Rather, we perform the differentiation in a distributional sense, but we show that the derivative of a Lipschitz function from $[0,T]$ into an arbitrary Banach space is not too badly behaved, and can be manipulated in many ways like a pointwise defined function.

§8 proves the converse result that given a family of generalized laws $(\sigma_t)_{t\in[0,T]}$ satisfying certain assumptions, one can construct laws $(\mu_t)_{t\in[0,T]}$ satisfying the differential equation. Rather than producing the functions F_{μ_t} analytically as in [Bau04, Jek20], we directly construct a process $(X_t)_{t\in[0,T]}$ with independent increments using operators on a version of the Fock space. This is generalization of the technique used for convolution semigroups in [GSS92, Mur97, Lu97, Spe98, PV13]. The Fock space construction in §8 has been well-studied in the free and Boolean cases, but the monotone Fock space for subordination families was new in [Jek20].

Although the proof is technical, the independence of the increments arises in a completely natural way, because the Fock space itself can be decomposed as the independent product of Fock spaces associated to the subintervals in some partition of $[0,T]$. Moreover, the proof that the laws $(\mu_t)_{t\in[0,T]}$ satisfy the differential equation is quite short once we know the independence of increments.

Finally, in §9, we explain some examples and applications of these results. In particular, the results of §7 and §8 produce bijective correspondences between the subordination families for the four types non-commutative independence, which is a generalization of the Bercovici-Pata bijections for convolution semigroups.

We also explain a short proof of the free, boolean, and monotone central limit theorem (for laws with bounded support) that is based on "coupling" two operators on the same Hilbert space. This proof seems to have been overlooked until [JL19], perhaps because it heavily depends on the assumption of bounded support and perhaps because it has no analogue in classical probability. We also discuss a similar coupling technique from [Jek20] for processes with independent increments using the Fock space construction, and we illustrate it with an example of Loewner chains driven by a function on $\mathbb{R}$.

CHAPTER 2

Background: $\mathcal{B}$-valued non-commutative probability

2.1 C*-algebras

As background, we recall some fundamentals of the theory of C*-algebras. We do not give proofs for many of the statements. We refer to Blackadar [Bla06, Chapter II] for an encyclopedic list of results, proof sketches, and references.

2.1.1 C*-algebras and $*$-homomorphisms

Definition 2.1.1. A *$*$-algebra* over $\mathbb{C}$ is an algebra over $\mathbb{C}$ together with a map $a \mapsto a^*$ such that $(a^*)^* = a$, the $*$ operation is conjugate-linear, and $(ab)^* = b^* a^*$. If $\mathcal{A}$ and $\mathcal{B}$ are $*$-algebras, then a $*$-homomorphism $\rho : \mathcal{A} \to \mathcal{B}$ is a homomorphism such that $\rho(a^*) = \rho(a)^*$.

Definition 2.1.2. A *(unital)* C*-*algebra* is a unital $*$-algebra $\mathcal{A}$ over $\mathbb{C}$ together with a norm $\|\cdot\|$ such that

(1) $(\mathcal{A}, \|\cdot\|)$ is a Banach space.

(2) $\|ab\| \leq \|a\|\|b\|$.

(3) $\|a^* a\| = \|a\|^2$.

Theorem 2.1.3. *Let $\mathcal{H}$ be a Hilbert space. If $\mathcal{A}$ is a subalgebra of $B(\mathcal{H})$ which is closed under adjoints and closed in operator norm, then $\mathcal{A}$ is a* C*-*algebra, where the $*$-operation is the adjoint and the norm is the operator norm. Conversely, every* C*-*algebra is isometrically $*$-isomorphic to such a C^*-algebra of operators on a Hilbert space.*

Proposition 2.1.4. *Suppose that $\mathcal{A}$ and $\mathcal{B}$ are* C*-*algebras.*

(1) If $\rho : \mathcal{A} \to \mathcal{B}$ is a $*$-homomorphism, then $\|\rho(a)\| \leq \|a\|$ for every $a \in \mathcal{A}$.

(2) If $\rho : \mathcal{A} \to \mathcal{B}$ is an injective $*$-homomorphism, then $\|\rho(a)\| = \|a\|$ for every $a \in \mathcal{A}$.

(3) If $\mathcal{A}$ is a C*-*algebra, then there is only one norm on $\mathcal{A}$ which satisfies the* C*-*algebra conditions.*

2.1.2 Positivity and states

Definition 2.1.5. An element a of a C*-algebra $\mathcal{A}$ is said to be *positive* if a can be written as x^*x for some $x \in \mathcal{A}$. We also write this condition as $a \geq 0$. Furthermore, we write $a \geq b$ if $a - b \geq 0$.

Definition 2.1.6. A linear functional $\phi : \mathcal{A} \to \mathbb{C}$ is *positive* if $a \geq 0$ implies $\phi(a) \geq 0$.

Definition 2.1.7. A *state* on a C*-algebra $\mathcal{A}$ is a positive linear functional with $\phi(1) = 1$. We denote the set of states by $S(\mathcal{A})$.

Proposition 2.1.8. *Let $\mathcal{A}$ be a C*-algebra.*

(1) Suppose that $\mathcal{A}$ is a C-algebra acting on a Hilbert space $\mathcal{H}$. An element $a \in \mathcal{A}$ is positive if and only if a is a positive operator on $\mathcal{H}$.*

(2) If ϕ is a positive linear functional, then $\|\phi\|_{\mathcal{A}^*} = |\phi(1)|$. In particular, the norm of a state is 1.

(3) If $a \in \mathcal{A}$ is self-adjoint, then

$$\|a\| = \sup_{\phi \in S(\mathcal{A})} |\phi(a)|.$$

(4) If $a \in \mathcal{A}$, then a is self-adjoint if and only if $\phi(a)$ is real for every state ϕ.

(5) If $a \in \mathcal{A}$, then we have $a \geq 0$ if and only if $\phi(a) \geq 0$ for every state ϕ.

2.1.3 The GNS construction

Given a state ϕ on a C*-algebra $\mathcal{A}$, one can define a sesquilinear form on $\mathcal{A}$ by $\langle a, b \rangle_\phi = \phi(a^*b)$. This form is nonnegative definite, and hence it satisfies the Cauchy-Schwarz inequality. If $\mathcal{K}_\phi = \{a : \phi(a^*a) = 0\}$, then the completion of $\mathcal{H}/\mathcal{K}_\phi$ with respect to $\|a\|_\phi = \phi(a^*a)^{1/2}$ is a Hilbert space, which we denote by $L^2(\mathcal{A}, \phi)$.

Moreover, every $a \in \mathcal{A}$ defines a bounded operator on $L^2(\mathcal{A}, \phi)$ by left multiplication. Indeed, because $a \mapsto \phi(b^*ab)$ is a positive functional and $\|a\|^2 - a^*a \geq 0$, we have $\|ab\|_\phi^2 = \phi(b^*a^*ab) \leq \|a\|^2 \phi(b^*b) = \|a\|^2 \|b\|_\phi^2$. Thus, the multiplication action of a is well-defined on the separation-completion $L^2(\mathcal{A}, \phi)$.

Therefore, there is a $*$-homomorphism $\pi_\phi : \mathcal{A} \to B(L^2(\mathcal{A}, \phi))$ given by $\pi_\phi(a)[b] = [ab]$, where $[b]$ is the equivalence class of b in the separation-completion. This is called the *Gelfand-Naimark-Segal representation* of $\mathcal{A}$ on $L^2(\mathcal{A}, \phi)$. Furthermore, as a consequence of Proposition 2.1.8 (3), we have the following representation of $\mathcal{A}$.

Theorem 2.1.9. *Let $\mathcal{H} = \bigoplus_{\phi \in S(\mathcal{A})} L^2(\mathcal{A}, \phi)$, and let $\pi : \mathcal{A} \to B(\mathcal{H})$ be the direct sum of the GNS representations π_ϕ. Then π is an isometric $*$-isomorphism.*

This construction is the basis of the fact that every C*-algebra can be represented concretely on a Hilbert space.

2.1.4 Matrices over a C*-algebra

Suppose that $\mathcal{A}$ is a C*-algebra and let us realize $\mathcal{A}$ as an algebra of operators on the Hilbert space $\mathcal{H}$ as in Theorem 2.1.3. Then a matrix $x \in M_{n\times m}(\mathcal{A}) = \mathcal{A} \otimes M_{n\times m}(\mathbb{C})$ can be viewed as an operator $\mathcal{H}^n \to \mathcal{H}^m$, and we denote by $\|x\|$ its operator norm. Note that $M_{n\times m}(\mathcal{A})$ is already complete in the operator norm.

In particular, $M_n(\mathcal{A})$ is a C*-algebra. Moreover, Proposition 2.1.4 (3) implies that $M_n(\mathcal{A})$ has a unique norm and thus the norm is independent of our choice of representation for $\mathcal{A}$ on a Hilbert space $\mathcal{H}$. Furthermore, the norm on $M_{n\times m}(\mathcal{A})$ is also independent of the representation because if $x \in M_{n\times m}(\mathcal{A})$ then the operator norm satisfies $\|x\|^2 = \|x^*x\|$, and $x^*x \in M_m(\mathcal{A})$ hence $\|x^*x\|$ is independent of the choice of representation.

Furthermore, there is a coordinate-free characterization of positivity in $M_n(\mathcal{A})$ in terms of positivity in $\mathcal{A}$.

Lemma 2.1.10. *Let $A \in M_n(\mathcal{A})$. Then the following are equivalent:*

(1) $A \geq 0$ in $\mathcal{A}$.

(2) For every $v \in M_{1\times n}(\mathcal{A})$, we have $v^*Av \geq 0$ in $\mathcal{A}$.

Proof. As in Theorem 2.1.9, we can represent $\mathcal{A}$ as a concrete C^*-algebra of operators on $\mathcal{H} := \bigoplus_{\phi\in S(\mathcal{A})} \mathcal{H}_\phi$, where $\mathcal{H}_\phi = L^2(\mathcal{A}, \phi)$.

We can view A as an operator $\mathcal{H}^n \to \mathcal{H}^n$ and v as an operator $\mathcal{H} \to \mathcal{H}^n$. If $A \geq 0$, then v^*Av is positive by the basic theory of operators on Hilbert space, and hence $v^*Av \geq 0$ in $\mathcal{A}$.

Conversely, suppose that (2) holds. Observe that

$$\mathcal{H}^n = \bigoplus_{\phi\in S(\mathcal{A})} \mathcal{H}_\phi^n,$$

and the action of A on $\mathcal{H}^n$ is the direct sum of its actions on each $\mathcal{H}_\phi^n$. So it suffices to show that $A|_{\mathcal{H}_\phi^n}$ is positive for each state ϕ. We know that for each $v \in M_{1\times n}(\mathcal{A}) \cong \mathcal{A}^n$, we have

$$\phi(v^*Av) \geq 0.$$

Let $[v]$ denote the vector $([v_1], \dots, [v_n])$ as an equivalence class in $\mathcal{H}_\phi^n$. Then $\langle [v], A[v]\rangle \geq 0$. Such vectors $[v]$ are dense in $\mathcal{H}_\phi$ by construction and hence $A|_{\mathcal{H}_\phi} \geq 0$ as desired. $\square$

2.2 Right Hilbert $\mathcal{B}$-modules

We begin with the $\mathcal{B}$-valued analogue of a Hilbert space. Right Hilbert $\mathcal{B}$-modules were introduced by Kaplansky [Kap53], Paschke [Pas73], and Rieffel [Rie74]. For further detail, see [Lan95]. A list of theorems and references can be found in [Bla06, §II.7].

2.2.1 Definition, inner products, separation-completion

Definition 2.2.1. Let $\mathcal{H}$ be a right $\mathcal{B}$-module. Then an $\mathcal{B}$*-valued pre-inner product* on $\mathcal{H}$ is a map $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathcal{B}$ such that

(1) *Right* $\mathcal{B}$*-linearity:* We have

$$\langle \xi, \zeta_1 b_1 + \zeta_2 b_2 \rangle = \langle \xi, \zeta_1 \rangle b_1 + \langle \xi, \zeta_2 \rangle b_2.$$

for ξ, ζ_1, $\zeta_2 \in \mathcal{H}$ and b_1, $b_2 \in \mathcal{B}$.

(2) *Symmetry:* We have $\langle \xi, \zeta \rangle^* = \langle \zeta, \xi \rangle$.

(3) *Nonnegativity:* $\langle \xi, \xi \rangle \geq 0$ in $\mathcal{B}$ for every $\xi \in \mathcal{H}$.

If in addition, $\langle \xi, \xi \rangle = 0$ implies that $\xi = 0$, then we say $\langle \cdot, \cdot \rangle$ is an $\mathcal{A}$*-valued inner product.*

Observation 2.2.2. *An* $\mathcal{A}$*-valued pre-inner product satisfies* $\langle \zeta_1 a_1 + \zeta_2 a_2, \xi \rangle = a_1^* \langle \zeta_1, \xi \rangle + a_2^* \langle \zeta_2, \xi \rangle$.

Lemma 2.2.3. *Let* $\mathcal{H}$ *be a right* $\mathcal{B}$*-module with an* $\mathcal{B}$*-valued pre-inner product, and define* $\|\xi\| = \|\langle \xi, \xi \rangle\|^{1/2}$.

(1) $\langle \zeta, \xi \rangle \langle \xi, \zeta \rangle \leq \|\xi\|^2 \langle \zeta, \zeta \rangle$.

(2) $\|\langle \xi, \zeta \rangle\| \leq \|\xi\| \|\zeta\|$.

(3) $\|\xi\| = \|\langle \xi, \xi \rangle\|^{1/2}$ *defines a semi-norm on* $\mathcal{H}$.

(4) $\|\xi\| = \sup_{\|\zeta\| \leq 1} \|\langle \xi, \zeta \rangle\|$.

Proof. Suppose that $\phi \in S(\mathcal{B})$. Then $\phi(\langle \xi, \zeta \rangle)$ is a scalar-valued pre-inner product and therefore satisfies the Cauchy-Schwarz inequality. Thus, we have

$$\begin{aligned}
\phi(\langle \zeta, \xi \rangle \langle \xi, \zeta \rangle) &= \phi(\langle \zeta, \xi \langle \xi, \zeta \rangle \rangle) \\
&\leq \phi(\langle \zeta, \zeta \rangle)^{1/2} \phi(\langle \xi \langle \xi, \zeta \rangle, \xi \langle \xi, \zeta \rangle \rangle)^{1/2} \\
&= \phi(\langle \zeta, \zeta \rangle)^{1/2} \phi(\langle \xi, \zeta \rangle^* \langle \xi, \xi \rangle \langle \xi, \zeta \rangle)^{1/2}
\end{aligned}$$

Next, note that $a \mapsto \phi(\langle \xi, \zeta \rangle^* a \langle \xi, \zeta \rangle)$ is positive linear functional on $\mathcal{B}$ and therefore

$$|\phi(\langle \xi, \zeta \rangle^* \langle \xi, \xi \rangle \langle \xi, \zeta \rangle)| \leq \|\langle \xi, \xi \rangle\| \phi(\langle \xi, \zeta \rangle^* \langle \xi, \zeta \rangle) = \|\xi\|^2 \phi(\langle \xi, \zeta \rangle^* \langle \xi, \zeta \rangle).$$

Altogether,

$$\phi(\langle \zeta, \xi \rangle \langle \xi, \zeta \rangle) \leq \phi(\langle \zeta, \zeta \rangle)^{1/2} \|\xi\| \phi(\langle \zeta, \xi \rangle \langle \xi, \zeta \rangle)^{1/2}.$$

We cancel the term $\phi(\langle \zeta, \xi \rangle \langle \xi, \zeta \rangle)^{1/2}$ from both sides and then square the inequality to obtain

$$\phi(\langle \zeta, \xi \rangle \langle \xi, \zeta \rangle) \leq \|\xi\|^2 \phi(\langle \zeta, \zeta \rangle).$$

Because $\langle \zeta, \xi \rangle \langle \xi, \zeta \rangle$ and $\|\xi\|^2 \langle \zeta, \zeta \rangle$ are self-adjoint elements of $\mathcal{B}$ and this inequality holds for every state ϕ, we have

$$\langle \zeta, \xi \rangle \langle \xi, \zeta \rangle \leq \|\xi\|^2 \langle \zeta, \zeta \rangle,$$

so (1) is proved. Inequality (2) follows by taking the norm of both sides in $\mathcal{B}$ and then taking the square root.

The norm on $\mathcal{H}$ is clearly positive-homogeneous. The triangle inequality holds because

$$\begin{aligned} \|\xi + \zeta\|^2 &= \|\langle \xi + \zeta, \xi + \zeta \rangle\| \\ &\leq \|\langle \xi, \xi \rangle\| + \|\langle \xi, \zeta \rangle\| + \|\langle \zeta, \xi \rangle\| + \|\langle \zeta, \zeta \rangle\| \\ &\leq (\|\xi\| + \|\zeta\|)^2. \end{aligned}$$

This proves (3). Moreover, (4) follows immediately from the Cauchy-Schwarz inequality (2). □

Definition 2.2.4. A *right Hilbert $\mathcal{B}$-module* is a right $\mathcal{B}$-module with an $\mathcal{B}$-valued pre-inner product such that $\mathcal{H}$ is a Banach space with respect to the semi-norm $\|\xi\| = \|\langle \xi, \xi \rangle\|^{1/2}$.

Lemma 2.2.5. *Let $\mathcal{H}$ be a right $\mathcal{B}$-module with an $\mathcal{B}$-valued pre-inner product. Define*

$$\mathcal{K} = \{\xi \in \mathcal{H} : \|\xi\| = 0\}.$$

Then $\langle \cdot, \cdot \rangle$ defines an inner product $\mathcal{H}/\mathcal{K}$, and the completion of $\mathcal{H}/\mathcal{K}$ with respect to the corresponding norm is a right Hilbert $\mathcal{B}$-module.

Proof. The Cauchy-Schwarz inequality implies that $\langle \cdot, \cdot \rangle$ yields a well-defined inner product on $\mathcal{H}/\mathcal{K}$. The right $\mathcal{B}$-action is bounded with respect to the norm of $\mathcal{H}$ since

$$\|\xi b\|^2 = \|\langle \xi b, \xi b \rangle\| = \|b^* \langle \xi, \xi \rangle b\| \leq \|\xi\|^2 \|b\|^2.$$

Thus, the right $\mathcal{B}$-action maps $\mathcal{K}$ into $\mathcal{K}$ and hence passes to a bounded action on the quotient. This in turn extends to the completion. The $\mathcal{B}$-valued inner product on $\mathcal{H}/\mathcal{K}$ extends to an $\mathcal{B}$-valued inner product on the completion because of the Cauchy-Schwarz inequality and the boundedness of the right $\mathcal{B}$-action. □

The module defined in the lemma above will be called the *separation-completion* of the right $\mathcal{B}$-module $\mathcal{H}$.

2.2.2 Orthogonality

Definition 2.2.6. If $\mathcal{H}$ is a right Hilbert $\mathcal{B}$-module, then we say that $\xi_1, \dots, \xi_n \in \mathcal{H}$ are *orthogonal* if $\langle \xi_i, \xi_j \rangle = 0$ for $i \neq j$.

Unlike the scalar case, there is no reason why orthonormal bases would exist in general. However, when we have orthogonal vectors, a version of the Pythagorean identity still holds

Observation 2.2.7. *If $\xi_1, \ldots, \xi_n$ are orthogonal, then*

$$\left\langle \sum_{j=1}^{n} \xi_j, \sum_{j=1}^{n} \xi_j \right\rangle = \sum_{j=1}^{n} \langle \xi_j, \xi_j \rangle,$$

and hence

$$\left\| \sum_{j=1}^{n} \xi_j \right\| \leq \left(\sum_{j=1} \|\xi_j\|^2 \right)^{1/2}.$$

2.2.3 Operators on right Hilbert $\mathcal{B}$-modules

Definition 2.2.8. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be right Hilbert $\mathcal{B}$-modules. A linear map $T : \mathcal{H}_1 \to \mathcal{H}_2$ is *bounded* if

$$\|T\| := \sup_{\|h\| \leq 1} \|Th\| < +\infty.$$

We say that T is *right-$\mathcal{A}$-linear* if $(Th)b = T(hb)$ for each $b \in \mathcal{B}$.

The adjoint of a linear operator is defined the same way as in the scalar case, except that there is a no guarantee that an adjoint exists.

Definition 2.2.9. Let $T : \mathcal{H}_1 \to \mathcal{H}_2$ be a bounded right-$\mathcal{B}$-linear map between right Hilbert $\mathcal{B}$-modules. Then T is *adjointable* if there exists $T^* : \mathcal{H}_2 \to \mathcal{H}_1$ such that

$$\langle Th_1, h_2 \rangle_{\mathcal{H}_2} = \langle h_1, T^* h_2 \rangle_{\mathcal{H}_1}.$$

In this case, we say that T^* is an adjoint for T.

Proposition 2.2.10.

(1) If $T : \mathcal{H}_1 \to \mathcal{H}_2$ is adjointable, then the adjoint is unique.

(2) If $T : \mathcal{H}_1 \to \mathcal{H}_2$ and $S : \mathcal{H}_2 \to \mathcal{H}_3$ are adjointable, then $(ST)^* = T^* S^*$.

(3) If T is adjointable, then T^* is adjointable and $T^{**} = T$.

(4) $\|T^* T\| = \|T\|^2 = \|T^*\|^2$.

Proof. (1) Suppose that S and S' are two adjoints for T. Then for every h_1 and h_2, we have

$$\langle h_1, (S - S')h_2 \rangle = \langle Th_1, h_2 \rangle - \langle Th_1, h_2 \rangle = 0.$$

For each h_2, we can take $h_1 = (S - S')h_2$ to conclude that $Sh_2 = S'h_2$.

(2) Given that the adjoint is unique, this equality follows from the fact that

$$\langle STh_1, h_3 \rangle = \langle Th_1, S^* h_3 \rangle = \langle h_1, S^* T^* h_3 \rangle.$$

(3) Note that

$$\langle T^* h_2, h_1 \rangle = \langle h_1, T^* h_2 \rangle^* = \langle T h_1, h_2 \rangle^* = \langle h_2, T h_1 \rangle.$$

(4) Observe that

$$\begin{aligned} \|T\| = \sup_{\|h_1\| \le 1} \|T h_1\| &= \sup_{\|h_1\|, \|h_2\| \le 1} \|\langle T h_1, h_2 \rangle\| \\ &= \sup_{\|h_1\|, \|h_2\| \le 1} \|\langle h_1, T^* h_2 \rangle\| = \sup_{\|h_1\|, \|h_2\| \le 1} \|\langle T^* h_2, h_1 \rangle\| = \|T^*\|. \end{aligned}$$

Moreover, using the Cauchy-Schwarz inequality,

$$\|T^* T\| = \sup_{\|h_1\|, \|h_1'\| \le 1} \|\langle T^* T h_1, h_1' \rangle\| = \sup_{\|h_1\|, \|h_1'\| \le 1} \|\langle T h_1, T h_1' \rangle\| = \left(\sup_{\|h_1\| \le 1} \|T h_1\| \right)^2 = \|T\|^2.$$

□

Definition 2.2.11. We denote the $*$-algebra of bounded, adjointable, right-$\mathcal{B}$-linear operators $\mathcal{H} \to \mathcal{H}$ by $B(\mathcal{H})$.

2.3 C*-correspondences

Now we introduce the $\mathcal{B}$-valued analogue of a representation of a C*-algebra on a Hilbert space.

Definition 2.3.1. Let $\mathcal{A}$ and $\mathcal{B}$ be C*-algebras. An $\mathcal{A}$-$\mathcal{B}$-*correspondence* is a right Hilbert $\mathcal{B}$-module $\mathcal{H}$ together with $*$-homomorphism $\pi : \mathcal{A} \to B(\mathcal{H})$.

In this case, for $a \in \mathcal{A}$, we write $a\xi := \pi(a)\xi$, and thus view $\mathcal{H}$ as a $\mathcal{A}$-$\mathcal{B}$-bimodule. The left and right actions commute because by definition $B(\mathcal{H})$ consists of right-$\mathcal{B}$-linear operators. A Hilbert $\mathcal{A}$-$\mathcal{B}$-correspondence can be thought of as a representation of a C*-algebra $\mathcal{A}$ on a $\mathcal{B}$-valued Hilbert space. Of course, a $\mathbb{C}$-$\mathcal{B}$-correspondence is equivalent to a right Hilbert $\mathcal{B}$-module.

2.3.1 Direct sums

Given a family of $\mathcal{A}$-$\mathcal{B}$-correspondences $\{\mathcal{H}_i\}_{i \in I}$, we define the *direct sum* $\bigoplus_{i \in I} \mathcal{H}_i$ as the completion of the algebraic direct sum with respect to the $\mathcal{B}$-valued inner product

$$\left\langle \sum_{i \in I} \xi_i, \sum_{i \in I} \zeta_i \right\rangle = \sum_{i \in I} \langle \xi_i, \zeta_i \rangle_{\mathcal{H}_i},$$

where $\sum_{i\in I}\xi_i$ and $\sum_{i\in I}\zeta_i$ are elements of the algebraic direct sum represented as sums of $\xi_i \in \mathcal{H}_i$ and $\zeta_i \in \mathcal{H}_i$ with only finitely many nonzero terms.

We must still verify that this definition makes sense. It is straightforward to check that this is an inner product, and therefore the completion is well-defined as a right Hilbert $\mathcal{B}$-module by Lemma 2.2.5. But it remains to show that left $\mathcal{A}$-action is bounded and extends to the completion. Let $a \in \mathcal{A}$ and let $\sum_{i\in I}\xi_i$ be in the algebraic direct sum. Then $\|a\|^2 - a^*a \geq 0$ in $\mathcal{A}$, and hence

$$\|a\|^2\langle\xi_i,\xi_i\rangle - \langle a\xi_i, a\xi_i\rangle = \langle\xi_i,(\|a\|^2 - a^*a)\xi_i\rangle \geq 0,$$

which implies that

$$\left\langle a\sum_{i\in I}\xi_i, a\sum_{i\in I}\xi_i\right\rangle = \sum_{i\in I}\langle a\xi_i, a\xi_i\rangle \leq \|a\|^2\sum_{i\in I}\langle\xi_i,\xi_i\rangle = \|a^2\|\left\langle\sum_{i\in I}\xi_i,\sum_{i\in I}\xi_i\right\rangle.$$

Therefore, the $\mathcal{A}$-action is bounded and so extends to the completion.

The direct sum operation is commutative and associative, up to natural isomorphism.

2.3.2 Tensor products

Suppose we are given an $\mathcal{A}$-$\mathcal{B}$-correspondence $\mathcal{H}$ and a $\mathcal{B}$-$\mathcal{C}$-correspondence $\mathcal{K}$. Then we define the *tensor product* $\mathcal{H}\otimes_{\mathcal{B}}\mathcal{K}$ by equipping the algebraic tensor product with the pre-inner product

$$\langle\xi_1\otimes\zeta_1,\xi_2\otimes\zeta_2\rangle = \langle\zeta_1,\langle\xi_1,\xi_2\rangle\zeta_2\rangle$$

and then forming the separation-completion as in Lemma 2.2.5.

Let us expound the definition in more detail and verify that the construction makes sense. Let $\mathcal{V}$ be the algebraic tensor product of $\mathcal{H}$ and $\mathcal{K}$ over $\mathcal{B}$. That is, $\mathcal{V}$ is the vector space spanned by $\xi\otimes\zeta$, where $\xi\in\mathcal{H}$ and $\zeta\in\mathcal{K}$, modulo the span of vectors of the form

$$\xi\otimes(\zeta_1+\zeta_2) - \xi\otimes\zeta_1 - \xi\otimes\zeta_2, \qquad (\xi_1\otimes\xi_2)\otimes\zeta - \xi_1\otimes\zeta - \xi_2\otimes\zeta, \xi b\otimes\zeta - \xi\otimes b\zeta,$$

where $b\in\mathcal{B}$. Note that $\mathcal{V}$ is a $\mathcal{A}$-$\mathcal{C}$-correspondence with the actions given by

$$a(\xi\otimes\zeta) = a\xi\otimes\zeta, \qquad (\xi\otimes\zeta)c = \xi\otimes\zeta c.$$

We equip $\mathcal{V}$ with a $\mathcal{C}$-valued form $\langle\cdot,\cdot\rangle$ given by

$$\langle\xi_1\otimes\zeta_1,\xi_2\otimes\zeta_2\rangle = \langle\zeta_1,\langle\xi_1,\xi_2\rangle\zeta_2\rangle.$$

Observe that if we replace $\xi_j b\otimes\zeta_j$ with $\xi_j\otimes b\zeta_j$ for $j = 1$ or 2 and $b\in\mathcal{B}$, the result is unchanged due to the right $\mathcal{B}$-linearity of the inner product on $\mathcal{H}$; therefore, this $\mathcal{C}$-valued form on $\mathcal{V}$ is well-defined. It is straighforward to check that this $\mathcal{C}$-valued form is right $\mathcal{C}$-linear and symmetric.

In order to check that this is nonnegative, consider a sum of simple tensors $\sum_{j=1}^n \xi_j \otimes \zeta_j$. Note that

$$\left\langle \sum_i \xi_i \otimes \zeta_i, \sum_j \xi_j \otimes \zeta_j \right\rangle = \sum_{i,j} \langle \zeta_i, \langle \xi_i, \xi_j \rangle \zeta_j \rangle = \langle \vec{\zeta}, X\vec{\zeta} \rangle_{\mathcal{K}^n},$$

where $\vec{\zeta} = (\zeta_1, \ldots, \zeta_n) \in \mathcal{K}^n$ and X is the matrix $[\langle \xi_i, \xi_j \rangle]_{i,j}$ in $M_n(\mathcal{B})$. We claim that $X \geq 0$ in $M_n(\mathcal{B})$. This follows from Lemma 2.1.10 because for $v \in M_{n\times 1}(\mathcal{B})$, then

$$v^*Xv = \sum_{i,j} \langle \xi_i v_i, \xi_j v_j \rangle = \left\langle \sum_i \xi_i v_i, \sum_j \xi_j v_j \right\rangle \geq 0.$$

Thus, X can be written as B^*B for some $B \in M_n(\mathcal{B})$. Thus,

$$\langle \vec{\zeta}, X\vec{\zeta} \rangle_{\mathcal{H}^n} = \langle B\vec{\zeta}, B\vec{\zeta} \rangle_{\mathcal{H}^n} \geq 0.$$

This shows nonnegativity of the inner product.

Therefore, Lemma 2.2.5 shows that the separation-completion of $\mathcal{V}$ with respect to $\langle \cdot, \cdot \rangle$ is a well-defined right Hilbert $\mathcal{A}$-module $\mathcal{K} \otimes_{\mathcal{B}} \mathcal{H}$. Finally, we must verify that the left $\mathcal{A}$-action is well-defined. Let $a \in \mathcal{A}$. Then $\|a\|^2 - a^*a \geq 0$, so that $\|a\|^2 - a^*a = x^*x$ for some $x \in \mathcal{A}$. Thus, for a simple tensor $\sum_j \xi_j \otimes \zeta_j$, we have

$$\left\langle \sum_i \xi_i \otimes \zeta_i, (\|a\|^2 - a^*a) \sum_j \xi_j \otimes \zeta_j \right\rangle = \left\langle \sum_i x\xi_i \otimes \zeta_i, \sum_j x\xi_j \otimes \zeta_j \right\rangle \geq 0,$$

which implies that

$$\left\langle a \sum_i \xi_i \otimes \zeta_i, a \sum_j \xi_j \otimes \zeta_j \right\rangle \leq \|a\|^2 \left\langle \sum_i \xi_i \otimes \zeta_i, \sum_j \xi_j \otimes \zeta_j \right\rangle.$$

Hence, the action of a is bounded and thus passes to the separation-completion. Moreover, direct computation shows that the action of $\mathcal{A}$ is a $*$-homomorphism.

This shows that the tensor product is well-defined. Furthermore, it is straightforward to check that the tensor product is associative, that is, if $\mathcal{H}_j$ is an $\mathcal{A}_{j-1}$-$\mathcal{A}_j$-correspondence for $j = 1, 2, 3$, then

$$(\mathcal{H}_1 \otimes_{\mathcal{A}_1} \mathcal{H}_2) \otimes_{\mathcal{A}_2} \mathcal{H}_3 \cong \mathcal{H}_1 \otimes_{\mathcal{A}_1} (\mathcal{H}_2 \otimes_{\mathcal{A}_2} \mathcal{H}_3)$$

as an $\mathcal{A}_0$-$\mathcal{A}_3$-correspondence. In particular, we can unambiguously (up to natural isomorphism) write

$$\mathcal{H}_1 \otimes_{\mathcal{A}_1} \cdots \otimes_{\mathcal{A}_{n-1}} \mathcal{H}_n$$

as an $\mathcal{A}_0$-$\mathcal{A}_n$-correspondence when $\mathcal{H}_j$ is an $\mathcal{A}_{j-1}$-$\mathcal{A}_j$-correspondence for $j = 1, \ldots, n$. Moreover, tensor products distribute over direct sums in the obvious way. We also have the following useful property:

Lemma 2.3.2. *Let $\mathcal{H}$ be an $\mathcal{A}$-$\mathcal{B}$-correspondence, let $\mathcal{K}$ be a $\mathcal{B}$-$\mathcal{C}$-correspondence, and consider the tensor product $\mathcal{H} \otimes_{\mathcal{B}} \mathcal{K}$. If $\xi \in \mathcal{H}$ and $\zeta \in \mathcal{K}$, then*

$$\langle \xi \otimes \zeta, \xi \otimes \zeta \rangle \leq \|\langle \xi, \xi \rangle\| \langle \zeta, \zeta \rangle$$

and hence

$$\|\xi \otimes \zeta\| \leq \|\xi\| \|\zeta\|.$$

Proof. Use $\langle \xi \otimes \zeta, \xi \otimes \zeta \rangle = \langle \zeta, \langle \xi, \xi \rangle \zeta \rangle$ and the fact that $\|\xi\|^2 - \langle \xi, \xi \rangle$ is a positive operator in $\mathcal{B}$. □

2.4 Completely positive maps and the GNS construction

Now we will define the $\mathcal{B}$-valued analogue of positive linear functionals on an algebra $\mathcal{A}$ and the GNS construction. It turns out that positivity of a map $\sigma : \mathcal{B} \to \mathcal{A}$ is not a strong enough condition to make the GNS construction work. Rather, we need the notion of complete positivity. Complete positivity was first studied by Stinespring [Sti55], and the operator-valued GNS construction is closely related to the Stinespring dilation theorem and its extension by Kasparov [Kas80]. For further references, see [Bla06, §II.6.9-10, §II.7.5].

Definition 2.4.1. Let $\sigma : \mathcal{A} \to \mathcal{B}$ be a linear map. We denote by $\sigma^{(n)} : M_n(\mathcal{A}) \to M_n(\mathcal{B})$ the map given by applying σ entrywise. We say that σ is *completely positive* if $\sigma^{(n)}$ is positive for every n, that is, $\sigma^{(n)}(A^*A) \geq 0$ for every $A \in M_n(\mathcal{A})$.

Lemma 2.4.2. *Let $\mathcal{H}$ be an Hilbert $\mathcal{A}$-$\mathcal{B}$-correspondence and $\xi \in \mathcal{H}$. Then $\sigma(b) := \langle \xi, b\xi \rangle$ is a completely positive map $\mathcal{A} \to \mathcal{B}$.*

Proof. Choose a positive element A^*A in $M_n(\mathcal{A})$ and write $A = [a_{i,j}]$. By Lemma 2.1.10, to show that $\sigma^{(n)}(A^*A) \geq 0$, it suffices to show that for $v \in M_{n\times 1}(\mathcal{B})$, we have $v^*\sigma^{(n)}(A^*A)v \geq 0$. But

$$v^*\sigma^{(n)}(A^*A)v = \sum_{i,j} \langle \xi v_i, (A^*A)_{i,j} \xi v_j \rangle = \sum_{i,j,k} \langle a_{k,i} \xi v_i, a_{k,j} \xi v_j \rangle = \langle A(\xi v), A(\xi v) \rangle_{\mathcal{H}^n} \geq 0,$$

where $\xi v \in \mathcal{H}^n$ is the vector $(\xi v_1, \dots, \xi v_n)$ and A acts on $\mathcal{H}^n$ by matrix multiplication in the obvious way. □

Conversely, we will show that every completely positive map $\sigma : \mathcal{A} \to \mathcal{B}$ can be realized by a vector ξ in a Hilbert $\mathcal{A}$-$\mathcal{B}$-correspondence. We define an $\mathcal{A}$-$\mathcal{B}$-correspondence $\mathcal{A} \otimes_\sigma \mathcal{B}$ by equipping the algebraic tensor product $\mathcal{A} \otimes_{\text{alg}} \mathcal{B}$ over $\mathbb{C}$ with the pre-inner product

$$\langle a_1 \otimes b_1, a_2 \otimes b_2 \rangle = b_1^* \sigma(a_1^* a_2) b_2.$$

This pre-inner product is clearly right $\mathcal{B}$-linear and symmetric. To show that it is nonnegative, consider a vector

$$\xi = \sum_{j=1}^{n} a_j \otimes b_j$$

and note that

$$\langle \xi, \xi \rangle = \sum_{i,j} b_i^* \sigma(a_i^* a_j) b_j.$$

The matrix $C = [a_i^* a_j]$ can be written in the form A^*A and hence is positive in $M_n(\mathcal{A})$. Therefore, by complete positivity of σ, the matrix $[\sigma(a_i^* a_j)]$ is positive in $M_n(\mathcal{B})$. Then by Lemma 2.1.10, $\sum_{i,j} b_i^* \sigma(a_i^* a_j) b_j \geq 0$ in $\mathcal{A}$. This shows nonnegativity of the pre-inner product.

Thus, by Lemma 2.2.5, we can define the separation-completion $\mathcal{A} \otimes_\sigma \mathcal{B}$ as a right Hilbert $\mathcal{B}$-module. Finally, we claim that the left multiplication action of $\mathcal{A}$ on $\mathcal{A} \otimes_{\text{alg}} \mathcal{B}$ passes to the separation-completion. To do this, it suffices to show that this action is bounded with respect to $\langle \cdot, \cdot \rangle$.

The argument is the same as in the construction of the tensor product for bimodules. Given $a \in \mathcal{A}$, we have $\|a\|^2 - a^*a \geq 0$ and hence it can be written as x^*x for some $x \in \mathcal{B}$. Using complete positivity, one argues that $\langle x\xi, x\xi \rangle \geq 0$ whenever $\xi = \sum_{j=1}^{n} a_j \otimes b_j$. Thus, we conclude that $\|a\xi, a\xi\| \leq \|a\|^2 \langle \xi, \xi \rangle$. In summary, we have shown that the following definition makes sense.

Definition 2.4.3. Let $\sigma : \mathcal{A} \to \mathcal{B}$ be completely positive. We denote by $\mathcal{A} \otimes_\sigma \mathcal{B}$ the Hilbert $\mathcal{A}$-$\mathcal{B}$-correspondence defined as the separation-completion of the algebraic tensor product $\mathcal{A} \otimes \mathcal{B}$ over $\mathbb{C}$ with respect to the pre-inner product $\langle a_1 \otimes b_1, a_2 \otimes b_2 \rangle = b_1^* \sigma(a_1^* a_2) b_2$.

Moreover, a direct computation shows the following.

Lemma 2.4.4. *Let $\sigma : \mathcal{A} \to \mathcal{B}$ be completely positive. Let ξ be the vector $1 \otimes 1$ in $\mathcal{A} \otimes_\sigma \mathcal{B}$. Then $\sigma(a) = \langle \xi, a\xi \rangle$. In particular, a map σ is completely positive if and only if it can be expressed as $\sigma(a) = \langle \xi, a\xi \rangle$ for some vector ξ in a Hilbert $\mathcal{A}$-$\mathcal{B}$-correspondence $\mathcal{H}$.*

Finally, let us point out, that just as in the case of states, completely positive maps are automatically bounded (in fact, completely bounded).

Lemma 2.4.5. *Let $\sigma : \mathcal{A} \to \mathcal{B}$ be completely positive. If $A \in M_n(\mathcal{A})$, then $\|\sigma^{(n)}(A)\| \leq \|\sigma(1)\| \|A\|$.*

Proof. First, consider $a \in \mathcal{A}$ (for the case $n = 1$). Let $\mathcal{H} = \mathcal{A} \otimes_\sigma \mathcal{B}$ and $\xi = 1 \otimes 1$. Then by Cauchy-Schwarz,

$$\|\sigma(a)\| = \|\langle \xi, a\xi \rangle\| \leq \|\xi\| \|a\xi\| \leq \|a\| \|\xi\|^2 = \|a\| \|\sigma(1)\|.$$

For $n > 1$, note that if σ is completely positive, then $\sigma^{(n)}$ is also completely positive and hence by the preceding argument $\|\sigma^{(n)}(A)\| \leq \|A\| \|\sigma^{(n)}(1)\| = \|A\| \|\sigma(1)\|$. □

2.5 $\mathcal{B}$-valued probability spaces

Completely positive maps are the $\mathcal{B}$-valued analogue of positive linear functionals on C*-algebras and measures on compact Hausdorff spaces. The analogue of a state or probability measure is a *$\mathcal{B}$-valued expectation*, also known as a *conditional expectation* $\mathcal{A} \to \mathcal{B}$. The probabilistic viewpoint on this construction is due largely to [Voi85, §5].

Definition 2.5.1. Let $\mathcal{B} \subseteq \mathcal{A}$ be unital C*-algebras. An *$\mathcal{B}$-valued expectation* $E : \mathcal{A} \to \mathcal{B}$ is a unital, positive $\mathcal{B}$-$\mathcal{B}$-bimodule map.

By "unital" we mean that $E[1] = 1$ and by "$\mathcal{B}$-$\mathcal{B}$-bimodule map," we mean that $E[ba] = bE[a]$ and $E[ab] = E[a]b$ for $b \in \mathcal{B}$ and $a \in \mathcal{A}$. The unital condition is the analogue of the normalization of a state or probability measure, and the $\mathcal{B}$-$\mathcal{B}$-bimodule property is the analogue of the property that $E[f(X)g(Y)|X] = f(X)E[g(Y)|X]$ in classical probability theory.

Remark 2.5.2. Complete positivity is automatic in this case. Indeed, if $\mathcal{B} \subseteq \mathcal{A}$ and $\Phi : \mathcal{A} \to \mathcal{B}$ is a positive $\mathcal{B}$-$\mathcal{B}$-bimodule map, then Φ is completely positive. To see this, consider a positive element $A^*A \in M_n(\mathcal{A})$ and given $v \in M_{1\times n}(\mathcal{B})$. Then we have

$$v^* E^{(n)}[A^*A]v = E[v^*A^*Av] \geq 0$$

since $v^*A^*Av \geq 0$ in $\mathcal{A}$ by Lemma 2.1.10. Since $v^*E^{(n)}[A^*A]v \geq 0$ for every v, Lemma 2.1.10 implies that $E^{(n)}[A^*A] \geq 0$.

Definition 2.5.3. A *$\mathcal{B}$-valued probability space* is a pair $(\mathcal{A}, E)$, where $\mathcal{A}$ is a C*-algebra unitally containing $\mathcal{B}$ (in other words, the inclusion $\mathcal{B} \to \mathcal{A}$ is specified as part of the data, though suppressed in the notation), and $E : \mathcal{A} \to \mathcal{B}$ is a conditional expectation, such that the representation of $\mathcal{A}$ on $\mathcal{A} \otimes_E \mathcal{B}$ is faithful, that is, the $*$-homomorphism $\mathcal{A} \to B(\mathcal{A} \otimes_E \mathcal{B})$ is injective.

This last condition that the representation is faithful is a type of non-degeneracy condition. For example, in the case where $\mathcal{B} = \mathbb{C}$ and $\mathcal{A} = C(X)$ for a compact Hausdorff space X and E is integration against a given probability measure P, the faithfulness condition says that (closed) support of P in X is all of X. In general, this condition says intuitively that all information about the algebra $\mathcal{A}$ can be captured from the expectation E. This is a reasonable assumption because in non-commutative probability theory, we only care about aspects of the algebra $\mathcal{A}$ that are observable from E.

We have seen that a completely positive map $\mathcal{A} \to \mathcal{B}$ can always be represented as $a \mapsto \langle \xi, a\xi \rangle$ for a vector ξ in an $\mathcal{A}$-$\mathcal{B}$-correspondence. To facilitate construction of non-commutative probability spaces, we present the following characterization of when $a \mapsto \langle \xi, a\xi \rangle$ is a conditional expectation in terms of the vector ξ, which comes from [Liu18, Lemma 2.10].

Definition 2.5.4. Let $\mathcal{H}$ be a Hilbert $\mathcal{B}$-$\mathcal{B}$-correspondence. A vector $\xi \in \mathcal{H}$ is said to be a *unit vector* if $\langle \xi, \xi \rangle = 1$. We say that ξ is *$\mathcal{B}$-central* if $b\xi = \xi b$ for $b \in \mathcal{B}$.

Lemma 2.5.5. *Let $\mathcal{B} \subseteq \mathcal{A}$ be a unital inclusion of unital* C*-algebras, let $\mathcal{H}$ be an $\mathcal{A}$-$\mathcal{B}$-correspondence, and let $\xi \in \mathcal{H}$. Then the following are equivalent:*

(1) The map $E : \mathcal{A} \to \mathcal{B} : a \mapsto \langle \xi, a\xi \rangle$ is a conditional expectation.

(2) $\langle \xi, b\xi \rangle = b$ for all $b \in \mathcal{B}$.

(3) ξ is a $\mathcal{B}$-central unit vector.

Proof. If (1) holds, then E is a unital $\mathcal{B}$-$\mathcal{B}$-bimodule map and hence $E|_{\mathcal{B}} = \mathrm{id}_{\mathcal{B}}$, which means exactly (2).

Suppose (2) holds. By substituting $b = 1$ in (2), we see that ξ is a unit vector. Next, to show that ξ is $\mathcal{B}$-central, fix $b \in \mathcal{B}$. Then we have

$$\begin{aligned}\langle b\xi - \xi b, b\xi - \xi b \rangle &= \langle b\xi, b\xi \rangle - \langle \xi b, b\xi \rangle - \langle b\xi, \xi b \rangle + \langle \xi b, \xi b \rangle \\ &= \langle \xi, b^* b\xi \rangle - b^* \langle \xi, b\xi \rangle - \langle \xi, b^* \xi \rangle b + b^* \langle \xi, \xi \rangle b \\ &= b^* b - b^* b - b^* b + b^* b = 0,\end{aligned}$$

hence $b\xi = \xi b$.

Finally, if (3) holds, then E is unital since $E[1] = \langle \xi, \xi \rangle = 1$ and it is a $\mathcal{B}$-$\mathcal{B}$-bimodule map since

$$E[b_1 a b_2] = \langle \xi, b_1 a b_2 \xi \rangle = \langle b_1^* \xi, a b_2 \xi \rangle = \langle \xi b_1^*, a \xi b_2 \rangle = b_1 \langle \xi, a\xi \rangle b_2 = b_1 E[a] b_2. \qquad \square$$

Remark 2.5.6. The previous lemma implies that if $\mathcal{B} \subseteq \mathcal{A}$ and $\Phi : \mathcal{A} \to \mathcal{B}$ is completely positive, then Φ is a $\mathcal{B}$-$\mathcal{B}$-bimodule map if and only if $\Phi|_{\mathcal{B}} = \mathrm{id}_{\mathcal{B}}$. This is because every such Φ can be realized by a vector ξ through the GNS construction.

We also remark that if ξ is a $\mathcal{B}$-central unit vector, then $\mathcal{B}\xi$ is automatically a direct summand of $\mathcal{H}$.

Lemma 2.5.7. *Let $\mathcal{H}$ be a $\mathcal{B}$-$\mathcal{B}$-correspondence and let $\xi \in \mathcal{H}$ be an $\mathcal{B}$-central unit vector.*

(1) $\mathcal{B}\xi$ and $\mathcal{H}^\circ := \{\zeta : \langle \xi, \zeta \rangle = 0\}$ are $\mathcal{B}$-$\mathcal{B}$-correspondences.

(2) $\mathcal{H} = \mathcal{B}\xi \oplus \mathcal{H}^\circ$.

(3) $\mathcal{B}\xi$ is isomorphic as a $\mathcal{B}$-$\mathcal{B}$-correspondence to the bimodule $\mathcal{B}$ with the inner product given by $\langle b_1, b_2 \rangle = b_1^* b_2$.

Proof. Note that $\mathcal{B}\xi$ is an $\mathcal{B}$-$\mathcal{B}$-bimodule because it is a left $\mathcal{B}$-module and $b\xi = \xi b$. Moreover, $\mathcal{H}^\circ$ is a $\mathcal{B}$-$\mathcal{B}$-bimodule because if $\zeta \in \mathcal{H}^\circ$ and $b \in \mathcal{B}$, then

$$\langle \xi, \zeta b \rangle = \langle \xi, \zeta \rangle b = 0$$

and

$$\langle \xi, b\zeta \rangle = \langle b^* \xi, \zeta \rangle = \langle \xi b^*, \zeta \rangle = b \langle \xi, \zeta \rangle = 0.$$

Moreover, any $\zeta \in \mathcal{H}$ can be written as

$$\zeta = \langle \xi, \zeta \rangle \xi + (\zeta - \langle \xi, \zeta \rangle \xi),$$

where the first term is in $\mathcal{B}\xi$ and the second term is in $\mathcal{H}^\circ$ because

$$\langle \xi, \zeta - \langle \xi, \zeta \rangle \xi \rangle = \langle \xi, \zeta - \xi \langle \xi, \zeta \rangle \rangle = \langle \xi, \zeta \rangle - \langle \xi, \xi \rangle \langle \xi, \zeta \rangle = 0.$$

Therefore, $\mathcal{H} = \mathcal{B}\xi \oplus \mathcal{H}^\circ$. Because this direct sum decomposition holds, the individual terms $\mathcal{B}\xi$ and $\mathcal{H}^\circ$ must be closed subspaces and hence are $\mathcal{B}$-$\mathcal{B}$-correspondences.

Finally, we can define a map $\phi : \mathcal{B} \to \mathcal{B}\xi$ by $b \mapsto b\xi$. This map is clearly surjective. Using the fact that ξ is $\mathcal{B}$-central, one checks that this map preserves the inner product and is an isomorphism of $\mathcal{B}$-$\mathcal{B}$-correspondences. □

2.6 $\mathcal{B}$-valued laws and generalized laws

We now turn to the definition of $\mathcal{B}$-valued laws (and generalized laws). The results of this section are based on [Voi95, PV13, AW16].

Recall that if X is a real random variable on a probability space (Ω, P), then the *law of* X is the measure μ_X on $\mathbb{R}$ given by $\int f \, d\mu_X = E[f(X)]$. Similarly, if $\phi : \mathcal{A} \to \mathbb{C}$ is a state and $X \in \mathcal{A}$ is self-adjoint, then the *law of* X *with respect to* ϕ is the measure μ_X given by $\int f \, d\mu_X = \phi[f(X)]$. In either case, if the measure μ_X is compactly supported, then it is uniquely specified by its moments $\int t^n \, d\mu_X(t) = \phi(X^n)$, that is, it suffices to consider polynomial test functions. In the $\mathcal{B}$-valued setting, there is no clear way to express these moments in terms of a measure, we will simply *define* the law of X by its action on polynomial test functions.

Definition 2.6.1. We denote by $\mathcal{B}\langle X \rangle$ the algebra of non-commutative polynomials in a formal variable X with coefficients in $\mathcal{B}$, that is, the universal (non-commutative) algebra generated by $\mathcal{B}$ and an indeterminate X. (As a vector space, $\mathcal{B}\langle X \rangle$ is the linear span of terms of the form $b_0 X b_1 X \dots b_{k-1} X b_k$.) We endow $\mathcal{A}\langle X \rangle$ with the $*$-operation determined by

$$(b_0 X b_1 X \dots b_{k-1} X b_k)^* = b_k^* X b_{k-1}^* \dots X b_1^* X b_0^*.$$

Definition 2.6.2. Let $(\mathcal{A}, E)$ be a $\mathcal{B}$-valued probability space and x a self-adjoint element of $\mathcal{A}$. The *law of* x is the map $\mu_x : \mathcal{B}\langle X \rangle \to \mathcal{B}$ given by $p(X) \mapsto E[p(x)]$.

In probability theory, it is a standard fact that every probability measure on $\mathbb{R}$ is the law of some random variable. Indeed, the random variable given by the identity function on the probability space $(\mathbb{R}, \mu)$ will have the law μ. Thus, laws which arise from random variables are characterized abstractly as measures. In operator-valued non-commutative probability, there is also an abstract characterization of laws, and a way to explicitly construct a random variable which realizes a given law, which is a version of the GNS construction.

Definition 2.6.3. An *$\mathcal{B}$-valued law* is a linear map $\mu : \mathcal{B}\langle X \rangle \to \mathcal{B}$ such that

(1) *μ is completely positive:* For any $P(X) \in M_n(\mathcal{B}\langle X \rangle)$ we have $\mu^{(n)}(P(X)^* P(X)) \geq 0$ in $M_n(\mathcal{B})$.

(2) *μ is exponentially bounded:* There exist some $M > 0$ and $R > 0$ such that

$$\|\mu(b_0 X b_1 X \dots b_{k-1} X b_k)\| \leq M R^k \|b_0\| \dots \|b_k\| \text{ for all } b_0, \dots, b_k \in \mathcal{B}.$$

(3) *μ is unital:* $\mu(1) = 1$.

(4) *μ is a $\mathcal{B}$-$\mathcal{B}$-bimodule map:* $\mu(b_1 p(X) b_2) = b_1 \mu(p(X)) b_2$ for $b_1, b_2 \in \mathcal{B}$.

Definition 2.6.4. Let $\mu : \mathcal{B}\langle X \rangle \to \mathcal{B}$. If $\|\mu(b_0 X b_1 X \dots b_{k-1} X b_k)\| \leq M R^k \|b_0\| \dots \|b_k\|$, then we say that R is an *exponential bound for μ*. Finally, we define the *radius of μ* as

$$\operatorname{rad}(\mu) := \inf\{R : R \text{ is an exponential bound for } \mu\}.$$

The characterization of $\mathcal{B}$-valued laws is proved in [PV13, Proposition 1.2], and it is similar to earlier results such as [AGZ09, Proposition 5.2.14].

Theorem 2.6.5. *A linear map $\mu : \mathcal{B}\langle X \rangle \to \mathcal{B}$ is a $\mathcal{B}$-valued law if and only if there exists a $\mathcal{B}$-valued probability space $(\mathcal{A}, E)$ and a self-adjoint $x \in \mathcal{A}$ with $\mu = \mu_x$. Moreover, for each μ, we can choose x such that $\|x\| = \operatorname{rad}(\mu)$.*

The most substantial part of the proof will work in greater generality, and we will need the more general result later when we work with analytic transforms associated to $\mathcal{B}$-valued laws.

Theorem 2.6.6. *Let $\mathcal{B}$ and $\mathcal{C}$ be C^*-algebras and $\sigma : \mathcal{B}\langle X \rangle \to \mathcal{C}$ a linear map. Then the following are equivalent:*

(1) σ is completely positive and exponentially bounded.

(2) There exists a $\mathcal{B}$-$\mathcal{C}$-correspondence $\mathcal{H}$, a vector $\xi \in \mathcal{H}$, and a self-adjoint operator $x \in B(\mathcal{H})$ with $\sigma(f(X)) = \langle \xi, f(x)\xi \rangle$ for all $f \in \mathcal{B}\langle X \rangle$.

(3) There exists a C^*-algebra $\mathcal{A}$ and a $*$-homomorphism $\rho : \mathcal{B}\langle X \rangle \to \mathcal{A}$, and a completely positive map $\overline{\sigma} : \mathcal{A} \to \mathcal{C}$ such that $\sigma = \overline{\sigma} \circ \rho$.

Moreover, for a completely positive and exponentially bounded σ, the operator x can be chosen such that $\|x\| = \operatorname{rad}(\sigma)$, and we have

$$\|\sigma(b_0 X b_1 \dots X b_k)\| \leq \|\sigma(1)\| \operatorname{rad}(\sigma)^k \|b_0\| \dots \|b_k\|. \tag{2.1}$$

Proof. (1) $\implies$ (2). We define a $\mathcal{C}$-valued pre-inner-product on $\mathcal{B}\langle X\rangle \otimes_{\text{alg}} \mathcal{C}$ by

$$\langle f_1(X) \otimes c_1, f_2(X) \otimes c_2 \rangle_\mu = c_1^* \mu(f_1(X)^* g_2(X)) c_2.$$

As we saw earlier with tensor products and the GNS construction, the complete positivity of σ implies that the pre-inner-product is nonnegative. Therefore, the Cauchy-Schwarz inequality holds and we can define the separation-completion with respect to this pre-inner-product, which we denote by $\mathcal{B}\langle X\rangle \otimes_\sigma \mathcal{C}$.

The space $\mathcal{H} := \mathcal{B}\langle X\rangle \otimes_\sigma \mathcal{C}$ is a $\mathcal{B}$-$\mathcal{C}$-correspondence with the left action of $\mathcal{B}$ defined in the natural way. Indeed, to show that the left multiplication by $b \in \mathcal{B}$ extends to the separation completion, it suffices to show that $\|b\xi\| \leq \|b\|\|\xi\|$ for $\xi \in \mathcal{B}\langle X\rangle \otimes_{\text{alg}} \mathcal{C}$. This is done by writing $\|b\|^2 - b^*b = y^*y$ for some $y \in \mathcal{B}$ as before.

Next, we claim that the linear operator given by $f(X) \otimes c \mapsto Xf(X) \otimes c$ on the algebraic tensor product passes to a well-defined and bounded operator x on the separation-completion $\mathcal{B}\langle X\rangle \otimes_\sigma \mathcal{C}$. Let R be an exponential bound for σ and let $T > R$. Unfortunately, we cannot claim that $T^2 - X^2$ is a positive element of $\mathcal{B}\langle X\rangle$, or that it can be written as $g(X)^* g(X)$ for some $g(X) \in \mathcal{B}\langle X\rangle$, since $\mathcal{B}\langle X\rangle$ does not have the same completeness properties as a C*-algebra. However, we can fix this problem by looking at a certain power-series completion of $\mathcal{B}\langle X\rangle$ and defining $g(X)$ as the power series for $\sqrt{T^2 - X^2}$.

For a monomial $b_0 X b_1 \dots X b_k$, we denote

$$\mathfrak{p}(b_0 X b_1 \dots X b_k) = R^k \|b_0\| \dots \|b_k\|.$$

Then for $f(X) \in \mathcal{B}\langle X\rangle$, we define

$$\|f(X)\|_R = \inf \left\{ \sum_{j=1}^n \mathfrak{p}(f_j) : f_j \text{ monomials and } f = \sum_{j=1}^n f_j \right\}.$$

Let $\mathcal{B}\langle\langle X\rangle\rangle_R$ be the completion of $\mathcal{B}\langle X\rangle$ in this norm. One checks easily that

$$\|f(X)g(X)\|_R \leq \|f(X)\|_R \|g(X)\|_R,$$

and this inequality extends to the completion, which makes $\mathcal{B}\langle\langle X\rangle\rangle_R$ a Banach algebra. Similarly, the $*$-operation on $\mathcal{B}\langle X\rangle$ extends to the completion. By standard results from complex analysis, the function $g(t) = \sqrt{T^2 - t^2}$ has a power series expansion

$$g(t) = \sum_{j=0}^\infty \alpha_j t^j$$

which converges for $|t| < T$. In particular, the series converges absolutely for $t = R$, which implies that

$$g(X) = \sum_{j=0}^{\infty} \alpha_j X^j$$

converges absolutely in $\mathcal{B}\langle\langle X\rangle\rangle_R$. Moreover, because of the absolute convergence and the Banach algebra properties, we can compute $\psi(X)^2$ by multiplying the series term by term and hence conclude that $g(X)^2 = T^2 - X^2$. Because R is an exponential bound for σ, we know that $\|\sigma(f(X))\| \leq M\|f(X)\|_R$, where M is a constant such that $\|b_0Xb_1 \dots Xb_k\| \leq MR^k\|b_0\| \dots \|b_k\|$. Hence, σ extends to a linear map $\mathcal{B}\langle\langle X\rangle\rangle_R \to \mathcal{C}$, which is still completely positive, and hence the pre-inner-product $\langle \cdot, \cdot \rangle$ extends to $\mathcal{B}\langle\langle X\rangle\rangle_R \otimes_{\text{alg}} \mathcal{C}$. Then for each vector $\zeta \in \mathcal{B}\langle\langle X\rangle\rangle_R \otimes_{\text{alg}} \mathcal{C}$, we have

$$\langle \zeta, (T^2 - X^2)\xi \rangle = \langle g(X)\zeta, g(X)\zeta \rangle \geq 0,$$

which implies that $\|X\zeta\| \leq T\|\zeta\|$, and in particular, this holds for $\zeta \in \mathcal{B}\langle X\rangle \otimes_{\text{alg}} \mathcal{C}$. By taking $T \searrow R$, we have $\|X\zeta\| \leq R\|\zeta\|$, which means that the multiplication operator by X is bounded with respect to the pre-inner-product and hence extends to the separation-completion.

The operator x thus defined is clearly self-adjoint. Moreover, letting $\xi = 1 \otimes 1 \in \mathcal{B}\langle X\rangle \otimes_\sigma \mathcal{C}$, we have

$$\langle \xi, f(x)\xi \rangle = \sigma(f(X)) \text{ for } f \in \mathcal{B}\langle X\rangle.$$

Since R was an arbitrary exponential bound, we have $\|x\| \leq \operatorname{rad}(\sigma)$, and thus,

$$\|\sigma(b_0Xb_1 \dots Xb_k)\| = \|\langle \xi, b_0xb_1 \dots xb_k\xi \rangle\| \leq \|\xi\|^2\|x\|^k\|b_0\| \dots \|b_k\| \leq \|\sigma(1)\| \operatorname{rad}(\sigma)^k\|b_0\| \dots \|b_k\|.$$

This proves (1) $\implies$ (2) as well as the last claim of the theorem.

(2) $\implies$ (3). Fix $\mathcal{H}$, ξ, and x as in (2), and let $\mathcal{A} = B(\mathcal{H})$. Let π be the $*$-homomorphism $\mathcal{B} \to B(\mathcal{H})$ given by the left $\mathcal{B}$-module structure. Then there is a unique $*$-homomorphism $\rho : \mathcal{B}\langle X\rangle \to B(\mathcal{H})$ satisfying $\rho|_{\mathcal{B}} = \pi$ and $\rho(X) = x$. Moreover, the map $\overline{\sigma} : B(\mathcal{H}) \to \mathcal{C}$ given by $a \mapsto \langle \xi, a\xi \rangle$ is completely positive and satisfies $\overline{\sigma} \circ \rho = \sigma$.

(3) $\implies$ (1). If (3) holds, then σ is completely positive because it is the composition of the two completely positive maps ρ and $\overline{\sigma}$. Moreover, it is exponentially bounded because

$$\|\sigma(b_0Xb_1 \dots Xb_k)\| = \|\overline{\sigma}(b_0\rho(X)b_1 \dots \rho(X)b_k)\| \leq \|\sigma(1)\|\|\rho(X)\|^k\|b_0\| \dots \|b_k\|$$

by Lemma 2.4.5. □

Proof of Theorem 2.6.5. Suppose that μ is a $\mathcal{B}$-valued law. Let $\mathcal{H}$, ξ, and x be as in Theorem 2.6.6 (2) for $\sigma = \mu$. Since $\langle \xi, b\xi \rangle = \mu(b) = b$, Lemma 2.5.5 implies that ξ is a $\mathcal{B}$-central unit vector, hence $\mathcal{A} = B(\mathcal{H})$ and $E = \langle \xi, (\cdot)\xi \rangle$ form a $\mathcal{B}$-valued probability space. And clearly $\mu_x = \mu$. Conversely, if $\mu = \mu_x$ for some x in a $\mathcal{B}$-valued probability space $(\mathcal{A}, E)$, then by Theorem 2.6.5, μ is completely positive and exponentially bounded. Since $E(1) = 1$ and E is a $\mathcal{B}$-$\mathcal{B}$-bimodule map, the same holds for μ. □

From the use of Lemma 2.5.5 above, the following corollary is obvious.

Corollary 2.6.7. *Let* $\sigma : \mathcal{B}\langle X\rangle \to \mathcal{B}$ *be completely positive and exponentially bounded. Then* σ *is a law if and only if* $\sigma|_{\mathcal{B}} = \mathrm{id}_{\mathcal{B}}$.

Lemma 2.6.8. *If* σ *and* τ *are generalized laws, then* $\sigma + \tau$ *is also a generalized law, and we have* $\mathrm{rad}(\sigma + \tau) = \max(\mathrm{rad}(\sigma), \mathrm{rad}(\tau))$.

Proof. It is straightforward to check from the definition that $\sigma+\tau$ is a generalized law. Next, if $p(X) = b_0 X b_1 \dots X b_k$, then

$$\begin{aligned}
\|(\sigma + \tau)(p(X))\| &\leq \|\sigma(p(X))\| + \|\tau(p(X))\| \\
&\leq \|\sigma(1)\| \,\mathrm{rad}(\sigma)^k \|b_0\| \dots \|b_k\| + \|\tau(1)\| \,\mathrm{rad}(\tau)^k \|b_0\| \dots \|b_j\| \\
&\leq (\|\sigma(1)\| + \|\tau(1)\|) \max(\mathrm{rad}(\sigma), \mathrm{rad}(\tau))^k \|b_0\| \dots \|b_k\|,
\end{aligned}$$

which shows that $\mathrm{rad}(\sigma + \tau) \leq \max(\mathrm{rad}(\sigma), \mathrm{rad}(\tau))$. On the other hand,

$$\begin{aligned}
\|\sigma(p(X))\| &\leq \|\sigma(1)\|^{1/2} \|\sigma(p(X)^* p(X))\|^{1/2} \\
&\leq \|(\sigma + \tau)(1)\|^{1/2} \|(\sigma + \tau)(p(X)^* p(X))\|^{1/2} \\
&\leq \|(\sigma + \tau)(1)\|^{1/2} \left(\|(\sigma + \tau)(1)\| \,\mathrm{rad}(\sigma + \tau)^{2k} \|b_0\|^2 \dots \|b_k\|^2 \right)^{1/2} \\
&= \|(\sigma + \tau)(1)\| \,\mathrm{rad}(\sigma + \tau)^k \|b_0\| \dots \|b_k\|,
\end{aligned}$$

where we have used the fact that $0 \leq \sigma(p(X)^* p(X)) \leq (\sigma + \tau)(p(X)^* p(X))$. But the above estimate implies that $\mathrm{rad}(\sigma) \leq \mathrm{rad}(\sigma + \tau)$, and of course $\mathrm{rad}(\tau) \leq \mathrm{rad}(\sigma + \tau)$ by symmetry. □

CHAPTER 3

Background: Fully matricial functions

3.1 Introduction

One of the key tools in scalar-valued non-commutative probability is the Cauchy-Stieltjes transform of a random variable X given by

$$G_X(z) = E[(z - X)^{-1}],$$

which is an complex-analytic function for z in the upper half-plane and in a neighborhood of ∞ (provided that X is bounded). The law of X can be recovered from the power series expansion of G_X at ∞ because

$$G_X(z^{-1}) = E[(z^{-1} - X)^{-1}] = \sum_{k=0}^{\infty} z^{k+1} E[X^k],$$

which is essentially the moment generating function for the law of X.

In this chapter, we describe a $\mathcal{B}$-valued analytic function theory suitable for $\mathcal{B}$-valued non-commutative probability, and in the next, we analyze the $\mathcal{B}$-valued Cauchy-Stieltjes transform. It should not be surprising at this point that our notion of analyticity needs to take into account matrix amplifications. One concrete motivation for this is that, without taking matrix amplifications, the Cauchy-Stieltjes transform is insufficient to encode the $\mathcal{B}$-valued law of a random variable X.

One would naïvely define the Cauchy-Stieltjes transform G_X as a function on an open subset of $\mathcal{B}$ given by $G_X(z) = E[(z - X)^{-1}]$. Looking at the power series of $G_X(z^{-1})$ at 0, we have

$$G_X(z^{-1}) = E[(z^{-1} - X)^{-1}] = E[(1 - zX)^{-1} z] = \sum_{k=0}^{\infty} E[(zX)^k z].$$

From this, we can recover all moments of the form $E[zXz \dots Xz]$. However, to know the law of X, we would need to consider all moments of the form $E[z_1 X z_2 \dots X z_k]$. Of course, for the Cauchy-Stieltjes transform *not* to encode the law of X would severely handicap analytic methods for operator-valued non-commutative probability.

But fortunately this problem is resolved by matrix amplification. We can consider the sequence of functions $G_X^{(n)}$ with domain in $M_n(\mathcal{B})$ given by $G_X^{(n)}(z) = E^{(n)}[(z - X^{(n)})^{-1}]$, where

$X^{(n)}$ is the diagonal matrix with entries given by X. To recover the moment $E[z_1 X z_2 \dots X z_n]$ for $z_j \in \mathcal{B}$, we evaluate (the analytic extension of) $G_X^{(n+2)}(z^{-1})$ on the matrix

$$z = \begin{bmatrix} 0 & z_0 & 0 & \dots & 0 & 0 \\ 0 & 0 & z_1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & z_n \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

and obtain

$$\begin{aligned} G_X^{(n+2)}(z^{-1}) &= \sum_{k=0}^{\infty} E^{(n+2)}[(zX^{(n+2)})^k z] \\ &= \begin{bmatrix} 0 & z_0 & E[z_0 X z_1] & \dots & E[z_0 X z_1 \dots X z_{n-1}] & E[z_0 X z_1 \dots X z_n] \\ 0 & 0 & z_1 & \dots & E[z_1 X z_2 \dots X z_{n-1}] & E[z_1 X z_2 \dots X z_n] \\ 0 & 0 & 0 & \dots & E[z_2 X z_3 \dots X z_{n-1}] & E[z_2 X z_3 X \dots X z_n] \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & z_n \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}, \end{aligned}$$

where $E[z_0 X z_1 \dots X z_n]$ can be recovered as the top right entry.

Thus, an analytic function F ought to be a sequence of functions $F^{(n)}$ defined on $n \times n$ matrices over $\mathcal{B}$. But we also need to guarantee that these functions "fit together consistently." More precisely, we will require that F respects direct sums and conjugation by invertible scalar matrices (see Definition 3.2.3).

Remarkably, these algebraic conditions, together with a local boundedness condition which is uniform in n, are sufficient to imply the existence of local power series expansions for the function $F^{(n)}$. The terms in these power series expansions are given by multilinear forms, much like the power series expansion for $G_X(z^{-1})$ is obtained from the multilinear forms $\mu(z_0 X z_1 \dots X z_n)$. Moreover, just as in the case of $G_X(z^{-1})$, these multilinear forms are computed by evaluating $F^{(n)}$ on certain upper triangular matrices.

The study of such *non-commutative* or *fully matricial* functions originated in the 1970's with the work of Joseph Taylor [Tay72], [Tay73]. Dan Voiculescu studied fully matricial functions in the context of the free difference quotient and generalized resolvents [Voi00], [Voi04], [Voi10]. Mihai Popa and Victor Vinnikov clarified the connection between fully matricial function theory in the abstract and the various analytic transforms associated to non-commutative laws [PV13], which we will discuss in detail in the later chapters.

We have opted for a self-contained development of the theory of fully matricial functions, though somewhat restricted in scope. We are indebted to the systematic work of Kaliuzhnyi-Verbovetskyi and Vinnikov [KV14], although we have not presented the material in the same way. We write with the analogy to complex analysis always in mind, and with an eye towards the results of Williams and Anshelevich on the Cauchy-Stieltjes transform [Wil17], [AW16], which we will discuss in the next chapter. We follow Voiculescu in using the term "fully matricial" rather than "non-commutative" since it gives a more concrete description of the definition.

3.2 Fully matricial domains and functions

In order to state the definition, we use the following notation.

(1) We identify $M_n(\mathcal{B})$ with $\mathcal{B} \otimes M_n(\mathbb{C})$.

(2) If $z \in M_n(\mathcal{B})$, then we denote

$$z^{(m)} = z \otimes 1_m = \begin{bmatrix} z & 0 & \dots & 0 & 0 \\ 0 & z & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & z & 0 \\ 0 & 0 & \dots & 0 & z \end{bmatrix} \in M_{nm}(\mathcal{B}).$$

(3) If $z \in M_n(\mathcal{B})$ and $w \in M_m(\mathcal{B})$, then we denote

$$z \oplus w = \begin{bmatrix} z & 0 \\ 0 & w \end{bmatrix} \in M_{n+m}(\mathcal{B}).$$

(4) If $z \in M_n(\mathcal{B})$, then we denote $B^{(n)}(z, r) = \{w \in M_n(\mathcal{B}) : \|z - w\| < r\}$.

Definition 3.2.1. A *fully matricial domain* Ω *over* $\mathcal{B}$ is a sequence of sets $\Omega^{(n)} \subseteq \mathcal{B}^{(n)}$ satisfying the following conditions.

(1) Ω *respects direct sums:* If $z \in \Omega^{(n)}$ and $w \in \Omega^{(m)}$, then $z \oplus w \in \Omega^{(n+m)}$.

(2) Ω *is uniformly open:* If $z \in \Omega^{(n)}$, then there exists $r > 0$ such that $B^{(nm)}(z^{(m)}, r) \subseteq \Omega^{(nm)}$ for all m.

(3) Ω *is non-empty:* $\Omega^{(n)}$ is non-empty for some n.

Notation 3.2.2. We denote by $M_\bullet(\mathcal{B})$ the fully matricial domain $(M_n(\mathcal{B}))_{n \in \mathbb{N}}$.

Definition 3.2.3. Let Ω_1 and Ω_2 be fully matricial domains over $\mathcal{B}_1$ and $\mathcal{B}_2$ respectively. A fully matricial function $F : \Omega_1 \to \Omega_2$ is a sequence of functions $F^{(n)} : \Omega_1^{(n)} \to \Omega_2^{(n)}$ satisfying the following conditions.

(1) *F respects intertwinings:* Suppose that $z \in \Omega_1^{(n)}$, $w \in \Omega_1^{(m)}$, $T \in M_{n\times m}(\mathbb{C})$. If $zT = Tw$, then $F^{(n)}(z)S = TF^{(m)}(w)$.

(2) *F is uniformly locally bounded:* For each $x \in \Omega_1^{(n)}$, there exist r and $M > 0$ such that $B^{(nm)}(z^{(m)}, r) \subseteq \Omega_1^{(nm)}$ and $F^{(nm)}(B^{(nm)}(z^{(m)}, r)) \subseteq B^{(nm)}(0, M)$ for all m.

In order to check that a function F is fully matricial, it is often convenient to use the following equivalent characterization of the intertwining condition.

Lemma 3.2.4. *Let Ω_1 and Ω_2 be fully matricial domains and let $F : \Omega_1 \to \Omega_2$ be a sequence of functions. Then F respects intertwinings if and only if the following conditions hold.*

(1) F respects direct sums: *If $z \in \Omega^{(n)}$ and $w \in \Omega^{(m)}$, then $F^{(n+m)}(z \oplus w) = F^{(n)}(z) \oplus F^{(m)}(w)$.*

(2) F respects similarities: *Suppose that $z \in \Omega^{(n)}$, that $S \in M_n(\mathbb{C})$ is invertible, and that $SzS^{-1} \in \Omega^{(n)}$. Then $F^{(n)}(SzS^{-1}) = SF^{(n)}(z)S^{-1}$.*

Proof. First, assume that F respects intertwinings. To prove (1), fix $z \in \Omega^{(n)}$ and $w \in \Omega^{(m)}$. Then we have the block matrix equations

$$\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} z & 0 \\ 0 & w \end{bmatrix} = z \begin{bmatrix} 1 & 0 \end{bmatrix}$$
$$\begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} z & 0 \\ 0 & w \end{bmatrix} = w \begin{bmatrix} 0 & 1 \end{bmatrix}$$

Because F respects intertwinings, we have

$$\begin{bmatrix} 1 & 0 \end{bmatrix} F^{(n+m)}(z \oplus w) = F^{(n)}(z) \begin{bmatrix} 1 & 0 \end{bmatrix}$$
$$\begin{bmatrix} 0 & 1 \end{bmatrix} F^{(n+m)}(z \oplus w) = F^{(m)}(w) \begin{bmatrix} 0 & 1 \end{bmatrix},$$

which together imply that

$$F^{(n+m)}(z \oplus w) = \begin{bmatrix} F^{(n)}(z) & 0 \\ 0 & F^{(m)}(w) \end{bmatrix} = F^{(n)}(z) \oplus F^{(m)}(w).$$

Next, fix z and S as in (2). Let $w = SzS^{-1}$. Then $Sz = wS$ and hence $SF^{(n)}(z) = F^{(n)}(w)S$, which means that $F^{(n)}(SzS^{-1}) = SF^{(n)}(z)S^{-1}$.

Conversely, suppose that (1) and (2) hold and consider an intertwining $zT = Tw$ where $z \in \Omega_1^{(n)}$, $w \in \Omega_1^{(m)}$, $T \in M_{n\times m}(\mathbb{C})$. Then observe that

$$\begin{bmatrix} z & 0 \\ 0 & w \end{bmatrix} \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z & 0 \\ 0 & w \end{bmatrix},$$

and observe that

$$S = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}$$

is invertible. Hence, $S(z \oplus w)S^{-1} = z \oplus w$ and therefore by assumptions (1) and (2), we have $S(F^{(n)}(z) \oplus F^{(m)}(w))S^{-1} = F^{(n)}(z) \oplus F^{(m)}(w)$. In other words,

$$\begin{bmatrix} F^{(n)}(z) & 0 \\ 0 & F^{(m)}(w) \end{bmatrix} \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} F^{(n)}(z) & 0 \\ 0 & F^{(m)}(w) \end{bmatrix},$$

and hence, looking at the top right block, $F^{(n)}(z)T = TF^{(m)}(w)$. □

In order to reduce the number of superscripts cluttering up our paper, we will use the following notation.

(1) For $F : \Omega_1 \to \Omega_2$ and $z \in \Omega_1^{(n)}$, we will usually write $F(z)$ rather than $F^{(n)}(z)$, and the context will make clear the size of the matrix z.

(2) If Ω_1 and Ω_2 are fully matricial domains, then we write $\Omega_1 \subseteq \Omega_2$ to mean that $\Omega_1^{(n)} \subseteq \Omega_2^{(n)}$ for every n.

(3) We write $z \in \Omega$ to mean that $z \in \Omega^{(n)}$ for some n.

(4) For $\Gamma \subseteq \Omega_1$ and $F : \Omega_1 \to \Omega_2$, we denote by $F(\Gamma)$ the sequence of sets $F(\Gamma)^{(n)} = F^{(n)}(\Gamma^{(n)})$. We define $F^{-1}(\Gamma)$ for $\Gamma \subseteq \Omega_2$ similarly.

(5) For $z \in M_n(\mathcal{B})$, we denote by $B(z, r)$ the fully matricial domain

$$B^{(k)}(z, r) = \begin{cases} B^{(nm)}(z^{(m)}, r), & k = nm \\ \varnothing, & \text{otherwise.} \end{cases}$$

In this notation, the uniform openness condition of Definition 3.2.1 states that for every $z \in \Omega$, there exists $r > 0$ such that $B(z, r) \subseteq \Omega$. Moreover, $F : \Omega_1 \to \Omega_2$ is uniformly locally bounded as in Definition 3.2.3 if and only if for every $z \in \Omega_1$, there exist R and M such that $F(B(z, R)) \subseteq B(0, M)$.

3.3 Difference-differential calculus

Definition 3.3.1. Let $F : \Omega_1 \to \Omega_2$ be fully matricial, where Ω_j is fully matricial domain over $\mathcal{B}_j$. Suppose that $z_0 \in \Omega_1^{(n_0)}$, ..., $z_k \in \Omega_1^{(n_k)}$, suppose that $w_1 \in M_{n_0 \times n_1}(\mathcal{B})$, ...,

$w_k \in M_{n_{k-1}\times n_k}(\mathcal{B}_1)$, and suppose that the block matrix

$$Z := \begin{bmatrix} z_0 & w_1 & 0 & \dots & 0 & 0 \\ 0 & z_1 & w_2 & \dots & 0 & 0 \\ 0 & 0 & z_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & z_{k-1} & w_k \\ 0 & 0 & 0 & \dots & 0 & z_k \end{bmatrix}$$

is in $\Omega_1^{(n_0+\cdots+n_k)}$. Then we define

$$\Delta^k F(z_0, \dots, z_k)[w_1, \dots, w_k]$$

as the upper right $n_0 \times n_k$ block of $F(Z)$.

Lemma 3.3.2. *Let $z_0, \dots, z_k$ and $w_1, \dots, w_k$ and Z be as above, and assume that each of the submatrices*

$$Z_{i,j} := \begin{bmatrix} z_i & w_{i+1} & \dots & 0 & 0 \\ 0 & z_{i+1} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & z_{j-1} & w_j \\ 0 & 0 & \dots & 0 & z_j \end{bmatrix}$$

is in the domain of F for each $i < j$. Then we have

$$F(Z) = \begin{bmatrix} F(z_0) & \Delta F(z_0, z_1)[w_1] & \Delta^2 F(z_0, z_1, z_2)[w_1, w_2] & \dots & \Delta^k F(z_0, \dots, z_k)[w_1, \dots, w_k] \\ 0 & F(z_1) & \Delta F(z_1, z_2)[w_2] & \dots & \Delta^{k-1} F(z_1, \dots, z_k)[w_2, \dots, w_k] \\ 0 & 0 & F(z_2) & \dots & \Delta^{k-2} F(z_2, \dots, z_k)[w_3, \dots, w_k] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & F(z_k) \end{bmatrix} \tag{3.1}$$

Proof. We proceed by induction on k with the base case $k = 0$ being trivial. Let $k \geq 1$. Let n_j be the size of the matrix z_j and let $N_{i,j} = n_i + \cdots + n_j$. Then we have

$$Z \begin{bmatrix} 1_{N_{0,k-1}\times N_{0,k-1}} \\ 0_{n_k \times N_{0,k-1}} \end{bmatrix} = \begin{bmatrix} 1_{N_{0,k-1}\times N_{0,k-1}} \\ 0_{n_k \times N_{0,k-1}} \end{bmatrix} Z_{0,k-1}$$

and therefore

$$F(Z) \begin{bmatrix} 1_{N_{0,k-1}\times N_{0,k-1}} \\ 0_{n_k \times N_{0,k-1}} \end{bmatrix} = \begin{bmatrix} 1_{N_{0,k-1}\times N_{0,k-1}} \\ 0_{n_k \times N_{0,k-1}} \end{bmatrix} F(Z_{0,k-1}).$$

From this relation together with the induction hypothesis applied to $F(Z_{0,k-1})$, we deduce that

$$F(Z) = \begin{bmatrix} F(z_0) & \Delta F(z_0, z_1)[w_1] & \dots & \Delta^{k-1} F(z_0, \dots, z_{k-1})[w_1, \dots, w_{k-1}] & * \\ 0 & F(z_1) & \dots & \Delta^{k-2} F(z_1, \dots, z_{k-1})[w_2, \dots, w_{k-1}] & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & * \end{bmatrix}.$$

In other words, (3.1) is verified except in the last n_k columns. Similarly, by considering the intertwining

$$\begin{bmatrix} 0_{N_{1,k} \times n_0} & 1_{N_{1,k} \times N_{1,k}} \end{bmatrix} Z = Z_{1,k} \begin{bmatrix} 0_{N_{1,k} \times n_0} & 1_{N_{1,k} \times N_{1,k}} \end{bmatrix}$$

and applying the induction hypothesis to $F(Z_{1,k})$, we can verify (3.1) except in the first n_k rows. It remains to check (3.1) in the top right $n_0 \times n_k$ block; but this holds by definition of $\Delta^k F$. □

Lemma 3.3.3. *Suppose that $z_0 \in \Omega_1^{(n_0)}$, ..., $z_k \in \Omega_1^{(n_k)}$, suppose that $w_1 \in M_{n_0 \times n_1}(\mathcal{B})$, ..., $w_k \in M_{n_{k-1} \times n_k}(\mathcal{B})$, and let ζ_1, ..., $\zeta_k \in \mathbb{C}$. Then we have*

$$\Delta^k F(z_0, \dots, z_k)[\zeta_1 w_1, \dots, \zeta_k w_k] = \zeta_1 \dots \zeta_k F(z_0, \dots, z_k)[w_1, \dots, w_k]$$

provided that both expressions are defined under Definition 3.3.1.

Proof. We consider the intertwining

$$\begin{bmatrix} z_0 & \zeta_1 w_1 & \dots & 0 & 0 \\ 0 & z_1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & z_{k-1} & \zeta_k w_k \\ 0 & 0 & \dots & 0 & z_k \end{bmatrix} \begin{bmatrix} \zeta_1 \dots \zeta_k & 0 & \dots & 0 & 0 \\ 0 & \zeta_2 \dots \zeta_k & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \zeta_k & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \zeta_1 \dots \zeta_k & 0 & \dots & 0 & 0 \\ 0 & \zeta_2 \dots \zeta_k & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \zeta_k & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} z_0 & w_1 & \dots & 0 & 0 \\ 0 & z_1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & z_{k-1} & w_k \\ 0 & 0 & \dots & 0 & z_k \end{bmatrix},$$

apply the function F, and then examine the top right corner. □

Definition 3.3.4. If $z_j \in \Omega_1^{(n_j)}$ for $j = 0, \dots, k$, then we extend the definition of

$$\Delta^k F(z_0, \dots, z_k)[w_1, \dots, w_k]$$

to arbitrary values of $w_1 \in M_{n_0 \times n_1}(\mathcal{B})$, ..., $w_k \in M_{n_{k-1} \times n_k}(\mathcal{B})$ by setting

$$F(z_0, \dots, z_k)[w_1, \dots, w_k] = \frac{1}{\zeta_1 \dots \zeta_k} \Delta^k F(z_0, \dots, z_k)[\zeta_1 w_1, \dots, \zeta_k w_k],$$

where $\zeta_1, \ldots, \zeta_k \in \mathbb{C} \setminus \{0\}$ are chosen to be sufficiently small that

$$\begin{bmatrix} z_0 & \zeta_1 w_1 & \ldots & 0 & 0 \\ 0 & z_1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & z_{k-1} & \zeta_k w_k \\ 0 & 0 & \ldots & 0 & z_k \end{bmatrix} \in \Omega_1^{(n_0+\cdots+n_k)}.$$

Such a choice of $\zeta_1, \ldots, \zeta_k$ is possible because $\Omega_1^{(n_0+\cdots+n_k)}$ is open. Lemma 3.3.3 guarantees that this definition of $F(z_0, \ldots, z_k)[w_1, \ldots, w_k]$ is independent of the choice of $\zeta_1, \ldots, \zeta_k$ and is consistent with the earlier Definition 3.3.1.

Lemma 3.3.5. *$\Delta^k F(z_0, \ldots, z_k)[w_1, \ldots, w_k]$ is multilinear in $w_1, \ldots, w_k$.*

Proof. We have already shown that $\Delta^k F(z_0, \ldots, z_k)$ behaves correctly when we multiply one of the w_j's by a scalar, so it remains to show that $\Delta^k F(z_0, \ldots, z_k)[w_1, \ldots, w_k]$ is additive in each variable y_j. First, consider the case $k = 1$ in which we must show that

$$\Delta F(z_0, z_1)[w + w'] = \Delta F(z_0, z_1)[w] + \Delta F(z_0, z_1)[w'].$$

Choose $\zeta \in \mathbb{C} \setminus \{0\}$ sufficiently small that the matrices

$$\begin{bmatrix} z_0 & \zeta(w+w') \\ 0 & z_1 \end{bmatrix}, \quad \begin{bmatrix} z_0 & \zeta w \\ 0 & z_1 \end{bmatrix}, \quad \begin{bmatrix} z_0 & \zeta w' \\ 0 & z_1 \end{bmatrix}, \quad \begin{bmatrix} z_0 & 0 & \zeta w \\ 0 & z_0 & \zeta w' \\ 0 & 0 & z_1 \end{bmatrix}$$

are all in the domain of F. From the intertwining

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_0 & 0 & \zeta w \\ 0 & z_0 & \zeta w' \\ 0 & 0 & z_1 \end{bmatrix} = \begin{bmatrix} z_0 & \zeta w \\ 0 & z_1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

we deduce that

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} F\left(\begin{bmatrix} z_0 & 0 & \zeta w \\ 0 & z_0 & \zeta w' \\ 0 & 0 & z_1 \end{bmatrix} \right) = \begin{bmatrix} F(z_0) & \zeta \Delta F(z_0, z_1)[w] \\ 0 & F(z_1) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Similarly,

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} F\left(\begin{bmatrix} z_0 & 0 & \zeta w \\ 0 & z_0 & \zeta w' \\ 0 & 0 & z_1 \end{bmatrix} \right) = \begin{bmatrix} F(z_0) & \zeta \Delta F(z_0, z_1)[w'] \\ 0 & F(z_1) \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and thus altogether,

$$F\left(\begin{bmatrix} z_0 & 0 & \zeta w \\ 0 & z_0 & \zeta w' \\ 0 & 0 & z_1 \end{bmatrix}\right) = \begin{bmatrix} F(z_0) & 0 & \zeta\Delta F(z_0,z_1)[w] \\ 0 & z_0 & \zeta\Delta F(z_0,z_1)[w'] \\ 0 & 0 & z_1 \end{bmatrix}.$$

Finally, we use the intertwining

$$\begin{bmatrix} z_0 & \zeta(w+w') \\ 0 & z_1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} z_0 & 0 & \zeta w \\ 0 & z_0 & \zeta w' \\ 0 & 0 & z_1 \end{bmatrix}$$

to deduce that

$$\begin{bmatrix} F(z_0) & \zeta\Delta F(z_0,z_1)[w+w'] \\ 0 & F(z_1) \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} F(z_0) & 0 & \zeta\Delta F(z_0,z_1)[w] \\ 0 & F(z_0) & \zeta\Delta F(z_0,z_1)[w'] \\ 0 & 0 & F(z_1) \end{bmatrix},$$

which shows that $\zeta\Delta F(z_0,z_1)[w+w'] = \zeta\Delta F(z_0,z_1)[w] + \zeta\Delta F(z_0,z_1)[w']$ as desired.

The argument in the general case is similar. To show linearity of $F(z_0,\dots,z_k)[w_1,\dots,w_n]$ in w_j, we consider replacing w_j by $w_j + w_j'$. The block 3×3 matrix used above is replaced by

$$\begin{bmatrix}
z_0 & \zeta_1 w_1 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\
0 & z_1 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \dots & z_{j-2} & \zeta_{j-1}w_{j-1} & 0 & 0 & \dots & 0 & 0 \\
0 & 0 & \dots & 0 & z_{j-1} & 0 & \zeta_j w_j & \dots & 0 & 0 \\
0 & 0 & \dots & 0 & 0 & z_{j-1} & \zeta_j w_j' & \dots & 0 & 0 \\
0 & 0 & \dots & 0 & 0 & 0 & x_j & \dots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & z_{k-1} & \zeta_k w_k \\
0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & z_k
\end{bmatrix}$$

and the intertwiners are replaced by

$$1_{n_0+\dots+n_{j-1}} \oplus \begin{bmatrix} \alpha & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \oplus 1_{n_{j+1}+\dots+n_k}$$

where α, $\beta = 0, 1$ and where n_j is the size of the matrix z_j. □

3.4 Taylor-Taylor expansion

We have defined the derivative operators $\Delta^k F$ using the matricial structure of F. Now we will show that these same operators $\Delta^k F$ describe the differential and analytic properties of

F, and in fact that F has local power series expansions in terms of $\Delta^k F$. We begin with the finite Taylor-Taylor expansion. This formula is named for Brook Taylor, who developed the Taylor expansion in calculus, and Joseph L. Taylor, who pioneered the theory of non-commutative (fully matricial) functions [Tay72, Tay73].

Lemma 3.4.1 (Taylor-Taylor Formula). *Let $F : \Omega_1 \to \Omega_2$ be a fully matricial function. Let $z_* \in \Omega_1^{(n)}$ and $m > 1$ and suppose that $B^{(nm)}(z_*^{(m)}, r) \subseteq \Omega_1^{(nm)}$. If $z \in B_{r/\sqrt{2}}^{(n)}(z_*)$, then*

$$F(z) = \sum_{k=0}^{m-2} \Delta^k F(z_*, \dots, z_*)[z - z_*, \dots, z - z_*] + \Delta^{m-1} F(z, z_*, \dots, z_*)[z - z_*, \dots, z - z_*],$$

where the $k = 0$ term in the sum is to be interpreted as $F(z_*)$.

Proof. Observe that the $m \times m$ block matrix

$$Z = \begin{bmatrix} z & z - z_* & 0 & 0 & \dots & 0 & 0 \\ 0 & z_* & z - z_* & 0 & \dots & 0 & 0 \\ 0 & 0 & z_* & z - z_* & \dots & 0 & 0 \\ 0 & 0 & 0 & z_* & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & z_* & z - z_* \\ 0 & 0 & 0 & 0 & \dots & 0 & z_* \end{bmatrix}$$

is in $B^{(nm)}(z_*^{(m)}, r) \subseteq \Omega^{(nm)}$ provided that $\|z - z_*\| < r/\sqrt{2}$ (here we put the $\sqrt{2}$ because there are two entries in the top row which need to be changed). We have the intertwining relation

$$\begin{bmatrix} 1 & 1 \dots & 1 \end{bmatrix} Z = z \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix},$$

and therefore,

$$\begin{bmatrix} 1 & 1 \dots & 1 \end{bmatrix} F(Z) = F(z) \begin{bmatrix} 1 & 1 \dots & 1 \end{bmatrix}.$$

Looking at the rightmost block of $\begin{bmatrix} 1 & 1 \dots & 1 \end{bmatrix} F^{(nm)}(Z)$ and applying Lemma 3.3.2, we have

$$F(z_*) + \sum_{k=1}^{m-1} \Delta^k F(z_*, \dots, z_*)[z - z_*, \dots, z - z_*] + \Delta^m F(z, z_*, \dots, z_*)[z - z_*, \dots, z - z_*] = F^{(n)}(z).$$

□

Next, we give a non-commutative analogue of the Cauchy estimates from complex analysis, which will help us prove convergence of the infinite Taylor-Taylor series. In the following, $\|\Delta^k F(z_0, \dots, z_k)\|$ denotes the norm of $\Delta^k F(z_0, \dots, z_k)$ as a multilinear form between Banach spaces, that is,

$$\|\Delta^k F(z_0, \dots, z_k)\| = \sup_{\|w_j\| \leq 1} \left\| \Delta^k F(z_0, \dots, z_k)[w_1, \dots, w_k] \right\|.$$

Lemma 3.4.2. *Let $F : \Omega_1 \to \Omega_2$ be fully matricial. Let $Z = z_0 \oplus \cdots \oplus z_k$ where $z_j \in \Omega_1^{(n_j)}$ and let $N = n_0 + \cdots + n_k$. Suppose that $B^{(N)}(Z, R) \subseteq \Omega_1^{(N)}$ and $F(B^{(N)}(Z, R)) \subseteq B^{(N)}(0, M)$. Then*

$$\|\Delta^k F(z_0, \dots, z_k)\| \leq \frac{M}{R^k} \text{ for } k \geq 1. \tag{3.2}$$

Proof. Suppose that $\|w_1\| \leq 1, \dots, \|w_k\| \leq 1$. For $r < R$, we have

$$W := \begin{bmatrix} z_0 & rw_1 & \dots & 0 & 0 \\ 0 & z_1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & z_{k-1} & rw_k \\ 0 & 0 & \dots & 0 & z_k \end{bmatrix} \in B^{(N)}(Z, R),$$

and hence $\|F(W) - F(Z)\| < M$. Looking at the top right block of $F(W) - F(Z)$, we obtain

$$\|F(z_0, \dots, z_k)[rw_1, \dots, rw_k]\| \leq M.$$

Because this holds whenever $r < R$ and $\|w_j\| \leq 1$, we have proven (3.2). □

Lemma 3.4.3. *Let $F : \Omega_1 \to \Omega_2$ be fully matricial. Let $z_* \in \Omega_1^{(n)}$, and suppose $B(z_*, R) \subseteq \Omega_1$ and $F(B(z_*, R)) \subseteq B(0, M)$. Then for $z \in B^{(n)}(z_*, R)$, we have*

$$F(z) = \sum_{k=0}^{\infty} \Delta^k F(z_*, \dots, z_*)[z - z_*, \dots, z - z_*]. \tag{3.3}$$

Proof. It follows from Lemma 3.4.2 that the power series on the right hand side of (3.3) converges when $\|z - z^*\| < R$. It remains to show that the sum of the series is $F(z)$. If we assume that $\|z - z_*\| < R/\sqrt{2}$, then by Lemma 3.4.1,

$$F(z) = \sum_{k=0}^{m-1} \Delta^k F(z_*, \dots, z_*)[z - z_*, \dots, z - z_*] + \Delta^m F(z, z_*, \dots, z_*)[z - z_*, \dots, z - z_*].$$

Now if $\|z - z_*\| \leq R/2$, then we have $B(z_*, R/2) \subseteq B(z_*, R)$ and hence $F(B(z, R/2)) \subseteq B(0, M)$. Hence, by Lemma 3.4.2,

$$\|\Delta^m F(z, z_*, \dots, z_*)[z - z_*, \dots, z - z_*]\| \leq \frac{2M\|z - z_*\|^m}{(R/2)^m}$$

which vanishes as $m \to \infty$. Therefore, (3.3) holds when $\|z - z_*\| < R/2$.

To extend (3.3) to $y \in B^{(n)}(z_*, R)$, we use complex analysis. Fix $z \in B^{(n)}(z_*, R)$. Note that for any state ϕ on $\mathcal{B}^{(n)}$ and for $\zeta \in \mathbb{C}$ with $|\zeta| < R/2\|z - z_*\|$, the function

$$g(\zeta) = \phi \circ F(z + \zeta(z - z_*)) = \sum_{k=0}^{\infty} \zeta^k \phi \circ \Delta^k F(z_*, \dots, z_*)[z - z_*, \dots, z - z_*]$$

is a scalar-valued analytic function. Now because F has also has a local power series expansion centered at $z + \zeta(z - z_*)$ whenever $z_* + \zeta(z - z_*)$ is in the domain of F, we see that g is actually analytic for $|\zeta| < R/\|z - z_*\|$. It follows that the power series expansion for g centered at 0 converges to g when $|\zeta| < R/\|z - z_*\|$. Thus, taking $\zeta = 1$, we obtain

$$\phi \circ F(z) = \sum_{k=0}^{\infty} \phi \circ \Delta^k F(z_*, \dots, z_*)[z - z_*, \dots, z - z_*],$$

and because this holds for arbitrary states ϕ, we have proved (3.3). □

3.5 Matricial properties of $\Delta^k F$

We will now describe how $\Delta^k F(z_0, \dots, z_k)$ behaves when we replace one of the z_j's with a direct sum. As a consequence, we will evaluate $\Delta^k F(z_0^{(n)}, \dots, z_0^{(n)})$ as a type of matrix amplification of $\Delta^k F(z_0, \dots, z_0)$, and hence derive a Taylor-Taylor expansion around a point z which will hold not only on a ball $B^{(n)}(z_0, r)$, but on a fully matricial ball $B(z_0, r)$. As a first step, we describe how the direct sum property of F carries over to $\Delta^k F$.

Lemma 3.5.1. *For $j = 1, \dots, k-1$, we have*

$$\begin{aligned}
&\Delta^k F(z_0, \dots, z_{j-1}, z_j \oplus z_j', z_{j+1}, \dots, z_k) \left[w_1, \dots, w_{j-1}, \begin{bmatrix} w_j, w_j' \end{bmatrix}, \begin{bmatrix} w_{j+1} \\ w_{j+1}' \end{bmatrix}, w_{j+2}, \dots, w_k \right] \\
&\quad = \Delta^k F(z_0, \dots, z_{j-1}, z_j, z_{j+1}, \dots, z_k) [w_1, \dots, w_{j-1}, w_j, w_{j+1}, w_{j+2}, \dots, w_k] \\
&\qquad + \Delta^k F(z_0, \dots, z_{j-1}, z_j', z_{j+1}, \dots, z_k) \left[w_1, \dots, w_{j-1}, w_j', w_{j+1}', w_{j+2}, \dots, w_k \right].
\end{aligned}$$

In the endpoint case $j = 0$, the same holds with the terms w_j and w_j' left out, and the endpoint case $j = k$, the same holds with the w_{j+1} and w_{j+1}' left out.

Proof. To simplify notation, first assume $k = 2$ and $j = 1$. Using the intertwining

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_0 & \zeta_1 w_1 & 0 & 0 \\ 0 & z_1 & 0 & \zeta_2 w_2 \\ 0 & 0 & z_1' & \zeta_2 w_2' \\ 0 & 0 & 0 & z_2 \end{bmatrix} = \begin{bmatrix} z_0 & \zeta_1 w_1 & 0 \\ 0 & z_1 & \zeta_2 w_2 \\ 0 & 0 & z_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

we deduce that

$$\Delta^2 F(z_0, z_1 \oplus z_1', z_2) \left[\begin{bmatrix} w_1 & 0 \end{bmatrix}, \begin{bmatrix} w_2 \\ w_2' \end{bmatrix} \right] = \Delta^2 F(z_0, z_1, z_2)[w_1, w_2].$$

A similar argument shows that

$$\Delta^2 F(z_0, z_1 \oplus z_1', z_2) \left[\begin{bmatrix} 0 & w_1' \end{bmatrix}, \begin{bmatrix} w_2 \\ w_2' \end{bmatrix} \right] = \Delta^2 F(z_0, z_1', z_2)[w_1', w_2'].$$

Then by linearity of $\Delta^2 F(z_0, z_1 \oplus z_1', z_2)$ in the first w coordinate, we get

$$\Delta^2 F(z_0, z_1 \oplus z_1', z_2)\left[\begin{bmatrix} w_1 & w_1' \end{bmatrix}, \begin{bmatrix} w_2 \\ w_2' \end{bmatrix}\right] = \Delta^2 F(z_0, z_1, z_2)[w_1, w_2] + \Delta^2 F(z_0, z_1', z_2)[w_1', w_2'].$$

The argument for the general case when $1 \leq j \leq k-1$ is the same except that we must augment our intertwining matrix by taking the direct sum with copies of the identity at the top left and bottom right (compare the general case of Lemma 3.3.5). The endpoint cases $j = 0$ and $j = k$ have a similar but simpler argument. □

Now we generalize the previous lemma to replace each z_j by an arbitrary direct sum.

Lemma 3.5.2. *Let Z_j be the $R_j \times R_j$ block diagonal matrix*

$$Z_j = z_{j,1} \oplus \cdots \oplus z_{j,R_j},$$

where the block $z_{j,r}$ is $n_{j,r} \times n_{j,r}$ and j runs from 0 to k. Let W_j be an $R_{j-1} \times R_j$ block matrix where the (r, s) block $w_{j,r,s}$ has dimensions $n_{j-1,r} \times n_{j,s}$. Then $\Delta^k F(Z_0, \dots, Z_k)[W_1, \dots, W_k]$ is an $R_0 \times R_k$ block matrix where the (r, s) block is given by

$$\sum_{r_1,\dots,r_{k-1}} \Delta^k F(z_{0,r}, z_{1,r_1}, \dots, z_{k-1,r_{k-1}}, z_{k,s})[w_{1,r,r_1}, w_{2,r_1,r_2}, \dots, w_{k-1,r_{k-2},r_{k-1}}, w_{k,r_{k-1},s}].$$

Remark 3.5.3. In the last lemma, the conditions on the dimensions are such that it would make sense to multiply the matrices $Z_0 W_1 Z_1 \dots W_k Z_k$ together. The lemma asserts that block entries of $\Delta^k F(Z_0, \dots, Z_k)[W_1, \dots, W_k]$ is computed from $\Delta^k F$ evaluated on the $z_{j,r}$'s and $w_{j,r,s}$'s in the same way as we would evaluate the matrix product $Z_0 W_1 Z_1 \dots W_k Z_k$ in terms of products of the $z_{j,r}$'s and $w_{j,r,s}$'s.

Proof of Lemma 3.5.2. We fix k and proceed by induction on the total number of direct summands of the Z_j's. If some Z_j has more than one direct summand, we can break Z_j into the direct sum of $z_{j,1} \oplus \cdots \oplus z_{j,R_j - 1}$ and z_{j,R_j} and then apply Lemma 3.5.1, and thus reduce to an earlier stage of the induction. □

Next, we will explain how to express $\Delta^k F(z_0^{(m_0)}, \dots, z_k^{(m_k)})$ as a matrix amplification of $\Delta^k F(z_0, \dots, z_k)$, and in particular, we will be able to extend the Taylor-Taylor expansion at z_* to a fully matricial ball around z_*.

Definition 3.5.4. Let $\mathcal{V}_1$, ..., $\mathcal{V}_k$ and $\mathcal{V}$ be vector spaces and let $\Lambda : \mathcal{V}_1 \times \cdots \times \mathcal{V}_k \to V$ be a multilinear form. Choose indices m_0, ..., m_k. We then define the multilinear form

$$\Lambda^{(m_0,\dots,m_k)} : M_{m_0 \times m_1}(\mathcal{V}_1) \times \cdots \times M_{m_{k-1} \times m_k}(\mathcal{V}_k) \to M_{m_0 \times m_k}(\mathcal{V})$$

by

$$[\Lambda^{(m_0,\dots,m_k)}(v_1, \dots, v_k)]_{i,j} = \sum_{i=i_0, i_1, \dots, i_{k-1}, i_k = j} \Lambda[(v_1)_{i_0,i_1}, \dots, (v_k)_{i_{k-1},i_k}]$$

We will sometimes denote the matrix amplification $\Lambda^{(m_0,\dots,m_k)}$ simply by $\Lambda^{\#}$ when we do not wish to specify the indices m_0, ..., m_k.

In particular, let $F : \Omega_1 \to \Omega_2$, where Ω_j is a fully matricial domain over $\mathcal{B}_j$. Let $z_j \in \Omega_1^{(n_j)}$ for $j = 0, \dots, k$. Then we have a multilinear form

$$\Delta^k F(z_0, \dots, z_k) : M_{n_0 \times n_k}(\mathcal{B}_1) \times \cdots \times M_{n_{k-1} \times n_j}(\mathcal{B}_1) \to M_{n_0 \times n_k}(\mathcal{B}_2).$$

If we choose indices $m_0, \dots, m_k$ and identify $M_{m_{j-1} \times m_j}(M_{n_{j-1} \times n_j}(\mathcal{B}_1))$ with $M_{n_{j-1}m_{j-1} \times n_j m_j}(\mathcal{B}_1)$, then we have by Lemma 3.5.2 that

$$\Delta^k F(z_0^{(m_0)}, \dots, z_k^{(m_k)}) = \Delta^k F(z_0, \dots, z_k)^{(m_0, \dots, m_k)}.$$

The case where $z_0 = \cdots = z_k = z_*$ is particularly relevant for non-commutative power series expansions. We now state a version of the Cauchy estimates and Taylor-Taylor expansion that take into account matrix amplification, beginning with a norm for multilinear forms which is stable under matrix amplification.

Definition 3.5.5. Recall that the norm of a multilinear form on $M_{n_0 \times n_1}(\mathcal{B}) \times \cdots \times M_{n_{k-1} \times n_k}(\mathcal{B}) \to M_{n_0 \times n_k}(\mathcal{B})$ is given by

$$\|\Lambda\| = \sup_{\|w_1\|, \dots, \|w_k\| \le 1} \|\Lambda[w_1, \dots, w_k]\|.$$

so we define the *completely bounded norm* as

$$\|\Lambda\|_\# = \sup_{m_0, \dots, m_k} \|\Lambda^{(m_0, \dots, m_k)}\|.$$

We say that Λ is *completely bounded* if $\|\Lambda\|_\# < +\infty$.

The next corollaries follow immediately from Lemma 3.5.2

Corollary 3.5.6. *Suppose that $F : \Omega_1 \to \Omega_2$ is fully matricial and $B(z_*, R) \subseteq \Omega_1$ and $F(B(z_*, R)) \subseteq B(0, M)$. Then $\|\Delta^k F(z_*, \dots, z_*)\|_\# \le M/R^k$.*

Proof. Let $m_0, \dots, m_k \in \mathbb{N}$. Then $B(z_*^{(m_0)} \oplus \cdots \oplus z_*^{(m_k)}, R) \subseteq B(z_*, R) \subseteq \Omega_1$, so it follows from Lemma 3.4.2 that

$$\|\Delta^k F(z_*^{(m_0)}, \dots, z_*^{(m_k)})\| \le \frac{M}{R^k}.$$

This holds for all $m_0, \dots, m_k$, so we are done. □

Corollary 3.5.7. *Let $F : \Omega_1 \to \Omega_2$, let $z_* \in \Omega_1^{(n)}$, and suppose that $F(B(z_*, R)) \subseteq B(0, M)$. Then for $z \in B^{(nm)}(z_*^{(m)}, R)$, we have*

$$F(z) = \sum_{k=0}^\infty [\Delta^k F(z_*, \dots, z_*)]^{(m, \dots, m)}(z - z_*^{(m)}, \dots, z - z_*^{(m)}).$$

This amplified power series expansion will allow us to compute and to estimate the derivatives of F at points in $B(z_*, R)$.

Proposition 3.5.8. *Suppose that $F : \Omega_1 \to \Omega_2$ is fully matricial, $z_* \in \Omega^{(n)}$ and $B(z_*, R) \subseteq \Omega_1$ and $F(B(z_*, R)) \subseteq B(0, M)$. Let $z_0, \dots, z_k$ be points with $z_j \in B^{(nm_j)}(0, R)$. Then we have*

$$\begin{aligned}&\Delta^k F(z_*^{(m_0)} + z_0, \dots, z_*^{(m_k)} + z_k)[w_1, \dots, w_k] \\ &= \sum_{\ell_0, \dots, \ell_k \geq 0} \Delta^{\ell_0 + \dots + \ell_k + k} F(z_*^{(m_0)}, \dots, z_*^{(m_k)})[\underbrace{z_0, \dots, z_0}_{\ell_0}, w_1, \underbrace{z_1, \dots, z_1}_{\ell_1}, \dots, w_k, \underbrace{z_k, \dots, z_k}_{\ell_k}]. \end{aligned} \tag{3.4}$$

In particular,

$$\|\Delta^k F(z_*^{(m_0)} + z_0, \dots, z_*^{(m_k)} + z_k)\|_\# \leq \frac{M}{(R - \|z_0\|) \dots (R - \|z_k\|)} \tag{3.5}$$

and

$$\|\Delta^k F(z_*^{(m_0)} + z_0, \dots, z_*^{(m_k)} + z_k) - \Delta^k F(z_*^{(m_0)}, \dots, z_*^{(m_k)})\|_\# \leq \frac{M \sum_{j=1}^k \|z_j\|}{(R - \|z_0\|) \dots (R - \|z_k\|)}. \tag{3.6}$$

Proof. Since $z_* \in \Omega^{(n)}$, the equations and estimates only concern the functions $F^{(mn)}$ for $m \in \mathbb{N}$. Now $F^{(mn)}(z - z_*^{(m)})$ is defined on $B^{(mn)}(0^{(mn)}, R)$ for each m and defines a fully matricial function on the domain $B(0^{(n)}, R)$. Thus, we may assume without loss of generality that $z_* = 0^{(n)}$. Furthermore, we denote

$$\Lambda_k = \Delta^k F(z_*, \dots, z^*),$$

so that

$$F(z) = \sum_{k=0}^\infty \Lambda_k^\#(z, \dots, z) \text{ for } z \in B(0^{(n)}, R).$$

Before performing the computation, we first show that the series converges absolutely and estimate it. Observe that

$$\begin{aligned}\sum_{\ell_0, \dots, \ell_k \geq 0} &\left\| \Lambda_{\ell_0 + \dots + \ell_k + k}[\underbrace{z_0, \dots, z_0}_{\ell_0}, w_1, \underbrace{z_1, \dots, z_1}_{\ell_1}, \dots, w_k, \underbrace{z_k, \dots, z_k}_{\ell_k}] \right\| \\ &\leq \sum_{\ell_0, \dots, \ell_k \geq 0} \frac{M}{R^{\ell_0 + \dots + \ell_k + k}} \|z_0\|^{\ell_0} \dots \|z_k\|^{\ell_k} \|w_1\| \dots \|w_k\| \\ &= \frac{M \|w_1\| \dots \|w_k\|}{(R - \|z_0\|) \dots (R - \|z_k\|)},\end{aligned}$$

where the last equality follows from summing the geometric series.

Now let us show that sum converges to $\Delta^k F(z_0, \dots, z_k)[w_1, \dots, w_k]$. Consider the block upper triangular matrix Z with the entries $z_0, \dots, z_k$ on the diagonal, the entries $w_1, \dots,$

w_k just above the diagonal, and zeroes elsewhere. By rescaling, assume that $w_1, \dots, w_k$ are small enough that $\|Z\| < R$. Recall that $\Delta^k F(z_0, \dots, z_k)[w_1, \dots, w_k]$ is the upper right corner of $F(Z)$. The upper right block of $\Lambda_\ell^\#(Z, \dots, Z)$ is given by

$$\sum_{1=i_0, i_1, \dots, i_\ell = k+1} \Lambda_\ell^\#(Z_{i_0,i_1}, \dots, Z_{i_{\ell-1},i_\ell}),$$

where $Z_{i,j}$ denotes the (i,j) block of Z. Because Z is block upper triangular with the only nonzero entries being on the diagonal and directly above it, the only nonzero terms in the matrix expansion are of the form

$$\Lambda_\ell^\#(\underbrace{z_0, \dots, z_0}_{m_0}, w_1, \underbrace{z_1, \dots, z_1}_{m_1}, \dots, w_k, \underbrace{z_k, \dots, z_k}_{m_k}).$$

Thus, we have

$$\begin{aligned} &\Delta^k F(z_0, \dots, z_k)[w_1, \dots, w_k] \\ &\qquad = \sum_{\ell=0}^\infty \left(\sum_{\substack{m_0, \dots, m_k \geq 0 \\ m_0 + \dots + m_k + k = \ell}} \Lambda_\ell^\#(\underbrace{z_0, \dots, z_0}_{m_0}, w_1, \underbrace{z_1, \dots, z_1}_{m_1}, \dots, w_k, \underbrace{z_k, \dots, z_k}_{m_k}) \right), \end{aligned}$$

which is exactly (3.4) in the case $z_* = 0^{(n)}$.

We already showed that when $z_* = 0^{(n)}$,

$$\|\Delta^k F(z_0, \dots, z_k)\| \leq \frac{M}{(R - \|z_0\|) \dots (R - \|z_k\|)}.$$

Because the same reasoning applies to $\Delta^k(z_0^{(m_0)}, \dots, z_k^{(m_k)})$ and yields the same estimate, we have bounded $\|\Delta^k F(z_0, \dots, z_k)\|_\#$ and proven (3.5).

To prove (3.6), observe that

$$\begin{aligned} &\|\Delta^k F(z_0, \dots, z_k) - \Delta^k F(0^{(nm_0)}, \dots, 0^{(nm_k)})\| \\ &\leq \sum_{\substack{\ell_0, \dots, \ell_k \geq 0 \\ \ell_0 + \dots + \ell_k \geq 1}} \left\| \Lambda_{\ell_0 + \dots + \ell_k + k}[\underbrace{z_0, \dots, z_0}_{\ell_0}, w_1, \underbrace{z_1, \dots, z_1}_{\ell_1}, \dots, w_k, \underbrace{z_k, \dots, z_k}_{\ell_k}] \right\| \\ &\leq \sum_{\ell_0, \dots, \ell_k \geq 0} \frac{M}{R^{\ell_0 + \dots + \ell_k + k}} \|z_0\|^{\ell_0} \dots \|z_k\|^{\ell_k} \|w_1\| \dots \|w_k\| \\ &= M \|w_1\| \dots \|w_k\| \left(\frac{1}{(R - \|z_0\|) \dots (R - \|z_k\|)} - \frac{1}{R^k} \right), \\ &\leq M \|w_1\| \dots \|w_k\| \frac{\sum_{j=0}^k \|z_j\|}{(R - \|z_0\|) \dots (R - \|z_k\|)}. \end{aligned}$$

The same argument also applies to the matrix amplifications of $\Delta^k F(z_0, \dots, z_k)$ and hence we have bounded $\Delta^k F(z_0, \dots, z_k) - \Delta^k F(0^{(nm_0)}, \dots, 0^{nm_k})$ in $\|\cdot\|_\#$. □

In particular, (3.6) implies the following corollary.

Corollary 3.5.9. *Let $F : \Omega_1 \to \Omega_2$ be fully matricial. Then $F(z)$ and $\Delta^k F(z, \dots, z)$ are uniformly locally Lipschitz functions of z with respect to $\|\cdot\|_\#$. That is, for every $z_* \in \Omega_1^{(n)}$, there exists $r > 0$ such that $F(z)$ and $\Delta^k F(z, \dots, z)$ are Lipschitz on $B^{(mn)}((z_*)^{(m)}, r)$, with Lipschitz constants independent of m.*

Furthermore, the following lemma shows that the multlinear forms in this amplified power series expansion are unique. That is, any other sequence of multilinear forms Λ_k which satisfies the equation in Corollary 3.5.7 must be equal to $\Delta^k F(z_0, \dots, z_0)$. This lemma justifies many ways of computing the derivatives of a fully matricial function F. As long as we obtain a power series that converges to F and the manipulation works for every size of matrices, then we must have the correct answer. This includes for instance computing the power series for products and compositions of fully matricial functions.

Lemma 3.5.10. *Let $F : \Omega_1 \to \Omega_2$ and $z_* \in \Omega_1^{(n)}$. Let $\Lambda_k : M_n(\mathcal{B}_1)^k \to M_n(\mathcal{B}_2)$ be a sequence of multilinear forms. If for some $r > 0$, have*

$$F(z) = \sum_{k=0}^{\infty} \Lambda_k^{(m,\dots,m)}(z - z_*^{(m)}, \dots, z - z_*^{(m)})$$

for $z \in B^{(m)}(z_*, r)$ for all m, then $\Lambda_k = \Delta^k F(z_*, \dots, z_*)$. In fact, we need only assume that the expansion holds when $z - z_*^{(m)}$ is strictly upper triangular and in $B^{(nm)}(z_*^{(m)}, r_m)$ for some $r_m > 0$.

Proof. Fix k. Fix $w_1, \dots, w_k \in M_n(\mathcal{B})$ and let $\zeta_1, \dots, \zeta_k$ be small scalars. Let

$$W = \begin{bmatrix} 0 & \zeta_1 w_1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \zeta_2 w_2 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots 0 & \zeta_k w_k & \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}]$$

Then we observe that

$$F(z_*^{(k+1)} + W) = \begin{bmatrix} 0 & \Lambda_1(\zeta_1 w_1) & \Lambda_2(\zeta_1 w_1, \zeta_2 w_2) & \dots & \Lambda_k(\zeta_1 w_1, \dots, \zeta_k w_k) & \\ 0 & 0 & \Lambda_1(\zeta_2 w_2) & \dots & \Lambda_{k-1}(\zeta_2 w_2, \dots, \zeta_k w_k) & \\ 0 & 0 & 0 & \dots & \Lambda_{k-2}(\zeta_3 w_3, \dots, \zeta_k w_k) & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \dots \Lambda_1(\zeta_k w_k) & & \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}],$$

and by examining the upper right block, it follows that

$$\Delta^k F(z_*, \ldots, z_*)(\zeta_1 w_1, \ldots, \zeta_k w_k) = \Lambda_k(\zeta_1 w_1, \ldots, \zeta_k w_k).$$

Thus, $\Delta^k F(z_*, \ldots, z_*) = \Lambda_k$ as desired. □

3.6 Examples

3.6.1 Series of multilinear forms

Our first example is closely related to the material from the last section on the matrix amplifications of multilinear forms. We will characterize the fully matricial functions on the ball $B(0, R)$ (where 0 is the 1×1 zero matrix) as convergent series of multilinear forms. We remark that the corresponding notion of formal power series of multilinear forms was studied by Dykema [Dyk07].

Proposition 3.6.1. *Suppose that $\Lambda_k : \mathcal{B}_1^k \to \mathcal{B}_2$ is a completely bounded multilinear form and that $\limsup_{k\to\infty} \|\Lambda_k\|_\#^{1/k} \leq 1/R$. Then*

$$F^{(n)}(z) = \sum_{k=0}^{\infty} \Lambda_k^{(n)}[z, \ldots, z]$$

is a fully matricial function on $B(0, R)$ which satisfies $\Delta^k F(0, \ldots, 0) = \Lambda_k$. Moreover, F is uniformly bounded on $B(0, r)$ for each $r < R$. Conversely, if F is a fully matricial function on $B(0, R)$ which is uniformly bounded on $B(0, r)$ for each $r < R$, then F can be written in this form, where $\Lambda_k = \Delta^k F(0, \ldots, 0)$.

Proof. Let Λ_k be given with $\limsup_{k\to\infty} \|\Lambda_k\|_\#^{1/k} \leq 1/R$. Choose $r < R$ and let $r < r' < R$. Then for k greater than or equal to some N, we have $\|\Lambda_k\| \leq 1/r'$. This implies that for $\|z\| \leq r$, we have

$$\sum_{k=0}^{\infty} \|\Lambda_k^{(n,\ldots,n)}(z, \ldots, z)\| \leq \sum_{k=0}^{N-1} \|\Lambda_k\|_\# r^k + \sum_{k=N}^{\infty} \left(\frac{r}{r'}\right)^k < +\infty.$$

This shows that the series converges uniformly on $B(0, r)$ and defines a function F which is bounded on $B(0, r)$ for each $r < R$. To show that F is fully matricial, suppose $zT = Tw$ where $z \in B^{(n)}(0, R)$ and $w \in B^{(m)}(0, R)$ and $T \in M_{n\times m}(\mathbb{C})$. A direct computation from the definition of the matrix amplification of multilinear forms shows that

$$\Lambda_k^\#(z, \ldots, z)T = \Lambda_k^\#(z, \ldots, z, zT) = \Lambda_k^\#(z, \ldots, z, Tw) = \Lambda_k^\#(z, \ldots, z, zT, w) = \ldots$$
$$\cdots = \Lambda_k^\#(Tw, w, \ldots, w) = T\Lambda_k^\#(w, \ldots, w).$$

Therefore, $F(z)T = TF(w)$ as desired.

Now consider the converse direction. Suppose that $\|F(z)\| \leq M_r$ for $\|z\| \leq r < R$. By Corollary 3.5.6, we have $\|\Lambda_k\|_\# \leq M_r/r^k$, so that $\limsup_{k\to\infty} \|\Delta^k F(0,\dots,0)\|_\#^{1/k} \leq 1/r$. Thus holds for all $r < R$, and so $\limsup_{k\to\infty} \|\Delta^k F(0,\dots,0)\|_\#^{1/k} \leq 1/R$. Moreover, by Corollary 3.5.7, $F(z)$ is given as the sum of $\Delta^k F(0,\dots,0)^\#[z,\dots,z]$. □

3.6.2 Non-commutative polynomials

In particular, if $F(X) = b_0 X b_1 \dots X b_k$ is a monomial in $\mathcal{B}\langle X\rangle$, then there is a corresponding multilinear form

$$\Lambda : (z_1, \dots, z_k) \mapsto b_0 z_1 b_1 \dots z_k b_k.$$

Note that

$$\Lambda^{(n)}(z_1, \dots, z_k) = b_0^{(n)} z_1 b_1^{(n)} \dots z_k b_k^{(n)}.$$

Thus, $\|\Lambda^{(n)}\| \leq \|b_0\| \dots \|b_k\|$, so that Λ is completely bounded. Thus, we can define a fully matricial function by

$$F^{(n)}(z) = \Lambda^{(n)}(z, \dots, z) = b_0^{(n)} z b_1^{(n)} \dots z b_k^{(n)}.$$

By linearity, for every non-commutative polynomial $F(X) \in \mathcal{B}\langle X\rangle$, the function $F(z)$ is fully matricial on $M_\bullet(\mathcal{B})$. Moreover, the derivatives $\Delta^k F$ are computed as in Lemma 3.5.8. For example, if $F(z) = b_0 z b_1 \dots z b_\ell$ and if $z_0, \dots, z_\ell \in \mathcal{B}$, we have

$$\Delta^k F(z_0, \dots, z_k)[w_1, \dots, w_k] = \sum_{1 \leq \ell_1 < \ell_2 < \dots < \ell_k \leq \ell} (b_0 z_0 b_1 \dots z_0 b_{\ell_1 - 1}) w_1 (b_{\ell_1} z_1 b_{\ell_1+1} \dots z_1 b_{\ell_2 - 1}) \dots w_k (b_{\ell_k} z_k b_{\ell_k+1} \dots z_k b_\ell).$$

3.7 Algebraic operations

Proposition 3.7.1. *Suppose that $F, G : \Omega \to M_\bullet(\mathcal{A})$ are fully matricial. Then so are $F + G$ and FG.*

Proof. Note that if $zT = Tw$ for some scalar matrix T, then we have

$$(F + G)(z)T = F(z)T + G(z)T = TF(w) + TG(w) = T(F + G)(w),$$

and

$$(FG)(z)T = F(z)G(z)T = F(z)TG(w) = TF(w)G(w) = T(FG)(w),$$

so that $F + G$ and FG respect intertwinings. To show $F + G$ and FG are uniformly locally bounded, pick $z_0 \in \Omega^{(n)}$. Then F is bounded by M_1 on some ball $B(z_0, R_1)$ and G is bounded by M_2 on some ball $B(z_0, R_2)$. Letting $R = \min(R_1, R_2)$, we have

$$\|z - z_0^{(m)}\| \leq R \implies \|F(z) + G(z)\| \leq M_1 + M_2 \text{ and } \|F(z)G(z)\| \leq M_1 M_2.$$

□

Lemma 3.7.2. *The sequence of sets $\Omega^{(n)} = \{z \in M_n(\mathcal{B}) : z \text{ is invertible}\}$ is a matricial domain and the function*

$$\operatorname{inv} : \Omega \to \Omega : z \mapsto z^{-1}$$

is fully matricial.

Proof. Note that Ω respects direct sums and is nonempty. To show that Ω is uniformly open, suppose that $z \in \Omega^{(n)}$. Then we claim that $B(z, 1/\|z^{-1}\|)$ is contained in Ω. To see this note that if $w \in B(z, 1/\|z^{-1}\|)$, then the series

$$w^{-1} = [z - (z - w)]^{-1} = z^{-1}[1 - (z - w)z^{-1}]^{-1} = \sum_{k=0}^{\infty} z^{-1}[(z - w)z^{-1}]^k$$

converges and we have

$$\|w^{-1}\| \leq \frac{\|z^{-1}\|}{1 - \|z^{-1}\|\|z - w\|}.$$

This same estimate shows that inv is uniformly locally bounded.

To show that inv respects intertwinings, suppose that $zT = Tw$. Multiplying by z^{-1} on the left and w^{-1} on the right yields $Tw^{-1} = z^{-1}T$ or $\operatorname{inv}(z)T = T\operatorname{inv}(w)$. □

Proposition 3.7.3. *Suppose that $F : \Omega_1 \to \Omega_2$ and $G : \Omega_2 \to \Omega_3$ are fully matricial. Then so is $G \circ F$.*

Proof. To show that $G \circ F$ respects intertwinings, suppose that $zT = Tw$. Then $F(z)T = TF(w)$ and hence $G(F(z))T = TG(F(w))$. To show uniform local boundedness, pick a point z_0. By uniform local boundedness of G, we can choose R and $M > 0$ such that $G(B(F(z_0), R)) \subseteq B(0, M)$. But by Corollary 3.5.9, there exists an R' such that $F(B(z_0, R')) \subseteq B(F(z_0), R)$. Thus, $G \circ F$ is uniformly bounded on $B(z_0, R')$. □

3.8 Inverse function theorem

We now present an inverse function theorem for fully matricial functions. For background and related results, see [Voi04, §11.5], [AK13], [AK15], [AM16]. In particular, the following result is a version of [AK15, Theorem 1.4].

Theorem 3.8.1. *Let $z_* \in M_n(\mathcal{B}_1)$ and $w_* \in M_n(\mathcal{B}_2)$. Suppose that $F : B(z_*, R) \to B(w_*, M)$ is fully matricial with $F(z_*) = w_*$. Suppose that $\Lambda_1 = \Delta F(z_*, z_*)$ is invertible with $\|\Lambda_1^{-1}\|_\# \leq K$. Then there exist r_1 and r_2 such that the following holds.*

(1) For each $w \in B(w_*, r_2)$, there exists a unique $z \in B(z_*, r_1)$ with $F(z) = w$.

(2) The inverse function $F^{-1} : B(w_*, r_2) \to B(z_*, r_1)$ is fully matricial.

More precisely, we can take

$$r_1 = R\rho_1\left(\frac{MK}{R}\right), \qquad r_2 = \frac{R}{K}\rho_2\left(\frac{MK}{R}\right),$$

where

$$\rho_1(t) = 1 - \frac{t^{1/2}}{(1+t)^{1/2}} \qquad \rho_2(t) = 1 + 2t - 2t^{1/2}(1+t)^{1/2}.$$

Proof. First, consider the special case where $\mathcal{B}_1 = \mathcal{B}_2$, $z_* = 0^{(n)}$, $w_* = 0^{(n)}$, $R = 1$, and $\Lambda_1 = \mathrm{id}$. Let $\Lambda_k = \Delta^k F(0^{(n)}, \dots, 0^{(n)})$. For $w \in M_{mn}(\mathcal{B}_1)$, note that $F(z) = w$ if and only if z is a fixed point of the function

$$H_w(z) = w + z - F(z) = w - \sum_{k=2}^{\infty} \Lambda_k(z, \dots, z).$$

We want to show that if r and w are sufficiently small, then G_w defines a contraction $\overline{B}^{(nm)}(0, r) \to \overline{B}^{(nm)}(0, r)$ and hence has a unique fixed point in $\overline{B}^{(nm)}(0, r)$.

To determine when H_w is a contraction, we estimate $H_w(z) - H_w(z')$. Let

$$\Lambda_k = \Delta^k F(0^{(n)}, \dots, 0^{(n)}).$$

Then for $\|z'\|$ and $\|z\| \leq r$, we have

$$\begin{aligned}
\|H_w(z) - H_w(z')\| &\leq \sum_{k=2}^{\infty} \|\Lambda_k(z, \dots, z) - \Lambda_k(z', \dots, z')\| \\
&\leq \sum_{k=2}^{\infty} \sum_{j=0}^{k-1} \|\Lambda_k(\underbrace{z, \dots, z}_{j}, z - z', \underbrace{z', \dots, z'}_{k-1-j})\| \\
&\leq M \sum_{k=2}^{\infty} k r^{k-1} \|z - z'\| \\
&= M\left(\frac{1}{(1-r)^2} - 1\right) \|z - z'\|.
\end{aligned}$$

Therefore, H_w is a contraction provided that

$$M\left(\frac{1}{(1-r)^2} - 1\right) < 1$$

or equivalently $r < \rho_1(M)$.

To determine when $H_w(z)$ maps $B^{(mn)}(0,r)$ into itself, note that for $\|z\| \leq r$

$$\begin{aligned}\|H_w(z)\| &\leq \|w\| + \sum_{k=2}^{\infty} \|\Lambda_k(z,\dots,z)\| \\ &\leq \|w\| + \frac{Mr^2}{1-r}.\end{aligned}$$

Thus, we have $\|H_w(z)\| \leq r$ provided that

$$\|w\| \leq \psi(r) := r - \frac{Mr^2}{1-r}.$$

Altogether, we have shown that $r < \rho_1(M)$ and $\|w\| \leq \psi(r)$, then H_w is a strict contraction $B^{(mn)}(0,r) \to B^{(mn)}(0,r)$. Therefore, by the Banach fixed point theorem, H_w has a unique fixed point in $B^{(mn)}(0,r)$. We denote this fixed point by $G^{(mn)}(w)$. Thus, $G^{(mn)}$ is a function $\overline{B}^{(mn)}(0,\psi(r)) \to \overline{B}^{(mn)}(0,r)$ for $r < \rho_1(M)$. By uniqueness of the fixed point, the value of $G(w)$ is independent of the choice of r, so G defines a function on the union of the balls $\overline{B}^{(mn)}(0,\psi(r))$ for $r < \rho_1(M)$. But $\psi(\rho_1(M)) = \rho_2(M)$, and thus G defines a function $B(0^{(n)}, \rho_2(M)) \to B(0^{(n)}, \rho_1(M))$.

We claim that G is fully matricial. Consider a similarity $w' = SwS^{-1}$ where $z, z' \in B(0^{(n)}, \rho_2(M))$ and $S \in GL_n(\mathbb{C})$. For r sufficiently close to $\rho_1(M)$, we have $\|w\|, \|w'\| \leq \psi(r)$. Note that $F(SG(w)S^{-1}) = SF(G(w))S^{-1} = SwS^{-1}$ and thus by uniqueness of the fixed point for $H_{SwS^{-1}}$ on $B^{(mn)}(0,r)$, we have $G(SwS^{-1}) = SG(w)S^{-1}$. The argument for direct sums is similar.

This completes the proof in the special case where $\mathcal{B}_1 = \mathcal{B}_2$, $z_* = 0^{(n)}$, $w_* = 0^{(n)}$, $R = 1$, and $\Lambda_1 = \mathrm{id}$. Now consider a function F which satisfies the hypotheses of the theorem in the general case. Let

$$\widehat{F}^{(nm)}(z) = \frac{1}{R}(\Lambda_1^{-1})^{\#}[F^{(nm)}(Rz + z_*) - w_*].$$

Then $\widehat{F}$ is a fully matricial function $B(0^{(n)}, 1) \to B(0^{(n)}, MK/R)$. The previous argument yields an inverse function $\widehat{G} : B(0^{(n)}, \rho_2(MK/R)) \to B(0^{(n)}, \rho_1(MK/R))$. The inverse function to F is given by

$$G(w) = R\widehat{G}\left(\frac{1}{R}(\Lambda_1^{-1})^{\#}[w - w_*]\right) + z_*,$$

and this function is defined $B(w_*, (R/K)\rho_2(MK/R)) \to B(z_*, R\rho_1(MK/R))$. □

Remark 3.8.2. Curiously, $\psi(r)$ is maximized when $r = \rho_1(M)$. Thus, the choice of r which will guarantee that H_w maps $B^{(mn)}(0,r)$ into $B^{(mn)}(0,r)$ for the largest range of w is $r = \rho_1(M)$. This is the same as the largest choice of r which will guarantee that H_w is a contraction.

Remark 3.8.3. In fact, the proof never used directly the fact that $F(B(z_*, R)) \subseteq B(w_*, M)$. It only used the Cauchy esimate

$$\|\Delta^k F(z_*, \dots, z_*)\| \leq \frac{M}{R^k}.$$

Thus, the conclusion of the theorem holds when we replace the boundedness assumption by this Cauchy estimate.

Furthermore, the inverse function depends continuously on the original function F in the following sense.

Proposition 3.8.4. *Let $F, G : B(z_*, R) \to M_\bullet(\mathcal{B}_2)$ be fully matricial. Suppose that*

$$F(B(z_*, R)) \subseteq B(F(z_*), M), \qquad G(B(z_*, R)) \subseteq B(G(z_*), M).$$

Suppose that $\Delta F(z_*, z_*)$ and $\Delta G(z_*, z_*)$ are invertible with

$$\|\Delta F(z_*, z_*)^{-1}\|_\# \leq K, \qquad \|\Delta G(z_*, z_*)^{-1}\|_\# \leq K.$$

Let r_1 and r_2 be as in Theorem 3.8.1 and let

$$F^{-1} : B(F(z_*), r_2) \to B(z_*, r_1), \qquad G^{-1} : B(G(z_*), r_2) \to B(z_*, r_1)$$

be the inverse functions given by that theorem. If we have

$$\sup_{z \in B(z_*, r_1)} \|F(z) - G(z)\| \leq \frac{r_2}{3},$$

then

$$\sup_{w \in B(G(w), r_2/3)} \|F^{-1}(w) - G^{-1}(w)\| \leq \frac{9r_1}{2r_2^2} \sup_{z \in B(z_*, r_1)} \|F(z) - G(z)\|.$$

Proof. Let $w \in B(G(z_*), r_2/3)$. Note that $B(G(z_*) r_2/3) \subseteq B(F(z_*), 2r_2/3)$ and hence $F^{-1}(w)$ is defined. Now let $w' = F \circ G^{-1}(w)$ and note that

$$F^{-1}(w) - G^{-1}(w) = F^{-1}(w) - F^{-1} \circ F \circ G^{-1}(w) = F^{-1}(w) - F^{-1}(w').$$

Moreover, we have

$$\|w - w'\| = \|G \circ G^{-1}(w) - F \circ G^{-1}(w)\| \leq \sup_{z \in B(z_*, r_1)} \|F(z) - G(z)\| \leq \frac{r_2}{3}.$$

Now because F^{-1} maps $B(z_*, r_2)$ into $B(F(z_*), r_1)$, we have by Lemma 3.4.1 and Proposition 3.5.8

$$\begin{aligned} \|F^{-1}(w) - F^{-1}(w')\| &= \left\|\Delta[F^{-1}](w, w')[w - w']\right\| \\ &\leq \frac{r_1}{(r_2 - \|w - F(z_*)^{(m)}\|)(r_2 - \|w' - F(z_*)^{(m)}\|)} \|w - w'\|. \end{aligned}$$

But $w \in B(F(z_*), r_2/3)$ and $w' \in B(F(z_*), 2r_2/3)$ and therefore

$$\begin{aligned} \|F^{-1}(w) - F^{-1}(w')\| &\leq \frac{r_1}{(r_2 - r_2/3)(r_2 - 2r_2/3)} \|w - w'\| \\ &\leq \frac{9r_1}{2r_2^2} \sup_{z \in B(z_*, r_1)} \|F(z) - G(z)\|. \end{aligned}$$

□

3.9 Uniformly locally bounded families

In complex analysis, the identity theorem states that if two analytic functions on a connected domain Ω agree in a neighborhood of a point z_0, then they must agree on Ω. Another related result is that if a sequence of functions f_n is uniformly locally bounded, and if $f_n \to f$ in a neighborhood of a point, then $f_n \to f$ locally uniformly on Ω.

More generally, for a family of functions which is uniformly locally bounded, the topology of local uniform convergence on Ω is metrizable with the metric given by $\sup_{z \in B(z_0,r)} |f(z) - g(z)|$. In fact, for various choices of z_0 and r, we obtain equivalent metrics.

We will now describe the fully matricial analogues of these results.

Definition 3.9.1. A fully matricial domain Ω is *connected* if z and w are in $\Omega^{(n)}$, then there exists $m > 0$ such that $z^{(m)}$ and $w^{(m)}$ are in the same connected component of $\Omega^{(nm)}$.

Definition 3.9.2. We say that a family $\mathcal{F}$ of fully matricial functions $\Omega \to M_\bullet(\mathcal{B})$ is *uniformly locally bounded* if for every $z_* \in \Omega$, there exists $R > 0$ and $M > 0$ such that

$$\sup_{z \in B(z_*,R)} \|F(z)\| \leq M \text{ for all } F \in \mathcal{F}.$$

Definition 3.9.3. Let $\mathcal{F}$ be a uniformly locally bounded family of fully matricial functions $\Omega \to M_\bullet(\mathcal{B})$. For $z_* \in \Omega$, we denote

$$\operatorname{rad}(z_*, \mathcal{F}) = \sup \left\{ R > 0 : \sup_{F \in \mathcal{F}} \sup_{z \in B(0,R)} \|F(z)\| < +\infty \right\},$$

and we call $\operatorname{rad}(z_*, \mathcal{F})$ the *radius of uniform local boundedness of $\mathcal{F}$ at z_**.

Definition 3.9.4. Let $\mathcal{F}$ be uniformly locally bounded family of fully matricial functions $\Omega \to M_\bullet(\mathcal{B})$. For $z_* \in \Omega$ and $r < \operatorname{rad}(z_*, \mathcal{F})$, we define

$$d_{z_*,r}(F,G) = \sup_{z \in B(z_*,r)} \|F(z) - G(z)\|$$

and

$$d'_{z_*,r}(F,G) = \sum_{k=0}^\infty r^k \|\Delta^k F(z_*, \dots, z_*) - \Delta^k G(z_*, \dots, z_*)\|_\#.$$

Definition 3.9.5. Let d_1 and d_2 be metrics on a set $\mathcal{X}$. We say that $d_1 \lesssim d_2$ if the map $\operatorname{id}_X : (\mathcal{X}, d_2) \to (\mathcal{X}, d_1)$ is uniformly continuous. In other words, for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$d_2(x,y) < \delta \implies d_2(x,y) < \epsilon.$$

We say that d_1 and d_2 are *uniformly equivalent* or $d_1 \sim d_2$ if we have $d_1 \lesssim d_2$ and $d_2 \lesssim d_1$. Note that $\lesssim$ is transitive and $\sim$ is an equivalence relation.

Theorem 3.9.6. *Let Ω be a connected fully matricial domain. Let $\mathcal{F}$ be a uniformly locally bounded family of fully matricial functions $\Omega_1 \to M_\bullet(\mathcal{B}_2)$.*

(1) For $z_* \in \Omega$ and $r < \operatorname{rad}(z_*, \mathcal{F})$, the functions $d_{z_*,r}$ and $d'_{z_*,r}$ are metrics on $\mathcal{F}$.

(2) All the metrics in the collection $\{d_{z_*,r}, d'_{z_*,r} : z_* \in \Omega, r < \operatorname{rad}(z_*, \mathcal{F})\}$ are uniformly equivalent to each other.

Proof.

Step 1: From the definition $d_{z_*,r}$, we see that $d_{z_*,r}$ is finite, satisfies the triangle inequality, and satisfies $d_{z_*,r}(F,G) = d_{z_*,r}(G,F)$. To show that $d'_{z_*,r}$ is finite, choose R with $r < R < \operatorname{rad}(z_*, \mathcal{F})$. By applying the Cauchy estimates (3.5), we see that for some constant M, we have

$$\|\Delta^k(F-G)(z_*, \dots, z_*)\|_\# \leq \frac{2M}{R^k},$$

so for $r < R$, we have

$$\sum_{k=0}^\infty r^k \|\Delta^k(F-G)(z_*, \dots, z_*)\|_\# < +\infty.$$

It is also immediate that $d_{z_*,r}$ satisfies the symmetry and triangle inequality properties. In other words, $d_{z_*,r}$ and $d'_{z_*,r}$ are pseudometrics.

In the rest of the proof, we will use the notation $d \lesssim d'$ for pseudometrics d and d' as well as for metrics. The meaning is that for every $\epsilon > 0$, there exists $\delta > 0$ such that $d'(F,G) < \delta$ implies $d(F,G) < \epsilon$.

Step 2: We claim that $d_{z_*,r} \lesssim d'_{z_*,r}$. Note that for $z \in B^{(nm)}(z_*, r)$, we have

$$\begin{aligned}
\|(F-G)(z)\| &\leq \sum_{k=1}^\infty \left\| \Delta^k(F-G)(z_*, \dots, z_*)^\#[z - z_*^{(m)}, \dots, z - z_*^{(m)}] \right\| \\
&\leq \sum_{k=1}^\infty \left\| \Delta^k(F-G)(z_*, \dots, z_*) \right\|_\# r^k,
\end{aligned}$$

and hence $d_{z_*,r} \leq d'_{z_*,r}$.

Step 3: We claim that for $r_1, r_2 < \operatorname{rad}(z_*, \mathcal{F})$, we have $d'_{z_*,r_1} \lesssim d_{z_*,r_2}$. First, choose R with $r_1 < R < \operatorname{rad}(z_*, \mathcal{F})$, choose M such that

$$\sup_{z \in B(z_*,R)} \|F(z)\| \leq M$$

and note that by the Cauchy estimate (3.5), we have

$$\|\Delta^k(F-G)(z_*, \dots, z_*)\|_\# \leq \frac{2M}{R^k}.$$

By the same estimate we have

$$\|\Delta^k(F-G)(z_*,\dots,z_*)\|_\# \le \frac{d_{z_*,r_2}(F,G)}{r_2^k}$$

Thus, we have

$$\begin{aligned} d'_{z_*,r_1}(F,G) &= \sum_{k=0}^{\infty} r_1^k \|\Delta^k(F-G)(z_*,\dots,z_*)\|_\# \\ &\le \sum_{k=0}^{N-1} d_{z_*,r_2}(F,G)\left(\frac{r_1}{r_2}\right)^k + \sum_{k=N}^{\infty} 2M\left(\frac{r_1}{R}\right)^k \\ &= d_{z_*,r_2}(F,G)\frac{(r_1/r_2)^N-1}{r_1/r_2-1} + \frac{2M(r_1/R)^N}{1-r_1/R}. \end{aligned}$$

If $\epsilon > 0$, then by choosing N large enough, we can make the second term smaller than $\epsilon/2$. After we fix such an N, then if $d_{z_*,r_2}(F,G)$ is sufficiently small, then the first term will also be less than $\epsilon/2$. This shows that $d'_{z_*,r_1} \lesssim d_{z_*,r_2}$.

Step 4: Using Steps 2 and 3, we see that for $r_1, r_2 < \operatorname{rad}(z_*, \mathcal{F})$, we have

$$d_{z_*,r_1} \lesssim d'_{z_*,r_1} \lesssim d_{z_*,r_2},$$

so the pseudometrics $d_{z_*,r}$ are equivalent for different values of r. Similarly,

$$d'_{z_*,r_1} \lesssim d_{z_*,r_2} \lesssim d'_{z_*,r_2},$$

so the pseudometrics $d'_{z_*,r}$ are equivalent for different values of r. Finally, the pseudometrics $d_{z_*,r}$ and $d'_{z_*,r}$ are equivalent.

Step 5: Let us write $z \sim z'$ if the pseudometrics $d_{z,r}$ and $d_{z',r'}$ are equivalent for some r and r' (or equivalently for all r and r'). This defines an equivalence relation on Ω_1. We claim that each equivalence class is uniformly open. To see this, fix $z_* \in \Omega^{(n)}$. Choose an $R > 0$ and $M > 0$ such that

$$\sup_{F\in\mathcal{F}} \sup_{z\in B(0,R)} \|F(z)\| \le M.$$

Suppose that $z \in B^{(nm)}(z_*, R/3)$. Then $B(z, 2R/3) \subseteq B(z_*, R)$ and hence $2R/3 < \operatorname{rad}(z, \mathcal{F})$. Also, since $B(z, 2R/3) \subseteq B(z_*, R)$, we have

$$d_{z,2R/3} \le d_{z_*,R}.$$

On the other hand, we also have $\|z_*^{(m)} - z\| < R/3$ and hence $B(z_*^{(m)}, R/3) \subseteq B(z, 2R/3)$.

Using the fact that F preserves direct sums, we have

$$\begin{aligned} d_{z_*,R/3}(F,G) &= \sup_{w \in B(z_*,R/3)} \|F(w) - G(w)\| \\ &= \sup_{w \in B(z_*,R/3)} \|F(w^{(m)}) - G(w^{(m)})\| \\ &\leq \sup_{w' \in B(z_*^{(m)},R/3)} \|F(w) - G(w)\| \\ &\leq d_{z,2R/3}(F,G). \end{aligned}$$

Therefore, we have $d_{z_*,R/3} \leq d_{z,2R/3} \leq d_{z_*,R}$ and hence $z \sim z_*$.

Step 6: Now we show that any two points $z_1 \in \Omega^{(n_1)}$ and $z_2 \in \Omega^{(n_2)}$ are equivalent. Note that because Ω is connected, there exists m such that $z_1^{(n_2 m)}$ and $z_2^{(n_1 m)}$ are in the same connected component of $\Omega^{(n_1 n_2 m)}$. As a consequence of Step 5, the equivalence classes of points in $\Omega^{(n_1 n_2 m)}$ are open subsets of $\Omega^{(n_1 n_2 m)}$. Each equivalence class in $\Omega^{(n_1 n_2 m)}$ is also relatively closed because its complement is the union of the other equivalence classes. Because $z_1^{(n_2 m)}$ and $z_2^{(n_1 m)}$ are in the same connected component, we must have $z_1^{(n_2 m)} \sim z_2^{(n_1 m)}$. As another consequence of Step 5, we have $z_1 \sim z_1^{(n_2 m)}$ and $z_2 \sim z_2^{(n_1 m)}$ and therefore $z_1 \sim z_2$.

Step 7: We have now shown that all the pseudometrics in claim (2) are uniformly equivalent. As a consequence if $d_{z_*,r}(F,G) = 0$ for some z_* and r, then this holds for all z_* and r which implies that $F = G$. Therefore, each $d_{z_*,r}$ is a metric. ☐

Corollary 3.9.7 (Identity Theorem). *Let $\Omega \subseteq M_\bullet(\mathcal{B})$ be a connected fully matricial domain, and let $F, G : \Omega \to M_\bullet(\mathcal{B})$ be fully matricial functions and $z_* \in \Omega_1^{(n_0)}$. The following are equivalent:*

(1) $\Delta^k F(z_*, \dots, z_*) = \Delta^k G(z_*, \dots, z_*)$ for all k.

(2) $F = G$ on $B(z_*, r)$ for some $r > 0$.

(3) $F = G$ on Ω.

Proof. Note that the family $\{F, G\}$ is uniformly locally bounded. Hence, this follows immediately from Theorem 3.9.6 ☐

Another consequence of the theorem is that if a sequence $(F_m)_{m \in \mathbb{N}}$ is uniformly locally bounded, and if F_m converges uniformly in a neighborhood of a point, then it converges on all of Ω in the following sense.

Definition 3.9.8. We say that a sequence $(F_m)_{m \in \mathbb{N}}$ of fully matricial functions $\Omega \to M_\bullet(\mathcal{B})$ *converges uniformly locally to F* if for every $z_0 \in \Omega^{(n_0)}$, there exists $R > 0$ such that

$$\lim_{m \to \infty} \sup_{z \in B(z_0,R)} \|F_m(z) - F(z)\| = 0.$$

Lemma 3.9.9. *If F_m is fully matricial and $F_m \to F$ uniformly locally as $m \to \infty$, then F is fully matricial.*

Proof. Note that F respects intertwinings because $zT = Tw$, then

$$F(z)T = \lim_{m\to\infty} F_m(z)T = \lim_{m\to\infty} TF_m(w) = TF(w).$$

To show that F is uniformly locally bounded, fix z_0. There exists $R > 0$ and n such that $\sup_{z\in B(z_0,R)} \|F_m(z) - F(z)\| \leq 1$. Since F_n is fully matricial, there exists r and M such $\sup_{z\in B(z_*,r)} \|F_m(z)\| \leq M$. This implies that $\|F(z)\| \leq M + 1$ for $z \in B(z_*, \min(r, R))$. □

Corollary 3.9.10. *Let Ω be a connected fully matricial domain and let $F_m : \Omega \to M_\bullet(\mathcal{B})$ be a sequence of fully matricial functions which is uniformly locally bounded. Let $z_* \in \Omega^{(n)}$. Then the following are equivalent:*

(1) For every m, the sequence $\Delta^m F_n(z_*, \ldots, z_*)$ converges with respect to $\|\cdot\|_\#$.

(2) For some $r > 0$, the sequence $(F_m)_{m\in\mathbb{N}}$ converges uniformly on $B(z_*, r)$.

(3) There exists some fully matricial function F such that $F_m \to F$ uniformly locally on Ω_1.

Proof. Suppose that $R < \operatorname{rad}(z_*, \{F_m : m \in \mathbb{N}\})$. Using the Cauchy estimates, we see that the series

$$\sum_{k=0}^{\infty} R^k \|\Delta^k (F_m - F_{m'})(z_*, \ldots, z_*)\|_\#$$

converges absolutely and the rate of convergence is independent of m and m'. Therefore, (1) occurs if and only if $(F_m)_{m\in\mathbb{N}}$ is Cauchy in $d_{z_*,R}$. Because the metrics in Theorem 3.9.6 are uniformly equivalent, they preserve Cauchy sequences. Hence, $(F_m)_{m\in\mathbb{N}}$ is Cauchy in $d_{z,r}$ for every z and $r < \operatorname{rad}(z, \{F_m : m \in \mathbb{N}\})$. This is equivalent to (2) and equivalent to (3) in light of Lemma 3.9.9. □

CHAPTER 4

Background: The $\mathcal{B}$-valued Cauchy-Stieltjes transform

4.1 Introduction

Recall that the Cauchy-Stieltjes transform of a finite measure on the real line is $g_\mu(\zeta) = \int_{\mathbb{R}} (\zeta - t)^{-1}\, d\mu(t)$. The Cauchy-Stieltjes transforms of spectral measures are an important tool for non-commutative probability both for computation and for analytic estimates. Some of its most useful properties are the following.

(1) For a compactly supported measure μ on $\mathbb{R}$, the power series coefficients of g_μ at ∞ are the moments of μ.

(2) There are simple and sharp a priori estimates on g_μ and its derivatives; for instance, if $\operatorname{Im}\zeta \geq \epsilon$, then $|\partial_\zeta^n g_\mu(\zeta)| \leq \mu(\mathbb{R})/\epsilon^{n+1}$.

(3) There are straightforward analytic conditions that test whether a function g is the Cauchy-Stieltjes transform of some measure.

Properties (2) and (3) together mean that if an analytic function g satisfies some simple analytic conditions, then we obtain much more precise analytic information about g "for free."

This chapter will prove analogous properties to (1) - (3) above for the fully matricial Cauchy-Stieltjes transform of a $\mathcal{B}$-valued law. The main theorem will be the analytic characterization of Cauchy-Stieltjes transforms due to Williams [Wil17, Theorem 3.1]. As motivation for this result, and as an ingredient for the proof, we now state the analytic characterization of Cauchy-Stieltjes transforms in the scalar case. Here $\mathbb{H}_+ = \{\zeta \in \mathbb{C} : \operatorname{Im}\zeta > 0\}$ and $\mathbb{H}_- = \{\zeta \in \mathbb{C} : \operatorname{Im}\zeta < 0\}$ are the upper and lower half-planes.

Lemma 4.1.1. *Let $g : \mathbb{H}_+ \to \mathbb{C}$. The following are equivalent:*

(1) g is the Cauchy-Stieltjes transform of a measure μ supported in $[-M, M]$.

(2) g is analytic, g maps $\mathbb{H}_+$ into $\mathbb{H}_-$, and $\tilde{g}(\zeta) = g(1/\zeta)$ has an analytic extension to $B(0, 1/M)$ satisfying $\tilde{g}(0) = 0$ and $g(\overline{\zeta}) = \overline{g(\zeta)}$.

Proof. If $g(\zeta) = \int_{\mathbb{R}} (\zeta - t)^{-1} \, d\mu(t)$, then clearly g is an analytic function $\mathbb{H}_+ \to \mathbb{H}_-$. Moreover,

$$\tilde{g}(\zeta) = \int_{\mathbb{R}} \zeta(1 - t\zeta)^{-1} \, d\mu(t)$$

which is analytic on $B(0, 1/M)$, preserves complex conjugates, and vanishes at 0.

Conversely, suppose that g satisfies these analytic conditions. Recall that if u is bounded and continuous on $\overline{\mathbb{H}}_+$ and harmonic on $\mathbb{H}_+$, then

$$u(\zeta) = -\int_{\mathbb{R}} \frac{1}{\pi} \operatorname{Im}(\zeta - t)^{-1} u(t) \, dt;$$

this is because the integral on the right hand side is harmonic and bounded with the same limiting values as u on the boundary of $\mathbb{H}_+$. Letting $u_\delta(\zeta) = \operatorname{Im} g(t + i\delta)$, we have

$$\operatorname{Im} g(\zeta + i\delta) = -\int_{\mathbb{R}} \operatorname{Im}(\zeta - t)^{-1} \frac{1}{\pi} \operatorname{Im} g(t + i\delta) \, dt.$$

Now $-\int_{\mathbb{R}} (\zeta - t)^{-1} \pi^{-1} g(t + i\delta) \, dt$ is analytic on $\mathbb{H}_+$ and has the same imaginary part as $g(\zeta)$, so they must be equal up to adding a real constant. But both functions vanish as $\zeta \to \infty$ along the positive imaginary axis, and hence

$$g(\zeta + i\delta) = \int_{\mathbb{R}} (\zeta - t)^{-1} \, d\mu_\delta(t),$$

where

$$d\mu_\delta(t) = -\frac{1}{\pi} \operatorname{Im} g(t + i\delta) \, dt.$$

We want to define μ as a weak limit of μ_δ as $\delta \to 0$. To accomplish this, we first show that μ_δ does not have much mass outside $[-R, R]$ for $R > M$.

Because $\tilde{g}(\zeta)$ is analytic on $B(0, 1/M)$, we know that for $\epsilon > 0$, we have

$$|\zeta| < \frac{1}{R} \implies |\tilde{g}(\zeta)| \leq C_R$$

for some constant $C_R > 0$. Then by Schwarz's lemma for functions on the disk, we have

$$|\zeta| < \frac{1}{R} \implies |\tilde{g}(\zeta)| \leq C_R R |\zeta|.$$

Therefore,

$$|\zeta| > R \implies |g(\zeta)| \leq \frac{C_R'}{|\zeta|}.$$

Now $\operatorname{Im} g = 0$ on $\mathbb{R} \setminus [-M, M]$ and hence for $|t| > R$,

$$\begin{aligned} |\operatorname{Im} g(t + i\delta)| &= |\operatorname{Im} g(t + i\delta) - \operatorname{Im} g(t)| \\ &\leq |g(t + i\delta) - g(t)| \\ &\leq \delta \sup_{s \in [0, \delta]} |g'(t + is)|. \end{aligned}$$

Now $B(t+is, \frac{1}{2}(|t|-R)) \subseteq \{|\zeta| > R+\frac{1}{2}(|t|-R)\}$ where g is bounded by $C_R'/(R+(1/2)(|t|-R)) = 2C_R'/(|t|+R)$, and hence by the Cauchy estimates on derivatives,

$$|g'(t+is)| \leq \frac{2}{|t|-R} \frac{2C_R'}{|t|+R} = \frac{4C_R'}{|t|^2-R^2}.$$

Thus, since $\operatorname{Im} g(t) = 0$, we have

$$|\operatorname{Im} g(t+i\delta)| \leq \frac{4C_R'\delta}{|t|^2-R^2}.$$

In particular, letting $M < R' < R$, we have

$$\mu_\delta(\mathbb{R} \setminus [-R,R]) \leq \delta \int_{\mathbb{R}\setminus[-R,R]} \frac{4C_{R'}'}{|t|^2-(R')^2}\, dt,$$

where the integral is finite. Therefore, for each point ζ, we have

$$g(\zeta) = g(\zeta+i\delta) + O(\delta) = \int_{-R}^{R} (\zeta-t)^{-1}\, d\mu_\delta(z) + O(\delta)$$

Moreover,

$$\mu_\delta([-R,R]) \leq \int_{-R}^{R} \frac{2R^2}{t^2+R^2}\, d\mu_\delta(t) = -2R \operatorname{Im} g(iR+i\delta).$$

Thus, the measures $\mu_\delta|_{[-R,R]}$ have uniformly bounded mass, and hence this family of measures is precompact. Therefore, for each R, there exists a sequence δ_n such that $\mu_{\delta_n}|_{[-R,R]}$ converges to some limit μ supported on $[-R,R]$ as $\delta \to 0$. In the limit, we have

$$g(\zeta) = \int (\zeta-t)^{-1}\, d\mu(t).$$

Thus, for each $R > M$, we have $g = g_\mu$ for some μ supported on $[-R,R]$. The moments of μ are uniquely determined by the power series expansion of g at ∞, hence μ is unique. Then μ is supported in $[-R,R]$ for every $R > M$, so that μ is supported in $[-M,M]$. □

4.2 Definition

We have seen in §3.1 that the Cauchy-Stieltjes transform of a $\mathcal{B}$-valued law should be viewed as a fully matricial function over $\mathcal{B}$ rather than simply a $\mathcal{B}$-valued function. To give the full definition, we must first define the natural domain for the Cauchy-Stieltjes transform, which consists of operators with positive imaginary part. Thus, we begin with the basic properties of real and imaginary parts of operators.

Notation 4.2.1. For $z \in M_n(\mathcal{B})$, we define $\operatorname{Re}(z) = \frac{1}{2}(z+z^*)$ and $\operatorname{Im}(z) = \frac{1}{2i}(z-z^*)$.

Observation 4.2.2. *The operators* $\mathrm{Re}(z)$ *and* $\mathrm{Im}(z)$ *are self-adjoint and* $z = \mathrm{Re}(z) + i\,\mathrm{Im}(z)$. *Moreover, if* $\mathcal{H}$ *is a Hilbert space or a right Hilbert* $\mathcal{B}$*-module and* $z \in B(\mathcal{H})$ *and* $\xi \in \mathcal{H}$*, then*

$$\mathrm{Re}\langle \xi, z\xi\rangle = \langle \xi, \mathrm{Re}(z)\xi\rangle \qquad \mathrm{Im}\langle \xi, z\xi\rangle = \langle \xi, \mathrm{Im}(z)\xi\rangle.$$

Lemma 4.2.3. *Suppose that* $z \in M_n(\mathcal{B})$ *and* $\mathrm{Im}\, z \geq \epsilon$*, where* $\epsilon \in (0, \infty)$ *and the inequality holds in* $M_n(\mathcal{B})$*. Then* z *is invertible with* $\|z^{-1}\| \leq 1/\epsilon$ *and* $\mathrm{Im}(z^{-1}) \leq -\epsilon/\|z\|^2$*.*

Proof. Recall that $M_n(\mathcal{B})$ is a C^*-algebra and hence can be realized as a concrete C^*-algebra of operators on a Hilbert space $\mathcal{H}$. Then observe that for $\xi \in \mathcal{H}$, we have

$$\|\xi\|\|z\xi\| \geq |\langle \xi, z\xi\rangle| \geq \mathrm{Im}\langle \xi, z\xi\rangle = \langle \xi, (\mathrm{Im}\, z)\xi\rangle \geq \langle \xi, \epsilon\xi\rangle = \epsilon\|\xi\|^2,$$

which shows that $\|z\xi\| \geq \epsilon\|\xi\|$ and hence $\ker z = 0$ and $\mathrm{Ran}\, z$ is closed. On the other hand, we have $\mathrm{Im}\, z^* = -\mathrm{Im}\, z \leq -\epsilon$, so similar reasoning shows that $\|z^*\xi\| \geq \epsilon\|\xi\|$ which implies that $\ker z^* = 0$ and hence $\mathrm{Ran}(z) = \mathcal{H}$. Since $\ker z = 0$ and $\mathrm{Ran}(z) = \mathcal{H}$, it follows that z is invertible as a linear operator. Because of the estimate $\|z\xi\| \geq \epsilon\|\xi\|$, we know that z^{-1} is bounded with $\|z^{-1}\| \leq 1/\epsilon$.

Finally, to show that $\mathrm{Im}(z^{-1}) \leq -\epsilon/\|z\|^2$, note that for $\xi \in \mathcal{H}$, we have

$$\begin{aligned}\mathrm{Im}\langle \xi, z^{-1}\xi\rangle = \mathrm{Im}\langle zz^{-1}\xi, z^{-1}\xi\rangle &= \mathrm{Im}\langle z^{-1}\xi, z^*(z^{-1}\xi)\rangle \\ &= -\mathrm{Im}\langle z^{-1}\xi, z(z^{-1}\xi)\rangle \leq -\epsilon\|z^{-1}\xi\|^2 \leq \frac{-\epsilon}{\|z\|^2}\|\xi\|^2,\end{aligned}$$

using the fact that $\|\xi\| = \|zz^{-1}\xi\| \leq \|z\|\|z^{-1}\xi\|$. ☐

Definition 4.2.4 (Fully Matricial Upper/Lower Half-plane)**.** We define $\mathbb{H}_{+,\epsilon}^{(n)}(\mathcal{B}) = \{z \in M_n(\mathcal{B}) : \mathrm{Im}\, z \geq \epsilon\}$ and define $\mathbb{H}_+^{(n)}(\mathcal{B}) = \bigcup_{\epsilon>0} \mathbb{H}_{+,\epsilon}^{(n)}(\mathcal{B})$. Finally, we define the *fully matricial upper half-plane* as $\mathbb{H}_+(\mathcal{B}) = (\mathbb{H}_+^{(n)}(\mathcal{B}))_{n\in\mathbb{N}}$.

Similarly, we define $\mathbb{H}_{-,\epsilon}^{(n)}(\mathcal{B}) = \{z \in M_n(\mathcal{B}) : \mathrm{Im}\, z \leq -\epsilon\}$ and $\mathbb{H}_-^{(n)}(\mathcal{B}) = \bigcup_{\epsilon>0} \mathbb{H}_-^{(n)}(\mathcal{B})$.

Finally, we define $\overline{\mathbb{H}}_\pm^{(n)}(\mathcal{B}) = \{z \in M_n(\mathcal{B}) : \pm\mathrm{Im}\, z \geq 0\}$.

Observation 4.2.5. $\mathbb{H}_+(\mathcal{B})$ *and* $\mathbb{H}_-(\mathcal{B})$ *are connected fully matricial domains (although* $\overline{\mathbb{H}}_+(\mathcal{B})$ *and* $\overline{\mathbb{H}}_-(\mathcal{B})$ *are not because they fail to be open).*

Proof. To see that $\mathbb{H}_+(\mathcal{B})$ respects direct sums, suppose $z_1 \in \mathbb{H}_+^{(n_1)}(\mathcal{B})$ and $z_2 \in \mathbb{H}_+^{(n_2)}(\mathcal{B})$. Then $\mathrm{Im}\, z_1 \geq \epsilon_1$ and $\mathrm{Im}\, z_2 \geq \epsilon_2$ for some ϵ_1, $\epsilon_2 > 0$. Then $\mathrm{Im}(z_1 \oplus z_2) = \mathrm{Im}\, z_1 \oplus \mathrm{Im}\, z_2 \geq \min(\epsilon_1, \epsilon_2)$, so that $z_1 \oplus z_2 \in \mathbb{H}_+^{(n_1+n_2)}(\mathcal{B})$.

To see that $\mathbb{H}_+(\mathcal{B})$ is uniformly open, suppose that $z \in \mathbb{H}_+^{(n)}(\mathcal{B})$. If $\mathrm{Im}\, z \geq \epsilon > 0$, then we have $B(z, \epsilon) \subseteq \mathbb{H}_+(\mathcal{B})$. Indeed, if $z' \in B^{(mn)}(z^{(n)}, \epsilon)$, then

$$\mathrm{Im}\, z' \geq \mathrm{Im}(z^{(n)}) - \|z^{(m)} - z'\| = (\mathrm{Im}\, z)^{(n)} - \|z^{(m)} - z'\| \geq \epsilon - \|z^{(m)} - z'\| > 0.$$

Furthermore, each $\mathbb{H}_+^{(n)}(\mathcal{B})$ is non-empty and connected (in fact, convex), and hence $\mathbb{H}_+(\mathcal{B})$ is non-empty and connected. The argument for $\mathbb{H}_-(\mathcal{B})$ is symmetrical. ☐

Definition 4.2.6 (Cauchy-Stieltjes Transform)**.** Let $\sigma : \mathcal{B}\langle X\rangle \to \mathcal{B}$ be a generalized law. We define the *Cauchy-Stieltjes transform* G_σ as the sequence of functions $G_\sigma^{(n)} : \mathbb{H}_+^{(n)}(\mathcal{B}) \to \mathbb{H}_{-,0}^{(n)}(\mathcal{B})$ given by

$$G_\sigma^{(n)}(z) = \overline{\sigma}^{(n)} \left[(z - \overline{X}^{(n)})^{-1}\right],$$

where $\overline{X}$ is the operator of left multiplication by X on $\mathcal{B}\langle X\rangle \otimes_\sigma \mathcal{B}$ and $\overline{\sigma}(b) = \langle 1\otimes 1, b(1\otimes 1)\rangle_\sigma$ (as in Theorem 2.6.6).

Note here that the definition makes sense because if $z \in \mathbb{H}_+^{(n)}$, then for some $\epsilon > 0$, we have

$$\operatorname{Im}(z - \overline{X}^{(n)}) = \operatorname{Im}(z) \geq \epsilon,$$

which implies that $z - \overline{X}^{(n)}$ is invertible.

Lemma 4.2.7. *For a generalized law σ, the Cauchy-Stieltjes transform G_σ is a fully matricial function. We also have*

$$z \in \mathbb{H}_{+,\epsilon}^{(n)}(\mathcal{B}) \implies \|G_\sigma(z)\| \leq \frac{\|\sigma(1)\|}{\epsilon}. \tag{4.1}$$

Proof. Let $\mathcal{B}$ be the C*-algebra generated by $\mathcal{B}$ and $\overline{X}$. Note that the inclusion $\mathcal{B} \to \mathcal{B}$ is fully matricial. Moreover, the function $z \mapsto z - \overline{X}^{(n)}$ is the sum of two fully matricial functions, so it is fully matricial on $M_\bullet(\mathcal{B})$ and in particular on $\mathbb{H}_+(\mathcal{B})$. Since inv fully matricial, so is $(z - \overline{X}^{(n)})^{-1}$. Finally, $\overline{\sigma}$ is a completely bounded linear map and hence is fully matricial by Proposition 3.6.1, so $\overline{\sigma}^{(n)}[(z - X^{(n)})^{-1}]$ is fully matricial. □

In the future, we will simplify and slightly abuse notation by writing

$$G_\sigma^{(n)}(z) = \sigma^{(n)} \left[(z - X^{(n)})^{-1}\right],$$

that is, writing σ instead of $\overline{\sigma}$ even though σ is technically only defined on $\langle A\rangle\langle X\rangle$ and writing X for the multiplication operator $\overline{X}$.

4.3 Derivatives and expansion at ∞

Lemma 4.3.1. *Let $z_j \in M_{n_j}(\mathcal{B})$ and $w_j \in M_{n_{j-1}\times n_j}(\mathcal{B})$. Then*

$$\Delta^k G_\sigma(z_0, \dots, z_k)[w_1, \dots, w_k] \\ = (-1)^k \sigma^{(n_0\times n_k)}[(z_0 - X^{(n_0)})^{-1} w_1 (z_1 - X^{(n_1)})^{-1} \dots w_k (z - X^{(n_k)})^{-1}],$$

and in particular if $\operatorname{Im} z_j \geq \epsilon_j$ *for* $\epsilon_j \in (0, \infty)$*, then*

$$\|\Delta^k G_\sigma(z_0, \dots, z_k)\| \leq \frac{\|\sigma(1)\|}{\epsilon_0 \dots \epsilon_k}.$$

Proof. Denote

$$Z = (z_0 - X^{(n_0)}) \oplus \cdots \oplus (z_k - X^{n_k}).$$

Fix small scalars $\zeta_1, \ldots, \zeta_k$ and define

$$W = \begin{bmatrix} 0 & \zeta_1 w_1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \zeta_2 w_2 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots 0 & \zeta_k w_k & \\ 0 & 0 & 0 & \ldots & 0 & 0 \end{bmatrix}]$$

Note that if $\zeta_1, \ldots, \zeta_k$ are sufficiently small, then

$$\begin{aligned} \sigma^{(n)}[(Z + W - X^{(n)})^{-1}] &= \sigma^{(n)}[Z^{-1}(1 + W(Z - X^{(n)})^{-1})^{-1}] \\ &= \sum_{j=0}^{k} \sigma^{(n)}[(Z - X^{(n)})^{-1}(W(Z - X^{(n)})^{-1})^j], \end{aligned}$$

where the expansion is truncated because WZ^{-1} is nilpotent. By looking at the upper right block, we obtain the desired formula for $\Delta^k G_\sigma$, and the upper bound for $\|\Delta^k G_\sigma\|$ follows immediately using Lemma 4.2.3. □

Notation 4.3.2. We write $\tilde{G}_\sigma(z) = G_\sigma(z^{-1})$ for all z where it is defined.

Lemma 4.3.3. *Suppose that σ is a generalized law with* $\text{rad}(\sigma) \leq M$*. Then $\tilde{G}_\sigma$ has a fully matricial extension to $B(0, 1/M)$ given by*

$$\tilde{G}_\sigma(z) = \sigma^{(n)}[z(1 - X^{(n)}z)^{-1}] = \sum_{k=0}^{\infty} \sigma^{(n)}[z(X^{(n)}z)^k].$$

Proof. First, we observe that if $z^{-1} \in \mathbb{H}_+^{(n)}(\mathcal{B})$, then

$$\begin{aligned} \tilde{G}_\sigma(z) &= \sigma^{(n)}[(z^{-1} - X^{(n)})^{-1}] \\ &= \sigma^{(n)}[(z^{-1} - X^{(n)})^{-1}] \\ &= \sigma^{(n)}[((1 - X^{(n)}z)z^{-1})^{-1}] \\ &= \sigma^{(n)}[z(1 - X^{(n)}z)^{-1}]. \end{aligned}$$

However, the latter function is also defined whenever $\|z\| < 1/M$. Now we claim that this extension of $\tilde{G}_\sigma$ is fully matricial on the domain

$$\Omega^{(n)} := \left\{ z : (1 - \overline{X}^{(n)} z)^{-1} \text{ is invertible} \right\} \supseteq B(0, 1/M) \cup \mathbb{H}_+(\mathcal{B}) \cup \mathbb{H}_-(\mathcal{B}),$$

where $\overline{X}$ is the multiplication operator on $\mathcal{B}\langle X \rangle \otimes_\sigma \mathcal{B}$. The argument that $\Omega^{(n)}$ is a matricial domain is similar to the argument that invertible elements of a C^*-algebra form a matricial domain. Moreover, $\overline{\sigma}^{(n)}[z(1 - \overline{X}^{(n)} z)^{-1}]$ is fully matricial on Ω because it a built out of the inclusion $\mathcal{B} \to \mathcal{B}$ by translation, inverse, products, and application of $\overline{\sigma}$. □

Lemma 4.3.4. *We have*

$$\Delta^k \tilde{G}_\sigma(0^{(n_0)}, \dots, 0^{(n_k)})[w_1, \dots, w_k] = \sigma^{(n_0 \times n_k)}[w_1 X^{(n_1)} w_2 \dots X^{(n_{k-1})} w_k].$$

Proof. From the geometric series expansion, we have for $z \in B^{(n)}(0, 1/M)$ that

$$\tilde{G}_\sigma(z) = \sum_{k=0}^\infty \sigma^{(n)} \left[z (X^{(n)} z)^k \right].$$

If we let Λ_k be the multilinear form

$$\Lambda_k(z_1, \dots, z_k) = \sigma[z_1 X z_2 \dots X z_k],$$

then for every n and every $z \in B^{(n)}(0^{(n)}, 1/M)$, we have

$$\tilde{G}_\sigma(z) = \sum_{k=0}^\infty \Lambda_k^{(n)}(z, \dots, z).$$

Therefore, by Lemma 3.5.10, we have $\Lambda_k = \Delta^k \tilde{G}_\sigma(0, \dots, 0)$. The general formula for $\Delta^k \tilde{G}_\sigma(0^{(n_0)}, \dots, 0^{(n_k)})$ follows from Lemma 3.5.2. □

Lemma 4.3.5. *If $\|z\| < 1/M$, then we have*

$$\left\| \tilde{G}_\sigma(z) \right\| \leq \frac{\|\sigma(1)\| \|z\|}{1 - M\|z\|}$$

In particular,

$$\|z\| < 1/(M + \epsilon) \implies \left\| \tilde{G}_\sigma(z) \right\| \leq \frac{\|\sigma(1)\|}{\epsilon}.$$

Proof. This follows by applying the triangle inequality to the geometric series expansion. □

4.4 Analytic characterization

The following theorem is due to Williams [Wil17, Theorem 3.1] and Anshelevich-Williams [AW16, Theorem A.1]. While we have nothing to add in terms of the proof, we aim to improve the organization and explanation of details and thus to present the proof in a more "textbook-like" manner.

Theorem 4.4.1. *Let $G^{(n)} : \mathbb{H}_+^{(n)}(\mathcal{B}) \to M_n(\mathcal{B})$. The following are equivalent:*

(1) $G = G_\sigma$ for some generalized law σ with $\operatorname{rad}(\sigma) \leq M$ *if and only if the following conditions hold.*

(2) The following conditions hold:

(a) G is fully matricial.

(b) G maps $\mathbb{H}_+^{(n)}(\mathcal{B})$ into ${}^{(n)}_{-,0}(\mathcal{B})$.

(c) $\tilde{G}(z) = G(z^{-1})$ has a fully matricial extension to $B(0, 1/M)$.

(d) This extension satisfies $\tilde{G}(0) = 0$ and $\tilde{G}(z^*) = \tilde{G}(z)^*$.

(e) For every $\epsilon > 0$, there exists $C_\epsilon > 0$ such that $\|z\| \leq 1/(M+\epsilon)$ implies $\|\tilde{G}(z)\| \leq C_\epsilon$.

Proof of (1) $\implies$ (2). Assume that (1) holds. We have already shown that (a), (b), (c), and (e) hold in Lemmas 4.2.7, 4.3.3, 4.3.5. Moreover, (d) follows from power series expansion in Lemma 4.3.3. □

The proof of (2) $\implies$ (1) is more involved, so we will prove several lemmas before concluding the proof of the Theorem. First, we define the map $\sigma : \mathcal{B}\langle X\rangle \to \mathcal{B}$. The correct choice of σ is clear in light of Lemma 4.3.4.

Lemma 4.4.2. *Let G satisfy (2) of Theorem 4.4.1. Define $\sigma : \mathcal{B}\langle X\rangle \to \mathcal{B}$ by*

$$\sigma(z_0 X z_1 \dots X z_k) = \Delta^{k+1}\tilde{G}(0, \dots, 0)[z_0, \dots, z_k].$$

Then any $R > M$ is an exponential bound for $\sigma^{(n)}$.

Proof. Because $\|z\| \leq 1/R$ implies $\|\tilde{G}(z)\| \leq C_{R-M}$, we have by Lemma 3.4.2 that

$$\|\Delta^k \tilde{G}(0^{(n)}, \dots, 0^{(n)})\| \leq C_{R-M} R^k.$$

□

Next, we show that σ extends to the analytic completion of $\mathcal{B}\langle X\rangle$. Fix $R > M$. As in the proof of Theorem 2.6.5, we define a norm on $M_n(\mathcal{B}\langle X\rangle) = M_n(\mathcal{B})\langle X^{(n)}\rangle$ by

$$\|F(X^{(n)})\|_R = \inf\left\{\sum_{j=1}^n \mathfrak{p}(F_j) : F_j \text{ monomials and } f = \sum_{j=1}^n F_j\right\},$$

where $\mathfrak{p}(z_0 X^{(n)} z_1 \dots X^{(n)} z_k) = R^k \|z_0\| \dots \|z_k\|$ for $z_0, \dots, z_k \in M_n(\mathcal{B})$. We denote the completion by $\mathcal{B}\langle X\rangle_R^{(n)}$. Recall that this is a Banach $*$-algebra.

Lemma 4.4.3. *Fix $R > M$. Then the map $\sigma^{(n)}$ defined above extends to a bounded map $\mathcal{B}\langle X\rangle_R^{(n)} \to M_n(\mathcal{B})$. Moreover, if $\|z\|_R < 1/R$, then $1 - X^{(n)}z$ is invertible in $\mathcal{B}\langle X\rangle_R^{(n)}$ and we have*

$$\tilde{G}(z) = \sigma^{(n)}[z(1 - X^{(n)}z)^{-1}].$$

Proof. The first claim follows because $\|\sigma^{(n)}(F(X))\| \leq C_{R-M}\|F(X)\|_R$ since R is an exponential bound for σ. Next, suppose that $\|z\|_R \leq 1/R$. Then because the geometric series $(1 - X^{(n)}z)^{-1}$ converges in $\mathcal{B}\langle X\rangle_R^{(n)}$, we see that $1 - X^{(n)}z$ is invertible. Moreover, a direct power series computation shows that $\tilde{G}(z) = \sigma^{(n)}[z(1 - X^{(n)}z)^{-1}]$ after we invoke Lemma 3.5.7. □

With these preparations in order, we can begin to prove complete positivity of σ. We start out by proving that certain symmetric moments are positive.

Lemma 4.4.4. *Suppose that G satisfies (2) of Theorem 4.4.1 and define σ as in Lemma 4.4.2. Let B_0 and B_1 be self-adjoint elements of $M_n(\mathcal{B})$ with $B_1 \geq \epsilon > 0$. Then*

$$\sigma^{(n)}\left[\left(B_1(X^{(n)} + B_0)\right)^{2k} B_1\right] \geq 0.$$

Proof. Fix B_0 and B_1 and let ϕ be a state on $M_n(\mathcal{B})$. Consider the scalar-valued function $g : \mathbb{H}_+ \to \overline{\mathbb{H}}_-$ given by

$$g(\zeta) = \phi \circ G(B_1^{-1}\zeta - B_0).$$

Now we analyze the behavior of g at ∞. Note that $\zeta^{-1}B_1^{-1} - B_0$ is invertible in $\mathcal{B}\langle X\rangle_R^{(n)}$ if ζ is small enough. In fact, for sufficiently small ζ, we have $\|(\zeta^{-1}B_1^{-1} - B_0)^{-1}\| < 1/R$. Thus, we have

$$\begin{aligned} g(1/\zeta) &= \phi \circ \tilde{G}((B_1^{-1}\zeta^{-1} - B_0)^{-1}) \\ &= \phi \circ \sigma[(B_1^{-1}\zeta^{-1} - B_0)^{-1}(1 - X^{(n)}(B_1^{-1}\zeta^{-1} - B_0)^{-1})^{-1}] \\ &= \phi \circ \sigma[(B_1^{-1}\zeta^{-1} - B_0 - X^{(n)})^{-1}] \\ &= \phi \circ \sigma[B_1\zeta(1 - (X^{(n)} + B_0)B_1\zeta)^{-1}] \\ &= \sum_{k=0}^{\infty} \zeta^{k+1}\, \phi \circ \sigma[(B_1(X^{(n)} + B_0))^k B_1], \end{aligned}$$

where the intermediate steps are performed in $\mathcal{B}\langle X\rangle_R^{(n)}$. In particular, $\tilde{g}(\zeta) = g(1/\zeta)$ extends to be analytic in a neighborhood of 0. Because $\tilde{G}$ preserves adjoints, we have $g(\overline{\zeta}) = \overline{g(\zeta)}$. Therefore, g is the Cauchy-Stieltjes transform of some compactly supported measure ρ on $\mathbb{R}$. Moreover, by examining the power series coefficients of $\tilde{g}$ at 0, we have

$$\phi \circ \sigma\left[\left(B_1(X^{(n)} + B_0)\right)^{2k} B_1\right] = \int_{\mathbb{R}} t^{2k}\, d\rho(t) \geq 0.$$

Because this holds for every state ϕ, we have $\sigma[(B_1(X^{(n)} + B_0))^{2k}B_1] \geq 0$ by Proposition 2.1.8 (5). $\square$

Lemma 4.4.5. *Let G satisfy (2) and let σ be as above. Let $F(Y) = C_0YC_1 \dots YC_k$ be a monomial in $M_n(\mathcal{B})\langle Y\rangle$ and let $B_0 \in M_n(\mathcal{B})$ be self-adjoint. Then*

$$\sigma^{(n)}\left(F\left(X^{(n)} + B_0\right)^* F\left(X^{(n)} + B_0\right)\right) \geq 0.$$

Proof. Let us write $Y = X^{(n)} + B_0$. Consider the matrix

$$C_\delta = \begin{bmatrix} \delta & \delta^2 C_k^* & 0 & \dots & 0 & 0 & 0 \\ \delta^2 C_k & \delta & \delta^2 C_{k-1}^* & \dots & 0 & 0 & 0 \\ 0 & \delta^2 C_{k-1} & \delta & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \delta & \delta^2 C_2^* & 0 \\ 0 & 0 & 0 & \dots & \delta^2 C_2 & \delta & \delta^2 C_1^* \\ 0 & 0 & 0 & \dots & 0 & \delta^2 C_1 & \delta + \delta^{-4k} C_0^* C_0 \end{bmatrix}$$

Observe that if δ is sufficiently small, then $C_\delta \geq \epsilon$ for some $\epsilon \in (0, \infty)$. Indeed, the diagonal terms δ will be much larger than the off-diagonal terms, while the extra diagonal term $\delta^{-4k} C_k^* C_k$ is already positive. Therefore, by the previous lemma,

$$\sigma^{(n)} \left[\left(C_\delta Y^{(k+1)} \right)^{2k} C_\delta \right] \geq 0.$$

We claim that the top left $n \times n$ block of $(C_\delta Y^{(k+1)})^{2k} C_\delta$ is equal to $F(Y)^* F(Y) + O(\delta)$. To see this, consider what happens when we multiply out $(C_\delta (X^{(n(k+1))} + B_0^{(k+1)}))^{2k} C_\delta$ using matrix multipication, treating each $n \times n$ block as a unit. The top left block of the product will be the sum of terms of the form

$$(C_\delta)_{1,i_1} Y (C_\delta)_{i_1,i_2} Y \dots (C_\delta)_{i_{k-2},i_{k-1}} Y (C_\delta)_{i_{k-1},1}$$

since $Y^{(k+1)}$ is a block diagonal matrix. Because C_δ is tridiagonal, the sequence of indices must have $|i_{j-1} - i_j| \leq 1$. We can picture such a sequence as a path in the graph with vertices $\{1, \dots, k+1\}$ and edges between j and $j+1$ and a self-loop at each vertex j.

All the entries in C_δ are $O(\delta)$ except the bottom right entry with the term $\delta^{-4k} C_k^* C_k$. Thus, any path which yields a term larger than $O(\delta)$ must reach the last vertex $k+1$ and use the self-loop at the vertex $k+1$. But if we travel along the path at a speed ≤ 1, the only way we can get from 1 to $k+1$, use the self-loop at $k+1$, and get then back to 1 in $2k+1$ steps is to follow the path

$$1, 2, \dots, k, k+1, k+1, k, \dots, 2, 1.$$

So the only term in the sum which is not $O(\delta)$ is the term

$$(\delta^2 C_k^*) Y \dots (\delta^2 C_1^*) Y (\delta + \delta^{-4k} C_0^* C_0) Y (\delta^2 C_1) \dots Y (\delta^2 C_k) = F(Y)^* F(Y) + O(\delta).$$

Hence, the upper left entry of $(C_\delta Y^{(k+1)})^{2k} C_\delta$ is $F(Y)^* F(Y) + O(\delta)$. As a consequence,

$$\sigma^{(n)}(F(Y)^* F(Y)) + O(\delta) \geq 0,$$

and thus by taking δ to zero, we have $\sigma^{(n)}(F(Y)^* F(Y)) \geq 0$. □

To finish the proof that $\sigma^{(n)}(P(X)^*P(X)) \geq 0$ for every P, we will use the following matrix amplification trick to reduce to the case of a monomial.

Lemma 4.4.6. *Let $P(X) \in M_n(\mathcal{B}\langle X \rangle)$ be a polynomial of degree d. Denote*

$$\widehat{X} = \begin{bmatrix} X & 1 \\ 1 & X \end{bmatrix}.$$

Then for some m, there exist matrices $C_0, \ldots, C_d \in M_{2m}(\mathcal{B})$ such that

$$\begin{bmatrix} P(X) & 0 \\ 0 & 0 \end{bmatrix} = C_0 \widehat{X}^{(m)} C_1 \widehat{X}^{(m)} \ldots C_{d-1} \widehat{X}^{(m)} C_d.$$

Proof. Fix d. Let $\Gamma^{(n)}$ be the set of all polynomials $\mathcal{B}\langle X \rangle^{(n)}$ of degree $\leq d$ which can be expressed as in the conclusion of the lemma.

First, we claim that $\Gamma^{(1)}$ contains the monomials in $\mathcal{B}\langle X \rangle$. Let $p(X) = a_0 X a_1 \ldots X a_k$ be a monomial of degree $k \leq d$. Then we have

$$\begin{bmatrix} p(X) & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a_0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X & 1 \\ 1 & X \end{bmatrix} \begin{bmatrix} a_1 & 0 \\ 0 & 0 \end{bmatrix} \cdots \begin{bmatrix} X & 1 \\ 1 & X \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X & 1 \\ 1 & X \end{bmatrix} \begin{bmatrix} a_k & 0 \\ 0 & 0 \end{bmatrix}$$
$$\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X & 1 \\ 1 & X \end{bmatrix} \right)^{d-k} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Next, we claim that if $P(X) \in \Gamma^{(n)}$ and $e_{i,j}$ is the (i,j) matrix unit in $M_k(\mathbb{C})$, then the matrix $P(X) \otimes e_{i,j}$ with $P(X)$ in the (i,j) block and zeroes elsewhere is in $\Gamma^{(nk)}$. Given such a $P(X)$, there exist $C_1, \ldots, C_d$ in $M_{2m}(\mathcal{B})$ such that

$$\begin{bmatrix} P(X) & 0 \\ 0 & 0 \end{bmatrix} = C_0 \widehat{X}^{(m)} C_1 \widehat{X}^{(m)} \ldots C_{d-1} \widehat{X}^{(m)} C_d.$$

Then observe the $2(m + n(k-1))$ by $2(m + n(k-1))$ matrix equation:

$$\begin{bmatrix} P(X) \otimes e_{i,j} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1_n \otimes e_{i,1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} C_0 & 0 \\ 0 & 0 \end{bmatrix} \widehat{X}^{(m+n(k-1))} \begin{bmatrix} C_1 & 0 \\ 0 & 0 \end{bmatrix} \cdots$$
$$\cdots \widehat{X}^{(m+n(k-1))} \begin{bmatrix} C_d & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1_n \otimes e_{1,j} & 0 \\ 0 & 0 \end{bmatrix}.$$

We caution the reader that the blocks C_j are $2m \times 2m$ while the blocks $P(X) \times e_{i,j}$ and $1_n \otimes e_{i,j}$ are $nk \times nk$.

Finally, we claim that $\Gamma^{(n)}$ is closed under addition. Suppose that $P(X)$ and $Q(X)$ are in $\Gamma^{(n)}$. Then there exist integers r and s and matrices $B_1, \ldots, B_d \in M_{2r}(\mathcal{B})$ and $C_1, \ldots, C_d \in M_{2s}(\mathcal{B})$ such that

$$\begin{bmatrix} P(X) & 0 \\ 0 & 0 \end{bmatrix} = B_0 \widehat{X}^{(r)} B_1 \widehat{X}^{(r)} \ldots B_{d-1} \widehat{X}^{(r)} B_d$$

and

$$\begin{bmatrix} Q(X) & 0 \\ 0 & 0 \end{bmatrix} = C_0 \widehat{X}^{(s)} C_1 \widehat{X}^{(s)} \dots B_{d-1} \widehat{X}^{(s)} B_d.$$

Then observe that

$$\begin{bmatrix} P(X) + Q(X) & 0 \\ 0 & 0 \end{bmatrix} = S \begin{bmatrix} B_0 & 0 \\ 0 & C_0 \end{bmatrix} \widehat{X}^{(r+s)} \begin{bmatrix} B_1 & 0 \\ 0 & C_1 \end{bmatrix} \dots \widehat{X}^{(r+s)} \begin{bmatrix} B_d & 0 \\ 0 & C_d \end{bmatrix} S^*$$

Where

$$S = \begin{bmatrix} 1_{n\times n} & 0_{n\times(n-r)} & 1_{n\times n} & 0_{n\times(n-s)} \\ 0_{(r+s-n)\times n} & 0_{(r+s-n)\times(n-r)} & 0_{(r+s-n)\times n} & 0_{(r+s-n)\times(s-n)} \end{bmatrix}$$

Altogether, we have shown that $\Gamma = \bigcup_{n=1}^\infty \Gamma^{(n)}$ contains the 1×1 monomials of degree $\leq d$, is closed under $P \mapsto P \otimes E_{i,j}$, and is closed under addition. This implies that Γ contains all matrix polynomials of degree $\leq d$ as desired. □

Conclusion to the proof of Theorem 4.4.1. Suppose that G satisfies (2) of the theorem and let $\sigma : \mathcal{B}\langle X \rangle \to \mathcal{B}$ be given as in Lemma 4.4.2. To show that σ is completely positive, choose a polynomial $P(X) \in M_n(\mathcal{B}\langle X \rangle)$. Let

$$B_0 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \in M_2(\mathcal{B}).$$

Then by Lemma 4.4.6, we can write $P(X)$ in the form

$$\begin{bmatrix} P(X) & 0 \\ 0 & 0 \end{bmatrix} = C_0(X^{(2m)} + B_0^{(m)}) C_1 (X^{(2m)} + B_0^{(m)}) \dots C_{d-1}(X^{(2m)} + B_0^{(m)}) C_d,$$

where $C_j \in M_{2m}(\mathcal{B})$. Thus, by Lemma 4.4.5, we have

$$\sigma^{(2m)} \big[C_d^*(X^{(2m)} + B_0^{(m)}) C_{d-1}^* \dots (X^{(2m)} + B_0^{(m)}) C_0^* \\ C_0 (X^{(2m)} + B_0^{(m)}) \dots C_{d-1}(X^{(2m)} + B_0^{(m)}) C_d \big] \geq 0,$$

which implies that $\sigma^{(n)}(P(X)^* P(X)) \geq 0$.

Next, we have shown in Lemma 4.4.2 that σ exponentially bounded by R whenever $R > M$. Therefore, σ is a generalized law with $\operatorname{rad}(\sigma) \leq M$.

It remains to show that the Cauchy transform of σ is the original function G. It follows from Lemma 4.4.3 that $\tilde{G}(z) = \tilde{G}_\sigma(z)$ when $\|z\| < 1/R$. If we let $z_0 = 2iR$, then we have $z \in B^{(n)}(z_0, R)$ implies that $\operatorname{Im} z \geq R + \epsilon$ for some $\epsilon > 0$ which implies that $z^{-1} \in B^{(n)}(0, 1/R)$. Hence, we have $G = G_\sigma$ on $B(z_0, R)$. So by the identity theorem (Theorem 3.9.7), we have $G = \tilde{G}_\sigma$ on the whole matricial upper half-plane. □

We now give an analytic characterization of when the generalized law σ is a law, and hence complete the analytic characterization of the Cauchy-Stieltjes transforms of $\mathcal{B}$-valued laws.

Lemma 4.4.7. *Let σ be a $\mathcal{B}$-valued generalized law. Then the following are equivalent.*

(1) σ is a law.

(2) $\Delta\tilde{G}_\sigma(0,0)[z] = z$ for all $z \in \mathcal{B}$.

(3) For each n, $\lim_{\|z\|\to 0} z^{-1}\tilde{G}_\sigma(z) = 1_n$, where the limit occurs in norm and is taken over all invertible z in $M_n(\mathcal{B})$.

Proof. We have $\Delta\tilde{G}_\sigma(0,0)[z] = \sigma(z)$ for $z \in \mathcal{B}$. We also know by Corollary 2.6.7 that σ is a law if and only if $\sigma|_{\mathcal{B}} = \text{id}$. This implies that (1) $\Leftrightarrow$ (2).

(1) $\implies$ (3). If σ is a law, then

$$z^{-1}\tilde{G}_\sigma(z) = z^{-1}\sigma^{(n)}[z(1 - X^{(n)}z)^{-1}] = \sigma^{(n)}[(1 - X^{(n)}z)^{-1}],$$

which is fully matricial in a neighborhood of zero, and hence (3) holds.

(3) $\implies$ (1). Fix an invertible operator $z \in \mathcal{B}$. Then we have for $\zeta \in \mathbb{C}$ that

$$\lim_{\zeta\to 0} \frac{1}{\zeta}\tilde{G}_\sigma(\zeta z) = z \lim_{\zeta\to 0}(\zeta z)^{-1}\tilde{G}_\sigma(\zeta z) = z.$$

On the other hand,

$$\lim_{\zeta\to 0} \frac{1}{\zeta}\tilde{G}_\sigma(\zeta z) = \lim_{\zeta\to 0}\sum_{k=0}^{\infty} \zeta^k \sigma[z(Xz)^k] = \sigma[z].$$

Therefore, $\sigma[z] = z$. Any element of $\mathcal{B}$ can be written as a linear combination of invertible operators and hence $\sigma|_{\mathcal{B}} = \text{id}$, which means that σ is a law. □

We also have the following corollary of Theorem 4.4.1 which is helpful for estimating the radius of generalized laws.

Corollary 4.4.8. *Suppose that σ and τ are $\mathcal{B}$-valued generalized laws and $\operatorname{Im} G_\sigma(z) \geq \operatorname{Im} G_\tau(z)$. Then*

(1) $G_\tau(z) - G_\sigma(z)$ is the Cauchy-Stieltjes transform of some generalized law ρ.

(2) $\operatorname{rad}(\sigma) \leq \operatorname{rad}(\tau)$.

(3) For $\operatorname{Im} z \geq \epsilon$, we have $\|G_\sigma(z) - G_\tau(z)\| \leq \|\sigma(1) - \tau(1)\|/\epsilon$.

Proof. (1) Observe that $G_\tau - G_\sigma$ maps $\mathbb{H}_+(\mathcal{B})$ into $\overline{\mathbb{H}}_-(\mathcal{B})$. Moreover, $\tilde{G}_\tau - \tilde{G}_\sigma$ extends to be fully matricial in a neighborhood of 0 in a way which preserves adjoints. Therefore, there is a generalized law ρ such that $G_\tau - G_\sigma = G_\rho$. Now $\rho = \tau - \sigma$ as maps $\mathcal{B}\langle X\rangle \to \mathcal{B}$.

(2) This follows from Lemma 2.6.8 since $\tau = \sigma + \rho$.

(3) This follows by applying the estimate for Lemma 4.2.7 to $G_\rho(z)$. □

4.5 The F-transform

Definition 4.5.1. Let μ be a $\mathcal{B}$-valued law. We define the *F-transform*

$$F_\mu(z) = G_\mu(z)^{-1}.$$

Lemma 4.5.2. *F_μ is a fully matricial function $\mathbb{H}_+(\mathcal{B}) \to \mathbb{H}_+(\mathcal{B})$.*

Proof. Suppose that $z \in \mathbb{H}_+(\mathcal{B})$. If we have $\operatorname{Im} z \geq \epsilon$, then by Lemma 4.2.3, we have

$$\operatorname{Im}(z - X^{(n)})^{-1} \leq \frac{-\epsilon}{\|z - X^{(n)}\|^2}.$$

By complete positivity of μ and the fact that $\mu(1) = 1$, we have

$$\operatorname{Im} G_\mu(z) \leq \frac{-\epsilon}{\|z - X^{(n)}\|^2}.$$

This implies that $G_\mu(z) \in \mathbb{H}_-(\mathcal{B})$ and in particular $G_\mu(z)$ is invertible. Moreover, one checks from Lemma 3.7.2 that $\operatorname{inv} : \mathbb{H}_-(\mathcal{B}) \to \mathbb{H}_+(\mathcal{B})$ is fully matricial, and hence $F_\mu(z)$ is fully matricial. $\square$

4.5.1 A characterization of F-transforms

The following characterization of F-transforms will be useful in the later chapters for understanding the analytic transforms associated to non-commutative independence. A related characterization of $z - F_\mu(z)$ as the self-energy of some law was given in [PV13, Theorem 5.6], and [Wil17, Corollary 3.3], while the statement that $z - F_\mu(z)$ is the Cauchy-Stieltjes transform of a generalized law was proved in [PV13, Remark 5.7]. Compare also [BV93, Proposition 5.2], [SW97, Proposition 3.1] (scalar case), [ABF13, §7], [Jek20, Proposition 3.9], and [JL19, Lemma 9.4].

Theorem 4.5.3. *There is a bijection $\mu \leftrightarrow (b, \sigma)$ between $\Sigma(\mathcal{B})$ and $\mathcal{B}_{sa} \times \Sigma^{\text{gen}}(\mathcal{B})$ given by the relation*

$$F_\mu^{(n)}(z) = z - b^{(n)} - G_\sigma^{(n)}(z) \text{ for all } n. \tag{4.2}$$

Furthermore, we have $b = \mu(X)$ and $\sigma(z) = \mu(XzX) - \mu(X)z\mu(X)$ for $z \in \mathcal{B}$, and

$$\max(\|b\|, \operatorname{rad}(\sigma)) \leq \operatorname{rad}(\mu) \leq \max(\|b\|, \operatorname{rad}(\sigma)) + \|\sigma(1)\|^{1/2}. \tag{4.3}$$

Proof. First, observe that for each (b, σ), there is at most one μ that satisfies the relation, since μ is uniquely determined by F_μ. Conversely, for each μ, there is at most one choice of (b, σ). Indeed, if $b^{(n)} + G_\sigma^{(n)}(z) = (b')^{(n)} + G_{\sigma'}^{(n)}(z)$, then we recover $b = b'$ by evaluating this function at z^{-1} and then taking $z \to 0$. Then since $G_\sigma = G_{\sigma'}$, we have also $\sigma = \sigma'$.

Next, we show that for each μ, there exists (b, σ) satisfying the given relation. Let $(\mathcal{H}, \xi)$ be a $\mathcal{B}$-$\mathcal{B}$-correspondence with $\mathcal{B}$-central unit vector and let X an operator on $\mathcal{H}$ with $\mu_X = \mu$ and $\|X\| = \mathrm{rad}(\mu)$. Let P be the projection onto $\mathcal{B}\xi$ and $Q = 1 - P$. We claim that the relation $F_\mu^{(n)}(z) = z - b^{(n)} - G_\sigma^{(n)}(z)$ is satisfied with $b = \mu(X)$ and σ equal to the distribution of QXQ with respect to the vector $QX\xi$. We start by using the resolvent identity; for $z \in M_n(\mathcal{B})$ with $\operatorname{Im} z \geq \epsilon > 0$,

$$\begin{aligned}\left(z - X^{(n)}\right)^{-1} &= \left(z - P^{(n)}X^{(n)} - Q^{(n)}X^{(n)}\right)^{-1} \\ &= \left(z - Q^{(n)}X^{(n)}\right)^{-1} + \left(z - P^{(n)}X^{(n)} - Q^{(n)}X^{(n)}\right)^{-1} P^{(n)}X^{(n)} \left(z - Q^{(n)}X^{(n)}\right)^{-1} \\ &= \left(z - Q^{(n)}X^{(n)}\right)^{-1} + \left(z - X^{(n)}\right)^{-1} P^{(n)}X^{(n)} \left(z - Q^{(n)}X^{(n)}\right)^{-1}.\end{aligned}$$

Let $\xi^{(n)} = \xi \otimes I_n$ in $\mathcal{H}^{(n)} = M_n(\mathcal{H})$. Clearly, $P^{(n)}$ is the projection onto $M_n(\mathcal{B})\xi^{(n)}$. Then we have

$$\begin{aligned}&\left\langle \xi^{(n)}, \left(z - X^{(n)}\right)^{-1} \xi^{(n)} \right\rangle \\ &= \left\langle \xi^{(n)}, \left(z - Q^{(n)}X^{(n)}\right)^{-1} \xi^{(n)} \right\rangle + \left\langle \xi^{(n)}, \left(z - X^{(n)}\right)^{-1} \xi^{(n)} \right\rangle \left\langle \xi^{(n)}, X^{(n)} \left(z - Q^{(n)}X^{(n)}\right)^{-1} \xi^{(n)} \right\rangle,\end{aligned} \tag{4.4}$$

where we have used the relation $\langle \xi^{(n)}, SP^{(n)}T\xi^{(n)} \rangle = \langle \xi^{(n)}, S\xi^{(n)} \rangle \langle \xi^{(n)}, T\xi^{(n)} \rangle$ for $S, T \in B(\mathcal{H}^{(n)})$. Regarding the first term on the right hand side, we write

$$\left\langle \xi^{(n)}, \left(z - Q^{(n)}X^{(n)}\right)^{-1} \xi^{(n)} \right\rangle = \left\langle \left(z^* - X^{(n)}Q^{(n)}\right)^{-1} \xi^{(n)}, \xi^{(n)} \right\rangle,$$

but $(z^* - X^{(n)}Q^{(n)})(z^*)^{-1}\xi^{(n)} = z^*(z^*)^{-1}\xi^{(n)} = \xi^{(n)}$, and hence $(z^* - X^{(n)}Q^{(n)})^{-1}\xi^{(n)} = (z^*)^{-1}\xi^{(n)}$. Thus,

$$\left\langle \xi^{(n)}, \left(z - Q^{(n)}X^{(n)}\right)^{-1} \xi^{(n)} \right\rangle = \langle (z^*)^{-1}\xi^{(n)}, \xi^{(n)} \rangle = z^{-1}.$$

Therefore, (4.4) becomes

$$G_\mu^{(n)}(z) = z^{-1} + G_\mu^{(n)}(z) \left\langle \xi^{(n)}, X^{(n)} \left(z - Q^{(n)}X^{(n)}\right)^{-1} \xi^{(n)} \right\rangle. \tag{4.5}$$

Now we must evaluate $\langle \xi^{(n)}, X^{(n)} \left(z - Q^{(n)}X^{(n)}\right)^{-1} \xi^{(n)} \rangle$. By the resolvent identity,

$$\left(z - Q^{(n)}X^{(n)}\right)^{-1} = z^{-1} + \left(z - Q^{(n)}X^{(n)}\right)^{-1} Q^{(n)}X^{(n)} z^{-1}. \tag{4.6}$$

But we claim that

$$\left(z - Q^{(n)}X^{(n)}\right)^{-1} Q^{(n)} = Q^{(n)} \left(z - Q^{(n)}X^{(n)}Q^{(n)}\right)^{-1} Q^{(n)}. \tag{4.7}$$

To see this, fix a vector ζ and let $\eta = \left(z - Q^{(n)}X^{(n)}\right)^{-1} Q^{(n)}\zeta$. Then

$$z\eta - Q^{(n)}X^{(n)}\eta = Q^{(n)}\zeta,$$

and so $\eta = z^{-1}Q^{(n)}[X^{(n)}\eta + \zeta]$, which shows that η is in the image of $Q^{(n)}$. Thus,

$$\left(z - Q^{(n)}X^{(n)}Q^{(n)}\right)\eta = Q^{(n)}\zeta$$

and so

$$\eta = Q^{(n)}\eta = Q^{(n)}\left(z - Q^{(n)}X^{(n)}Q^{(n)}\right)^{-1}Q^{(n)}\zeta.$$

Now (4.6) and (4.7) imply that

$$\begin{aligned}
&\left\langle \xi^{(n)}, X^{(n)}\left(z - Q^{(n)}X^{(n)}\right)^{-1}\xi^{(n)}\right\rangle \\
&= \langle \xi^{(n)}, X^{(n)}z^{-1}\xi^{(n)}\rangle + \left\langle \xi^{(n)}, X^{(n)}Q^{(n)}\left(z - Q^{(n)}X^{(n)}\right)^{-1}Q^{(n)}X^{(n)}z^{-1}\xi^{(n)}\right\rangle \\
&= \mu(X)^{(n)}z^{-1} + \left\langle Q^{(n)}X^{(n)}\xi^{(n)}, \left(z - Q^{(n)}X^{(n)}Q^{(n)}\right)^{-1}Q^{(n)}X^{(n)}\xi^{(n)}\right\rangle z^{-1} \\
&= \left(b^{(n)} + G_\sigma^{(n)}(z)\right)z^{-1}.
\end{aligned}$$

Plugging this into (4.5), we get

$$G_\mu^{(n)}(z) = z^{-1} + G_\mu^{(n)}(z)\left(b^{(n)} + G_\sigma^{(n)}(z)\right)z^{-1}.$$

Then we multiply by $F_\mu^{(n)}(z)$ on the left and z on the right to get

$$z = F_\mu^{(n)}(z) + b^{(n)} + G_\sigma^{(n)}(z)$$

as desired. Clearly, $\|b\| = \|\mu(X)\| \leq \operatorname{rad}(\mu)$, and because σ is the law of QXQ, we have $\operatorname{rad}(\sigma) \leq \|QXQ\| \leq \|X\| = \operatorname{rad}(\mu)$. So we obtain the left inequality of (4.3).

Finally, we show that for each (b, σ), there is a μ satisfying the given relation. By the GNS construction, there is a $\mathcal{B}$-$\mathcal{B}$ correspondence $\mathcal{K}$, a self-adjoint operator Y on $\mathcal{K}$, and a vector ζ such that σ is the distribution of Y with respect to ζ. Let $\mathcal{H}$ be the $\mathcal{B}$-$\mathcal{B}$-correspondence $\mathcal{B} \oplus \mathcal{K}$. Define $X \in B(\mathcal{H})$ by

$$X(c \oplus \eta) = (bc + \langle \zeta, \eta\rangle) \oplus (\zeta c + Y\eta) \text{ for } c \in \mathcal{B}, \eta \in \mathcal{K}.$$

Let ξ be the $\mathcal{B}$-central unit vector $1 \oplus 0$ in $\mathcal{H}$. Let μ be the law of X with respect to ξ. If $P \in B(\mathcal{H})$ is the projection onto $\mathcal{B}\xi$ and Q is the projection onto $\mathcal{K}$, then QXQ with the domain and codomain restricted to $\mathcal{K}$ is exactly Y. Also, $\mu(X) = b$ and $QX\xi = \zeta$. Therefore, we are in the same situation as the previous argument (minus the assumption that $\|X\| = \operatorname{rad}(\mu)$), and thus (4.2) holds. To check the right inequality of (4.3), we write $X = PXP + QXQ + QXP + PXQ$, hence

$$\operatorname{rad}(\mu) \leq \|X\| \leq \|PXP + QXQ\| + \|PXQ + QXP\|.$$

Now $PXP + QXQ$ is a "block diagonal" operator composed of the blocks $PXP = bP$ and QXQ, which is essentially Y, and thus $\|PXP + QXQ\| \leq \max(\|b\|, \operatorname{rad}(\sigma))$. Meanwhile, using $PQ = QP = 0$ and using the C* identity several times, we get

$$\begin{aligned} \|PXQ + QXP\|^2 &= \|(PXQ + QXP)^*(PXQ + QXP)\| \\ &= \|PXQXP + QXPXQ\| \\ &= \max\left(\|PXQXP\|, \|QXPXQ\|\right) \\ &= \max\left(\|(QXP)^*(QXP)\|, \|(QXP)(QXP)^*\|\right) \\ &= \|QXP\|^2 \\ &= \|PXQXP\|. \end{aligned}$$

Now a direct computation shows that $PXQXP = \langle \zeta, \zeta \rangle P = \sigma(1)P$. Hence, $\|PXQ + QXP\| = \|\sigma(1)\|^{1/2}$, which combines with our previous estimates to show the right inequality of (4.3). □

Remark 4.5.4. We also have the estimate

$$\mathrm{rad}(\mu) \leq \left(\|b\|^2 + \mathrm{rad}(\sigma)^2 + 2\|\sigma(1)\|\right)^{1/2}.$$

To see this, start with Y and X as in the last part of the proof. Since PX and QX have orthogonal ranges, we have

$$\|PX + QX\|^2 \leq \|PX\|^2 + \|QX\|^2.$$

But then also

$$\|PX\|^2 = \|XP\|^2 = \|PXP + QXP\|^2 \leq \|PXP\|^2 + \|QXP\|^2,$$

and the same applies to QX. Therefore,

$$\|X\|^2 \leq \|PXP\|^2 + \|QXQ\|^2 + \|PXQ\|^2 + \|QXP\|^2 = \|b\|^2 + \mathrm{rad}(\sigma)^2 + 2\|\sigma(1)\|.$$

Remark 4.5.5. The theorem (perhaps with less sharp estimates on the radius) can alternatively be deduced by studying the analytic properties of F_μ and G_σ and invoking Theorem 4.4.1, as we suggested in [Jek20, Proposition 3.30]. Finding this alternative proof is a good exercise.

The object $\mathrm{Var}_\mu(a) := \mu(XaX) - \mu(X)a\mu(X)$ will be significant in the coming chapters. As a consequence of what we have just shown, Var_μ is a completely positive $\mathcal{B} \to \mathcal{B}$. This Var_μ is related to the variance in classical probability theory. Indeed, if $\mathcal{B} = \mathbb{C}$ and μ is a measure on the real line, then Var_μ is a map $\mathbb{C} \to \mathbb{C}$ which is simply multiplication by a positive scalar, and this positive scalar is the classical variance of μ. Thus, we will call Var_μ the *operator-valued variance of* μ.

Remark 4.5.6. The F-transform is related to theory of conformal mapping in complex analysis in the following way. Suppose that Ω is a connected region in the upper half-plane $\mathbb{H}$ such that $\mathbb{H} \setminus \Omega$ is bounded. Then there is a unique conformal map $F : \mathbb{H} \to \Omega$ satisfying

$F(z) = z + O(1/z)$. The coefficient t such that $F(z) = z + t/z + O(1/z^2)$ is known as the *half-plane capacity* of Ω. Now by Lemma 4.1.1, one can show that $F = F_\mu$ for some compactly supported measure μ. The normalization that $F(z) = z + O(1/z)$ then amounts to the mean of μ being zero, and the half-plane capacity of Ω is nothing but the variance of μ, which is also equal to the total mass of the measure σ that corresponds to μ as in Theorem 4.5.3.

4.5.2 Inversion of F-transforms

Our next result concerns the inverse of the F-transform, which will be relevant to the discussion of free convolution in §5.5 and §7.4.

Lemma 4.5.7. *Let μ be a $\mathcal{B}$-valued law. For each $\delta > \|\mathrm{Var}_\mu(1)\|^{1/2}$, the function $F_\mu^{(n)}$ is injective on $\mathbb{H}_{+,\delta}^{(n)}(\mathcal{B})$. The inverse function is fully matricial and sends $\mathbb{H}_{+,2\delta}^{(n)}(\mathcal{B})$ into $\mathbb{H}_{+,\delta}^{(n)}(\mathcal{B})$.*

Proof. By Theorem 4.5.3, there exists a self-adjoint b and a generalized law σ with $\mathrm{rad}(\sigma) \leq 2\,\mathrm{rad}(\mu)$ such that

$$F_\mu(z) = z - b - G_\sigma(z).$$

Let $\delta > \|\sigma(1)\|^{1/2} = \|\mathrm{Var}_\mu(1)\|^{1/2}$. Then we claim that F_μ has an inverse function $\Psi : \mathbb{H}_{+,2\delta}(\mathcal{B}) \to \mathbb{H}_{+,\delta}(\mathcal{B})$. We will construct Ψ by a contraction mapping principle just as in the inverse function theorem. We want to solve the equation

$$w = \Psi(w) - b - G_\sigma(\Psi(w)),$$

so that $\Psi(w)$ satisfies the fixed point equation

$$\Psi^{(n)}(w) = w + b^{(n)} + G_\sigma^{(n)}(\Psi(w)).$$

Let $H_w(z) = w + b + G_\sigma(z)$. Note that by Lemma 4.3.1, if $z, z' \in \mathbb{H}_{+,\delta}^{(n)}(\mathcal{B})$, then

$$\|H_w(z) - H_w(z')\| = \|G_\sigma(z) - G_\sigma(z')\| \leq \frac{\|\sigma(1)\|}{\delta^2}\|z - z'\|.$$

Therefore, H_w is a contraction provided that $\delta > \|\sigma(1)\|^{1/2}$. Moreover, if $\mathrm{Im}\, w \geq 2\delta$, then H_w maps $\mathbb{H}_{+,\delta}(\mathcal{B})$ into itself because

$$\mathrm{Im}\, H_w(z) = \mathrm{Im}\, w + \mathrm{Im}\, G_\sigma(z) \geq 2\delta - \frac{\|\sigma(1)\|}{\delta} \geq \delta.$$

Therefore, by the Banach fixed point theorem, H_w has a unique fixed point $\Psi(w)$ in $\mathbb{H}_{+,\delta}(\mathcal{B})$. We also have

$$\|\Psi(w) - w\| = \|b^{(n)} + G_\sigma(\Psi(w))\| \leq \|b\| + \frac{\|\sigma(1)\|}{\delta} \leq \|b\| + \delta.$$

Therefore, if we define

$$\Psi_0(w) = w, \qquad \Psi_{k+1}(w) = H_w(\Psi_k(w)),$$

then for $\operatorname{Im} w \geq 2\delta$,

$$\|\Psi_k(w) - \Psi(w)\| \leq \left(1 - \frac{\|\sigma(1)\|}{\delta^2}\right)^k (\|b\| + \delta).$$

In particular, Ψ_k converges uniformly locally to Ψ on $\bigcup_{\delta > \|\sigma(1)\|^{1/2}} \mathbb{H}_{+,2\delta}(\mathcal{B})$. It follows that $\Psi(w)$ is fully matricial. □

Lemma 4.5.8. *If μ is a $\mathcal{B}$-valued law, then $\Phi_\mu(z) := F_\mu^{-1}(z) - z$ is a well-defined fully matricial function $\bigcup_{\delta > \|\operatorname{Var}_\mu[1]\|^{1/2}} \mathbb{H}_{2\delta,+}(\mathcal{B}) \to \overline{\mathbb{H}}_-(\mathcal{B})$. If we write $F_\mu(z) = z - b - G_\sigma(z)$ for a generalized law σ and $b \in \mathcal{B}$ self-adjoint and if $\delta > \|\operatorname{Var}_\mu[1]\| = \|\sigma(1)\|$, then*

$$\|\Phi_\mu(z)\| \leq \|b\| + \frac{\|\sigma(1)\|}{\delta} \text{ for } \operatorname{Im} z \geq 2\delta \tag{4.8}$$

and

$$\|\Phi_\mu(z) - b - G_\sigma(z)\| \leq \frac{\|\sigma(1)\|}{2\delta^2}\left(\|b\| + \frac{\|\sigma(1)\|}{\delta}\right) \text{ for } \operatorname{Im} z \geq 2\delta. \tag{4.9}$$

Proof. In the notation of the previous proof, we have

$$\Phi_\mu(z) = \Psi(z) - z = b + G_\sigma(\Psi(z)).$$

Therefore, Φ_μ is a fully matricial function $\bigcup_{\delta > \|\sigma(1)\|^{1/2}} \mathbb{H}_{+,2\delta}(\mathcal{B}) \to \overline{\mathbb{H}}_-(\mathcal{B})$.

To check (4.8), use the identity $\Phi_\mu(z) = b + G_\sigma(\Psi(z))$. When $\operatorname{Im} z \geq 2\delta$ and $\delta > \|\sigma(1)\|^{1/2}$, then we have $\operatorname{Im} \Psi(z) \geq \delta$. Hence, by applying (4.1) to G_σ, we get (4.8)

To check (4.9), observe that

$$\Phi_\mu(z) - b - G_\sigma(z) = G_\sigma(\Psi(z)) - G_\sigma(z).$$

But by Lemma 4.3.1,

$$\begin{aligned} \|G_\sigma(\Psi(z)) - G_\sigma(z)\| &\leq \|\Delta G_\sigma(\Psi(z), z)\|_\# \|\Psi(z) - z\| \\ &\leq \frac{\|\sigma(1)\|}{\delta \cdot 2\delta}\left(\|b\| + \frac{\|\sigma(1)\|}{\delta}\right), \end{aligned}$$

where in the last line we substituted $\Psi(z) - z = \Phi_\mu(z)$ and used the previous estimate (4.8). □

Lemma 4.5.9. *The function* $\tilde{\Phi}_\mu(z)$ *has a fully matricial extension to* $B(0, (3-2\sqrt{2})/\operatorname{rad}(\mu))$ *satisfying*

$$\tilde{\Phi}_\mu(0) = \mu(X), \qquad \tilde{\Phi}_\mu(z^*) = \tilde{\Phi}_\mu(z)^*$$

and

$$\|\tilde{\Phi}_\mu(z) - \mu(X)^{(n)}\| \leq \frac{2\|\operatorname{Var}_\mu(1)\| \operatorname{rad}(\mu)}{\sqrt{2}-1}.$$

Proof. Consider the behavior of $\tilde{\Phi}_\mu = \tilde{\Phi}_\mu$ near zero. We have

$$\begin{aligned}\tilde{\Phi}_\mu(z) &= b^{(n)} + G_\sigma(\Psi(z^{-1})) \\ &= b^{(n)} + G_\sigma(F_\mu^{-1}(z^{-1})) \\ &= b^{(n)} + \tilde{G}_\sigma(\tilde{G}_\mu^{-1}(z)).\end{aligned}$$

By the inverse function theorem, since $D\tilde{G}_\mu(0) = \operatorname{id}$, we know that $\tilde{G}_\mu$ has a inverse function in a neighborhood of zero, and hence $\tilde{\Phi}_\mu$ is defined in a neighborhood of 0 and $\tilde{\Phi}_\mu(0) = b = \mu(X)$.

To get a more precise estimate on the size of the neighborhood, observe that for $R = 1/\operatorname{rad}(\mu)$, we have

$$\|\Delta^k \tilde{G}_\mu(0, \dots, 0)\|_\# \leq \frac{1}{R^{k-1}},$$

and therefore, we are in the setting of the inverse function theorem with $M = R$ and $K = 1$. Thus, by Theorem 3.8.1, $\tilde{G}_\mu^{-1}$ maps $B(0, R(3-2\sqrt{2})) \to B(0, R(1-1/\sqrt{2}))$. But note that

$$\left(1 - \frac{1}{\sqrt{2}}\right) R < \frac{1}{2} R = \frac{1}{2\operatorname{rad}(\mu)} \leq \frac{1}{\operatorname{rad}(\sigma)},$$

and hence $B(0, R(1-1/\sqrt{2}))$ is within the ball where $\tilde{G}_\sigma$ is defined, so that $\tilde{\Phi}_\mu = \tilde{G}_\sigma \circ \tilde{G}_\mu + b^{(n)}$ is defined on $B(0, R(3-2\sqrt{2}))$. Futhermore, $\tilde{G}_\sigma$ is bounded by

$$\frac{\|\sigma(1)\|}{(1/2)R - (1-1/\sqrt{2})R} = \frac{\|\sigma(1)\| \operatorname{rad}(\mu)}{1/2 + 1/\sqrt{2} - 1} = \frac{2\|\operatorname{Var}_\mu(1)\| \operatorname{rad}(\mu)}{\sqrt{2}-1}.$$ □

4.6 Convergence in moments

Definition 4.6.1. If σ is a generalized law, then we define the kth moment of σ as the multilinear form

$$\operatorname{Mom}_k(\sigma)[w_0, \dots, w_k] = \sigma(w_0 X w_1 \dots X w_k)$$

or equivalently

$$\operatorname{Mom}_k(\sigma) = \Delta^{k+1} \tilde{G}_\sigma(0, \dots, 0).$$

Definition 4.6.2. Let σ_n and σ be generalized laws. We say that $\sigma_n \to \sigma$ *in moments* if

$$\lim_{n\to\infty} \|\mathrm{Mom}_k(\sigma_n) - \mathrm{Mom}_k(\sigma)\|_{\#} = 0 \text{ for every } k,$$

where $\|\cdot\|_{\#}$ is the completely bounded norm for multilinear forms. Similarly, we say that $\{\sigma_n\}$ is *Cauchy in moments* if $\{\mathrm{Mom}_k(\sigma_n)\}$ is Cauchy for each k.

Definition 4.6.3. We denote by $\Sigma_M(\mathcal{B})$ the set of $\mathcal{B}$-valued laws with $\mathrm{rad}(\mu) \leq M$. We denote by $\Sigma_{M,K}^{\mathrm{gen}}(\mathcal{B})$ the set of $\mathcal{B}$-valued generalized laws σ with $\mathrm{rad}(\sigma) \leq M$ and $\|\sigma(1)\| \leq K$.

Lemma 4.6.4. *If $\{\sigma_n\}$ in $\Sigma_{M,K}^{\mathrm{gen}}(\mathcal{B})$ is Cauchy in moments, then it converges in moments. Also, $\Sigma_M(\mathcal{B})$ is a closed subset of $\Sigma_{M,1}^{\mathrm{gen}}(\mathcal{B})$ with respect to the convergence in moments.*

Proof. Clearly, the multilinear forms $\mathrm{Mom}_k(\sigma_n)$ converge to some multilinear form Λ_k. We can define $\sigma : \mathcal{B}\langle X\rangle \to \mathcal{B}$ by $\sigma(w_0 X w_1 \dots X w_k) = \Lambda_k(w_0, \dots, w_k)$. Then $\sigma_n[f(X)] \to \sigma[f(X)]$ for each $f(X) \in \mathcal{B}\langle X\rangle$ and hence σ is completely positive and exponentially bounded by M. Also, $\|\sigma(1)\| \leq K$ since $\|\sigma_n(1)\| \leq K$. Therefore, σ is a generalized law in $\Sigma_{M,K}^{\mathrm{gen}}(\mathcal{B})$ and $\sigma_n \to \sigma$ in moments.

To show that $\Sigma_M(\mathcal{B})$ is closed, note that the property of $\sigma : \mathcal{B}\langle X\rangle \to \mathcal{B}$ being a unital $\mathcal{B}$-$\mathcal{B}$-bimodule map is preserved under limits. □

Proposition 4.6.5. *If $r < 1/M$ and $\sigma, \tau \in \Sigma_{M,K}^{\mathrm{gen}}(\mathcal{B})$, define*

$$d_r(\sigma,\tau) = \sum_{k=0}^{\infty} r^{k+1} \|\mathrm{Mom}_k(\sigma) - \mathrm{Mom}_k(\tau)\|_{\#} = d_{0,r}(\tilde{G}_\sigma, \tilde{G}_\tau).$$

Then we have the following.

(1) d_r is a metric.

(2) The metrics d_r for different values of r are uniformly equivalent.

(3) $\{\sigma_n\} \subseteq \Sigma_{M,K}^{\mathrm{gen}}$ is convergent (resp. Cauchy) in moments if and only if it convergent (resp. Cauchy) in d_r.

(4) $\Sigma_{M,K}^{\mathrm{gen}}(\mathcal{B})$ is complete with respect to d_r.

Proof. Observe that $d_r(\sigma,\tau) = d_{0,r}(\tilde{G}_\sigma, \tilde{G}_\tau)$. It follows from Lemma 4.3.5 that $\mathcal{F} = \{\tilde{G}_\sigma : \sigma \in \Sigma_{M,K}^{\mathrm{gen}}(\mathcal{B})\}$ is a uniformly locally bounded family of fully matricial functions on $B(0, 1/M)$ and that $\mathrm{rad}(0, \mathcal{F}) = 1/M$. Therefore, claims (1) and (2) follow from Theorem 3.9.6.

(3) Note that

$$\|\mathrm{Mom}_k(\sigma) - \mathrm{Mom}_k(\tau)\|_{\#} \leq \frac{1}{r^k} d_r(\sigma,\tau).$$

Hence, convergence or Cauchyness in d_r implies convergence or Cauchyness in moments. Conversely, using standard geometric series estimates,

$$d_r(\sigma, \tau) \leq \sum_{k=0}^{N-1} r^k \|\mathrm{Mom}_k(\sigma) - \mathrm{Mom}_k(\tau)\|_\# + \frac{(rM)^N}{1 - rM}$$

and hence convergence or Cauchyness in moments implies convergence or Cauchyness in d_r.

(4) This follows from (3) and Lemma 4.6.4. □

Proposition 4.6.6.

(1) The collection $\mathcal{G}_{M,K} = \{G_\sigma : \sigma \in \Sigma_{M,K}^{\mathrm{gen}}(\mathcal{B})\}$ is a uniformly locally bounded family of fully matricial functions on $\mathbb{H}_+(\mathcal{B})$.

(2) For each $z \in \mathbb{H}_{+,\epsilon}(\mathcal{B})$ we have $\mathrm{rad}(z, \mathcal{G}) \geq \epsilon$.

(3) The metrics $d_{z,r}(G_\sigma, G_\tau)$ on $\mathcal{G}$ are uniformly equivalent to the metrics $d_r(\sigma, \tau)$.

(4) $\mathcal{G}_{M,K}$ with the topology of uniform local convergence is homeomorphic to $\Sigma_{M,K}^{\mathrm{gen}}$ with the topology of convergence in moments.

Proof. (1) and (2) follow from Lemma 4.2.7.

To prove (3), note that the metrics $d_{z,r}$ are all equivalent to each other by Theorem 3.9.6. Moreover, note that all the elements of $B(3iM, M)$ are invertible and $\mathrm{inv}(B(3iM, M)) \subseteq B(0, 1/2M)$, so that

$$d_{3iM,M}(G_\sigma, G_\tau) \leq d_{0,1/2M}(\tilde{G}_\sigma, \tilde{G}_\tau) = d_{1/2M}(\sigma, \tau).$$

Thus, $d_{3iM,M}(G_\sigma, G_\tau)$ can be estimated above by $d_{1/2M}(\sigma, \tau)$. For the converse direction, note that $\mathrm{inv}(B(1/2iM, 1/8M)) \subseteq B(2iM, 2M/3)$ and hence

$$d_{1/2iM,1/8M}(\tilde{G}_\sigma, \tilde{G}_\tau) \leq d_{2iM,2M/3}(G_\sigma, G_\tau).$$

By Theorem 3.9.6, the metric $d_{1/2iM,1/8M}$ is equivalent to $d_r(\sigma, \tau)$, and thus $d_r(\sigma, \tau)$ can be estimated from above by $d_{2iM,2M/3}(G_\sigma, G_\tau)$.

(4) is an immediate consequence of (3). □

CHAPTER 5

Background: Four non-commutative independences

5.1 Introduction

The discovery of free independence is due to Voiculescu [Voi85, Voi86] (and the combinatorial aspects of the theory were developed by Speicher [Spe94]). His key insight was that the free product operation on groups and the corresponding operator algebras could be viewed as a non-commutative version of probabilistic independence. The analogy between the classical and free theories included the following elements:

1. *Rule for specifying mixed moments:* To say that algebras $\mathcal{A}_1, \dots, \mathcal{A}_n$ are freely independent specifies rule for determining the mixed moments of variables in the larger algebra which they generate.

2. *Product space construction:* Any two algebras could be joined in an independent way. In ordinary probability theory, this is the role of the product measure spaces, corresponding to a tensor product of the L^2 spaces, on which the two individual algebras act by multiplication on first and second coordinate. In free probability theory, products of algebras act on the free product of the underlying Hilbert spaces, a construction related to Fock spaces in physics.

3. *Convolution operation and analytic transforms:* In ordinary probability theory, the law of a sum of independent random variables is the convolution of the two individual laws, and the convolution can be computed using the Fourier transform of the measure. Given (1), the law of the sum of independent random variables is determined by the individual laws, and so "free convolution" is well-defined. Voiculescu found that the R-transform played a similar role in free probability theory; namely, the R transform of the "free convolution" of two laws is the sum of the R-transforms.

This theory was extended to the operator-valued setting in [Voi85, §5] as well as [Voi95] and [Spe98].

Another type of non-commutative independence, called boolean independence, was introduced into non-commutative probability by Speicher and Woroudi [SW97], based on previous work by physicists. This independence had a rule for specifying mixed moments, a product space construction, and a convolution operation. For operator-valued Boolean independence, see [Pop09], [PV13, §2], [BPV13].

Finally, monotone independence was discovered by Muraki [Mur97], [Mur00], [Mur01], and adapted to the operator-valued setting by Popa [Pop08] and Hasebe and Saigo [HS14]. There was a parallel theory of moment computations, product spaces, and analytic transforms. Unlike free and boolean independence, monotone independence is sensitive to the order of algebras. Thus, the monotone convolution operation is not commutative. It corresponds to composition rather than addition of analytic transforms.

After the discovery of several types of independence, Speicher formulated axioms for independences which lead to a natural commutative binary product operation, and he showed that tensor, free, and boolean were the only three possilibities [Spe97]; Ben Ghorbal and Schürmann proved related results in the framework of category theory [BS02]. When the product is no longer required to be commutative, there are exactly two more possibilities, monotone independence and its mirror image anti-monotone independence, as proved by Muraki in 2003 [Mur03]. This in some sense classified the possible notions of independence. The analogous results in the operator-valued setting have not yet been studied.

Here we will focus on operator-valued free, boolean, monotone, and anti-monotone independence. We exclude classical or tensor independence because it does not adapt well to the $\mathcal{B}$-valued setting if $\mathcal{B}$ is not commutative, and because the other types of independence have closer similarities with each other. As much as possible, we will present theories of these four types in parallel.

5.2 Moment conditions for independence

Definition 5.2.1. Let $\mathcal{A} \supseteq \mathcal{B}$ be C*-algebras. We say that $\mathcal{C}$ is a (non-unital) $\mathcal{B}$-$*$-subalgebra of $\mathcal{A}$ if $\mathcal{A}$ if $\mathcal{C}$ is closed under addition, multiplication, and adjoints, and if $\mathcal{B}\mathcal{C} \subseteq \mathcal{A}$.

Definition 5.2.2 (Boolean independence)**.** Let $(\mathcal{A}, E)$ be a $\mathcal{B}$-valued probability space. Then $\mathcal{B}$-subalgebras $\mathcal{A}_1$, ..., $\mathcal{A}_N$ are said to be *boolean independent* if we have

$$E[a_1 \dots a_k] = E[a_1] \dots E[a_k]$$

whenever $a_j \in \mathcal{A}_{i(j)}$, provided that the consecutive indices $i(j)$ and $i(j+1)$ are distinct.

Definition 5.2.3 (Free independence)**.** Let $(\mathcal{A}, E)$ be a $\mathcal{B}$-valued probability space. Then subalgebras $\mathcal{A}_1$, ..., $\mathcal{A}_N$ containing $\mathcal{B}$ are said to be *freely independent* if we have

$$E[a_1 \dots a_k] = 0$$

whenever $a_j \in \mathcal{A}_{i(j)}$ with $E[a_j] = 0$, provided that the consecutive indices $i(j)$ and $i(j+1)$ are distinct.

Definition 5.2.4 (Monotone independence)**.** Let $(\mathcal{A}, E)$ be a $\mathcal{B}$-valued probability space. Then $\mathcal{B}$-subalgebras $\mathcal{A}_1$, ..., $\mathcal{A}_N$ are said to be *monotone independent* if we have

$$E[a_1 \dots a_k] = E[a_1 \dots a_{r-1} E[a_r] a_{r+1} \dots a_k]$$

whenever $a_j \in \mathcal{A}_{i(j)}$, provided that the index $i(r)$ is strictly greater than the consecutive indices $i(r-1)$ and $i(r+1)$ (if $r = 1$, we drop the condition on $i(r-1)$ and if $r = k$, we drop the condition on $i(r+1)$).

Definition 5.2.5 (Anti-monotone independence)**.** Let $(\mathcal{A}, E)$ be a $\mathcal{B}$-valued probability space. Then $\mathcal{B}$-subalgebras $\mathcal{A}_1$, …, $\mathcal{A}_N$ are said to be *anti-monotone independent* if we have

$$E[a_1 \dots a_k] = E[a_1 \dots a_{r-1} E[a_r] a_{r+1} \dots a_k]$$

whenever $a_j \in \mathcal{A}_{i(j)}$, provided that the index i_r is strictly less than the consecutive indices $i(r-1)$ and $i(r+1)$ (if $r = 1$, we drop the condition on $i(r-1)$ and if $r = k$, we drop the condition on $i(r+1)$).

Remark 5.2.6. Boolean and free independence are unchanged if we reorder the algebras $\mathcal{A}_1$, …, $\mathcal{A}_N$. However, monotone and anti-monotone independence are sensitive to order. Also, $\mathcal{A}_1$, …, $\mathcal{A}_N$ are anti-monotone independent if and only if $\mathcal{A}_n$, …, $\mathcal{A}_1$ are monotone independent.

Next, we define what it means for random variables to be independent. In the following, for a self-adjoint X in $\mathcal{A} \supseteq \mathcal{B}$, it will be convenient to denote by $\mathcal{B}\langle X \rangle$ the subalgebra of $\mathcal{A}$ generated by $\mathcal{B}$ and X. This object is strictly speaking not the same thing as the formal polynomial algebra $\mathcal{B}\langle X \rangle$, but this abuse of notation is already entrenched in algebra. We also denote by $\mathcal{B}\langle X \rangle_0$ the polynomials with no constant term, that is,

$$\mathcal{B}\langle X \rangle_0 = \operatorname{Span}\{b_0 X a_1 \dots X b_k : b_j \in \mathcal{B}, k \geq 1\}.$$

Definition 5.2.7. Self-adjoint random variables X_1, …, X_N in $(\mathcal{A}, E)$ are said to be *freely independent* if the algebras $\mathcal{B}\langle X_1 \rangle$, …, $\mathcal{B}\langle X_N \rangle$ are freely independent. Random variables X_1, …, X_N are said to be *boolean / monotone / anti-monotone independent* if the $\mathcal{B}$-algebras $\mathcal{B}\langle X_1 \rangle_0$, …, $\mathcal{B}\langle X_n \rangle_0$ are boolean / monotone / anti-monotone independent.

The definition of independence provides enough information to evaluate the expectation of any element of the $\mathcal{B}$-algebra generated by $\mathcal{A}_1$, …, $\mathcal{A}_N$. Here and in the rest of this chapter, we state the result for all types of independence simultaneously.

Lemma 5.2.8. *Suppose that $\mathcal{A}_1$, …, $\mathcal{A}_N$ are boolean (resp. free, monotone, anti-monotone) independent $\mathcal{B}$-subalgebras, and assume in the free case that they are unital. If $a_j \in \mathcal{A}_{i_j}$ for $j = 1$, …, k, then $E[a_1 \dots a_k]$ is uniquely determined by $E|_{\mathcal{A}_1}$, …, $E|_{\mathcal{A}_N}$.*

Proof for the boolean case. Starting with a string $a_1 \dots a_k$, we can group together consecutive terms if they come from the same algebra. After this regrouping and relabeling, we can assume that any two consecutive terms come from different algebras. Then by definition of boolean independence $E[a_1 \dots a_k] = E[a_1] \dots E[a_k]$. □

Proof for the free case. Let $\mathcal{C}$ be the formal $\mathcal{B}$-algebra generated by $\mathcal{A}_1, \dots, \mathcal{A}_N$, that is, the span of all strings of the form $a_1 \dots a_k$ where a_j and a_{j+1} come from distinct algebras. Let

$$\mathcal{D} = \mathcal{B} + \operatorname{Span}\{a_1 \dots a_k : E[a_j] = 0,\, a_j \in \mathcal{A}_{i(j)},\, i(j) \neq i(j+1)\}.$$

We claim that $\mathcal{C} = \mathcal{D}$.

We must show that every string $a_1 \dots a_k$ can be represented as a linear combination of the terms in $\mathcal{D}$. We prove this by induction on k, the base case $k = 0$ being trivial. In the inductive step, let $k \geq 1$ and consider a string $a_1 \dots a_k$ where $a_j \in \mathcal{A}_{i(j)}$ and $i(j) \neq i_(j+1)$. We can write $a_j = c_j + b_j$ where $b_j = E[a_j]$ and $c_j = a_j - b_j$ has expectation zero. Then

$$a_1 \dots a_k = (c_1 + b_1) \dots (c_k + b_k).$$

We expand the right hand side into 2^k terms using the distributive property. The first term $c_1 \dots c_k$ has the desired form. We claim that each of the other terms can be expressed as a word in $\mathcal{C}$ with length less than k (so that we can apply the inductive hypothesis). Each term is a product of some c_j's and some a_j's, but we can group each a_j together with all the terms before or after until we reach one of the c_j's. Then if two adjacent elements come from the same algebra $\mathcal{A}_i$, then we can group them together into one term. After applying as many such regrouping operations as possible, we have expressed this term as a string of the form $a_1' \dots a_{k'}'$ with $k' < k$ and the terms a_j' coming from different $\mathcal{A}_{i(j)}$'s with $i(j+1) \neq i(j)$. Then by the inductive hypothesis, this term is in $\mathcal{D}$.

This implies that every $c \in \mathcal{C}$ can be expressed as the sum of $b \in \mathcal{B}$ plus a linear combination of terms of the form $a_1 \dots a_k$, where $E[a_j] = 0$, $a_j \in \mathcal{A}_{i(j)}$, and $i(j) \neq i(j+1)$. This decomposition was reached using purely algebraic operations and knowledge of $E|_{\mathcal{A}_i}$ for each i. Using freeness, each term of the form $a_1 \dots a_k$ has expectation zero. Thus, $E[c] = b$. □

Proof for the (anti-)monotone case. In the monotone case, we proceed by induction on the length k of the string $a_1 \dots a_k$, where the base case $k = 1$ is trivial. By regrouping the terms if necessary, assume that consecutive terms come from different algebras. Then choose an index j such that $i(j)$ is maximal. By monotone independence,

$$E[a_1 \dots a_k] = E[a_1 \dots a_{j-1} E[a_j] a_{j+1} \dots a_k].$$

Since $E[a_j] \in \mathcal{B}$, this can be represented as a string of length $\leq k - 1$, to which we apply the induction hypothesis.

The anti-monotone case follows by symmetry from the monotone case. □

5.3 Construction of product spaces

In classical probability theory, one constructs the product $(\Omega, P) = (\Omega_1 \otimes \Omega_2, P_1 \otimes P_2)$ of two probability spaces (Ω_1, P_1) and (Ω_2, P_2). The algebras $\mathcal{A}_1 = L^\infty(\Omega_1, P_1)$ and $\mathcal{A}_2 =$

$L^\infty(\Omega_2, P_2)$ embed into $\mathcal{A} = L^\infty(\Omega, P)$ as subalgebras which are classically independent, that is, $E[b_1 b_2] = E[a_1]E[a_2]$ when $a_j \in L^\infty(\Omega_j, P_j)$. The algebra $\mathcal{A}$ is thus a certain completed tensor product of $\mathcal{A}_1$ and $\mathcal{A}_2$ with the state $\int \cdot \, dP$ being the tensor product of the two states $\int \cdot \, dP_1$ and $\int \cdot \, dP_2$. Moreover, the Hilbert space $L^2(\Omega, P)$ is the Hilbert-space tensor product of $L^2(\Omega_1, P_1)$ and $L^2(\Omega_2, P_2)$.

Similarly, in non-commutative probability, we seek to a way to independently join given $\mathcal{B}$-valued non-commutative probability spaces $(\mathcal{A}_1, E_1)$, ..., $(\mathcal{A}_N, E_N)$. We construct the joint algebra by first constructing a joint Hilbert space, in the same way that classical independence arises from tensor products of Hilbert spaces. The constructions of these product spaces can be found in the following references: scalar-valued boolean [Ber06, §2], operator-valued boolean [PV13, Remark 2.3], scalar-valued free [Voi86, §3], scalar-valued (anti-)monotone [Mur00, §2], [Ber05], operator-valued (anti-)monotone [Pop08, §4].

5.3.1 Independent products of correspondences

Suppose that $(\mathcal{H}_1, \xi_1)$, ..., $(\mathcal{H}_N, \xi_N)$ are $\mathcal{B}$-$\mathcal{B}$-correspondences and ξ_j is a $\mathcal{B}$-central unit vector in $\mathcal{H}_j$. Denote $E_j[a] = \langle \xi_j, a\xi_j \rangle$ for $a \in B(\mathcal{H}_j)$. Let $\mathcal{H}_j^\circ$ be the orthogonal complement of ξ_j in $\mathcal{H}_j$, so that $\mathcal{H}_j = \mathcal{B}\xi_j \oplus \mathcal{H}_j^\circ$ by Lemma 2.5.7. For each ind $\in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\,\dagger\}$ (denoting boolean, free, monotone, and anti-monotone independence respectively), we define $\bigstar_{\text{ind}}[(\mathcal{H}_1, \xi_1), \dots, (\mathcal{H}_N, \xi_N)]$ to be the pair $(\mathcal{H}, \xi)$ of a $\mathcal{B}$-$\mathcal{B}$-correspondence and a central unit vector constructed as follows:

$$\mathcal{H} = \begin{cases} \mathcal{B}\xi \oplus \bigoplus_{j=1}^N \mathcal{H}_j^\circ, & \text{boolean case} \\ \mathcal{B}\xi \oplus \bigoplus_{k \geq 1} \bigoplus_{\substack{j_1, \dots, j_k \in [N] \\ j_r \neq j_{r+1}}} \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \dots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ, & \text{free case} \\ \mathcal{B}\xi \oplus \bigoplus_{k \geq 1} \bigoplus_{\substack{j_1, \dots, j_k \in [N] \\ j_r > j_{r+1}}} \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \dots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ, & \text{monotone case} \\ \mathcal{B}\xi \oplus \bigoplus_{k \geq 1} \bigoplus_{\substack{j_1, \dots, j_k \in [N] \\ j_r < j_{r+1}}} \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \dots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ, & \text{anti-monotone case.} \end{cases}$$

Here $\mathcal{B}\xi$ represents a copy of $\mathcal{B}$ as a $\mathcal{B}$-$\mathcal{B}$-correspondence with ξ corresponding to the vector 1.

Next, we define $*$-homomorphisms $\rho_{\text{ind},j} : B(\mathcal{H}_j) \to B(\mathcal{H})$ as follows. For each type of independence and each j, there is a natural decomposition of $\mathcal{H}$ as

$$\mathcal{H} \cong \mathcal{H}_j \otimes_{\mathcal{B}} \mathcal{M}_j \oplus \mathcal{N}_j$$

given as follows:

(1) In the boolean case, we observe that

$$\mathcal{H} = (\mathcal{B}\xi \oplus \mathcal{H}_j^\circ) \oplus \bigoplus_{i \neq j} \mathcal{H}_i^\circ \cong \mathcal{H}_j \oplus \bigoplus_{i \neq j} \mathcal{H}_i^\circ,$$

so we take $\mathcal{M}_j = \mathcal{B}$ and $\mathcal{N}_j = \bigoplus_{i \neq j} \mathcal{H}_i^\circ$.

(2) In the free case, we split the direct summands into the cases where $j_1 = j$ and $j_1 \neq j$. Thus, we write

$$\begin{aligned}\mathcal{H} &\cong \left(\bigoplus_{k \geq 0} \bigoplus_{\substack{j_1, \dots, j_k \in [N] \\ j_r \neq j_{r+1} \\ j_1 \neq j}} \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ \right) \oplus \mathcal{H}_j^\circ \otimes \left(\bigoplus_{k \geq 0} \bigoplus_{\substack{j_1, \dots, j_k \in [N] \\ j_r \neq j_{r+1} \\ j_1 \neq j}} \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ \right) \\ &\cong (\mathcal{B} \oplus \mathcal{H}_j^\circ) \otimes_{\mathcal{B}} \bigoplus_{k \geq 0} \bigoplus_{\substack{j_1, \dots, j_k \in [N] \\ j_r \neq j_{r+1} \\ j_1 \neq j}} \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ,\end{aligned}$$

where the $k = 0$ terms in the big direct sum are evaluated as $\mathcal{B}$ or $\mathcal{B}\xi$ by convention. Thus, we take $\mathcal{N}_j = 0$ and

$$\mathcal{M}_j = \bigoplus_{k \geq 0} \bigoplus_{\substack{j_1, \dots, j_k \in [N] \\ j_r \neq j_{r+1} \\ j_1 \neq j}} \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ.$$

(3) In the monotone case, we take

$$\mathcal{M}_j = \bigoplus_{k \geq 0} \bigoplus_{\substack{j_1, \dots, j_k \in [N] \\ j_r \neq j_{r+1} \\ j > j_1}} \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ,$$

where the $k = 0$ term is evaluated as $\mathcal{B}$, and

$$\mathcal{N}_j = \bigoplus_{k \geq 1} \bigoplus_{\substack{j_1, \dots, j_k \in [N] \\ j_r \neq j_{r+1} \\ j < j_1}} \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ.$$

The isomorphism $\mathcal{H} \cong \mathcal{H}_j \otimes \mathcal{M}_j \oplus \mathcal{N}_j$ comes from similar manipulations as in the previous cases. The term $\mathcal{H}_j \otimes_{\mathcal{B}} \mathcal{M}_j$ splits as $\mathcal{M}_j \oplus \mathcal{H}_j^\circ \otimes_{\mathcal{B}} \mathcal{M}_j$, and these two pieces yield respectively all the direct summands of $\mathcal{H}$ where the first index is $j_1 < j$ and those where the first index is j. Meanwhile, the $\mathcal{N}_j$ term produces all the direct summand where the first index is greater than j.

(4) The anti-monotone case is the same as the monotone case except with the directions of the inequalities reversed.

In each case, if we denote by Φ_j the (unitary) isomorphism $\mathcal{H} \to \mathcal{H}_j \otimes_{\mathcal{B}} \mathcal{M}_j \oplus \mathcal{N}_j$, we define the $*$-homomorphism $\rho_{\text{ind},j}$ by

$$\rho_{\text{ind},j}(a) = \Phi_j^*[(a \otimes \text{id}_{\mathcal{M}_j}) \oplus 0_{\mathcal{N}_j}]\Phi_j.$$

Note that $\rho_{\text{ind},j}$ is injective because $\rho_{\text{ind},j}(a)$ restricted to the direct summands $\mathcal{B} \oplus \mathcal{H}_j^\circ$ in $\mathcal{H}$ is a itself conjugated by the obvious isomorphism $\mathcal{B} \oplus \mathcal{H}_j^\circ \to \mathcal{H}_j$. Moreover, by the same token

$$\langle \xi, \rho_{\text{ind},j}(a)\xi \rangle = \langle \xi_j, a\xi \rangle,$$

which means that $\rho_{\text{ind},j}$ is expectation-preserving. It is also easy to check that $\rho_{\text{ind},j}$ is a $\mathcal{B}$-$\mathcal{B}$-bimodule map, where $B(\mathcal{H}_j)$ and $B(\mathcal{H})$ are given a $\mathcal{B}$-$\mathcal{B}$-bimodule structure through the embeddings $\mathcal{B} \to B(\mathcal{H}_j)$ and $\mathcal{B} \to B(\mathcal{H})$ given by the left $\mathcal{B}$-module structure of $\mathcal{H}_j$ and $\mathcal{H}$.

5.3.2 Independence on the product space

Theorem 5.3.1. *Let $(\mathcal{H}_j, \xi_j)$ for $j = 1, \dots, N$ be $\mathcal{B}$-$\mathcal{B}$ correspondences with $\mathcal{B}$-central unit vectors. Let $\text{ind} \in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\dagger\}$. Let $(\mathcal{H}, \xi) = \bigstar_{\text{ind}}[(\mathcal{H}_1, \xi_1), \dots, (\mathcal{H}_N, \xi_N)]$. Let $E : B(\mathcal{H}) \to \mathcal{B}$ be the expectation given by the vector ξ. Then the algebras $\rho_{\text{ind},1}(B(\mathcal{H}_1))$, $\dots$, $\rho_{\text{ind},N}(B(\mathcal{H}_N))$ are* ind*-independent in the $\mathcal{B}$-valued probability space $(B(\mathcal{H}), E)$.*

Although we have endeavored as much as possible to present the four independences in a unified way, the proof of this theorem will necessarily be done in cases since the moment conditions for each type of independence are idiosyncratic.

Proof of the boolean case. Let $k \geq 1$, and consider a product of terms $\rho_{j_1}(a_1)$, $\dots$, $\rho_{j_k}(a_k)$. We claim that

$$\rho_{j_1}(a_1) \dots \rho_{j_k}(a_k)\xi = E_{j_1}[a_1] \dots E_{j_k}[a_k]\xi + \zeta,$$

where $\zeta \in \mathcal{H}_{j_1}^\circ$, and we will prove this by induction. The base case $k = 1$ is immediate. Now suppose $k > 1$ and note by induction hypothesis,

$$\zeta := \rho_{j_2}(a_2) \dots \rho_{j_k}(a_k)\xi = E_{j_2}[a_2] \dots E_{j_k}[a_k]\xi + \zeta'$$

with $\zeta' \in \mathcal{H}_{j_2}^\circ$. Since $j_2 \neq j_1$, we have $\rho_{j_1}(a_1)\zeta' = 0$. Meanwhile, if we set $b = E_{j_2}[a_2] \dots E_{j_k}[a_k]$, then

$$\rho_{j_1}(a_1)\zeta = \rho_{j_1}(a_1)b\xi = \langle \xi_{j_1}, a_1 b \xi_{j_1} \rangle \xi + \zeta,$$

where $\zeta \in \mathcal{H}_{j_1}^\circ$ by virtue of the construction of $\rho_{j_1}(a_1)$ and the orthogonal decomposition of $\mathcal{H}_{j_1}$ into $\mathcal{B}\xi_{j_1}$ and $\mathcal{H}_{j_1}^\circ$. But note that

$$\langle \xi_{j_1}, a_1 b \xi_{j_1} \rangle \xi = \langle \xi_{j_1}, a_1 \xi_{j_1} b \rangle \xi = \langle \xi_{j_1}, a_1 \xi_{j_1} \rangle b\xi = E_{j_1}[a_1]E_{j_2}[a_2] \dots E_{j_k}[a_k]\xi,$$

which completes the induction step. It follows from this claim that

$$\begin{aligned} E[\rho_{j_1}(a_1)\dots\rho_{j_k}(a_k)] &= \langle \xi, \rho_{j_1}(a_1)\dots\rho_{j_k}(a_k)\xi\rangle \\ &= \langle \xi, E_{j_1}[a_1]\dots E_{j_k}[a_k]\xi\rangle + \langle \xi, \zeta\rangle \\ &= E_{j_1}[a_1]\dots E_{j_k}[a_k], \end{aligned}$$

which demonstrates boolean independence. □

Proof of the free case. Let $k \geq 1$, and consider a product of terms $\rho_{j_1}(a_1)$, ..., $\rho_{j_k}(a_k)$ where $E_{j_r}(a_r) = 0$ for each r. We claim that

$$\rho_{j_1}(a_1)\dots\rho_{j_k}(a_{j_k})\xi \in \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \dots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ,$$

which we will prove by induction on k. In the case $k = 1$, we express $a_{j_1}\xi_{j_1}$ in $\mathcal{H}_{j_1}$ as $b\xi_{j_1} + \zeta$, where $\zeta \in \mathcal{H}_{j_1}$ and the coefficient $b = \langle \xi_{j_1}, a_{j_1}\xi_{j_1}\rangle$. But by assumption $b = 0$, so that $a_{j_1}\xi_{j_1} \in \mathcal{H}_{j_1}^\circ$. For $k > 1$, we know by inductive hypothesis that

$$\zeta := \rho_{j_2}(a_2)\dots\rho_{j_k}(a_{j_k})\xi \in \mathcal{H}_{j_2}^\circ \otimes_{\mathcal{B}} \dots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ.$$

This sits inside the direct summand

$$\mathcal{H}_{j_2} \otimes_{\mathcal{B}} \dots \otimes_{\mathcal{B}} \mathcal{H}_{j_k} \subseteq (\mathcal{B} \oplus \mathcal{H}_{j_1}^\circ) \otimes_{\mathcal{B}} \mathcal{H}_{j_2}^\circ \otimes_{\mathcal{B}} \dots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ. \subseteq \mathcal{H}_{j_1} \otimes_{\mathcal{B}} \mathcal{M}_{j_1}.$$

Because a_1 maps $\mathcal{B}\xi_{j_1}$ into $\mathcal{H}_{j_1}^\circ$, we know that $\rho_{j_1}(a_1)$ maps ζ into $\mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \dots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ$ as desired.

Therefore, we have $\rho_{j_1}(a_1)\dots\rho_{j_k}(a_{j_k})\xi \in \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \dots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ$, and hence

$$E[\rho_{j_1}(a_1)\dots\rho_{j_k}(a_{j_k})] = \langle \xi, \rho_{j_1}(a_1)\dots\rho_{j_k}(a_{j_k})\xi\rangle = 0,$$

which demonstrates free independence. □

Proof of the (anti-)monotone case. In order to show monotone independence, we must show that

$$E[\rho_{p_s}(a_s)\dots\rho_{p_1}(a_1)\rho_j(a)\rho_{q_1}(a_1')\dots\rho_{q_t}(a_t')] = E[\rho_{p_s}(a_s)\dots\rho_{p_1}(a_1)E_j[a]\rho_{q_1}(a_1')\dots\rho_{q_t}(a_t')],$$

provided that $j > p_1$ if $s > 0$ and $j > q_1$ if $t > 0$, where $a \in B(\mathcal{H}_j)$ and $a_i \in B(\mathcal{A}_{p_i})$ and $a_i' \in B(\mathcal{H}_{q_i})$, and where $E_j[a]$ on the right hand side denotes the multiplication by $E_j[a] \in \mathcal{H}$ using the left $\mathcal{B}$-action on $\mathcal{H}$. This claim is equivalent to

$$\langle \rho_{p_1}(a_1^*)\dots\rho_{p_s}(a_s^*)\xi, (\rho_j(a) - E_j[a])\rho_{q_1}(a_1')\dots\rho_{q_t}(a_t')\xi\rangle = 0.$$

Now we write

$$\begin{aligned} b &= E_j[a] \\ \zeta &= \rho_{p_1}(a_1^*)\dots\rho_{p_s}(a_s^*)\xi \\ \zeta' &= \rho_{q_1}(a_1')\dots\rho_{q_t}(a_t')\xi, \end{aligned}$$

and our goal is to show that $\langle \zeta, \rho_j(a)\zeta' \rangle = \langle \zeta, b\zeta' \rangle$.

Recall that in the definition of $\rho_{\mathrm{mono},j}$, we used the direct sum decomposition $\mathcal{H} \cong \mathcal{M}_j \oplus \mathcal{H}_j^\circ \otimes_{\mathcal{B}} \mathcal{M}_j \oplus \mathcal{N}_j$. With some of abuse of notation, let us pretend this isomorphism is an equality. We claim ζ and ζ' are in $\mathcal{M}_j$. This is clear for ζ if $s = 0$ and hence $\zeta = \xi$. On the other hand, if $s > 0$, this follows because the image of $\rho_1(b_1^*)$ is contained in

$$\mathcal{H}_{p_1} \otimes \mathcal{M}_{p_1} = \mathcal{B} \oplus \bigoplus_{k \geq 1} \bigoplus_{p_1 \geq j_1 > j_2 > \cdots > j_k} \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{H}_{j_k}^\circ \subseteq \mathcal{M}_j.$$

and $p_1 < j$. The argument for ζ' is identical.

By construction, for $y \in B(\mathcal{H}_j)$, the operator $\rho_{\mathrm{mono},j}(y)$ maps $\mathcal{M}_j$ into $(\mathcal{B} \oplus \mathcal{H}_j^\circ) \otimes_{\mathcal{B}} \mathcal{M}_j$. However, since $E_j[a - b] = 0$, $\rho_j(a - b)$ maps the space $\mathcal{M}_j$ into $\mathcal{H}_j^\circ \otimes \mathcal{M}_j$, which is orthogonal to $\mathcal{M}_j$. Since $\rho_j(b)|_{\mathcal{M}_j} = b|_{\mathcal{M}_j}$, we have $\rho_j(a - b)\zeta' = [\rho_j(a) - b]\zeta'$. Thus, $[\rho_j(a) - b]\zeta' \in \mathcal{H}_j^\circ \otimes_{\mathcal{B}} \mathcal{M}_j$ and $\zeta \in \mathcal{M}_j$ are orthogonal, so that $\langle \zeta, [\rho_j(a) - b]\zeta' \rangle = 0$, which proves the desired equality to show monotone independence. The anti-monotone case of course is symmetrical. □

5.4 Associativity

For ind $\in \{\mathrm{bool}, \mathrm{free}, \mathrm{mono}, \mathrm{mono}\dagger\}$, the product operation $\bigstar_{\mathrm{ind}}$ satisfies the following associativity properties. These results are well known and we consider them folklore. For $\mathcal{B}$-$\mathcal{B}$-correspondences $\mathcal{H}$ and $\mathcal{K}$ and a bounded adjointable operator $T : \mathcal{H} \to \mathcal{K}$, let us denote by $\mathrm{Ad}(T)$ the map $B(\mathcal{H}) \to B(\mathcal{K})$ given by $S \mapsto TST^*$. There is a canonical isomorphism

$$\Phi : \bigstar_{\mathrm{ind}}[(\mathcal{H}_1, \xi_1), (\mathcal{H}_2, \xi_2), (\mathcal{H}_3, \xi_3)] \to \bigstar_{\mathrm{ind}}[(\mathcal{H}_1, \xi_1), \bigstar_{\mathrm{ind}}[(\mathcal{H}_2, \xi_2), (\mathcal{H}_3, \xi_3)]]$$

such that the following diagrams commute:

$$\begin{array}{ccc} B(\mathcal{H}_1) & \xrightarrow{\rho_{\mathrm{ind},1}} & B\left(\bigstar_{\mathrm{ind}}[(\mathcal{H}_1, \xi_1), (\mathcal{H}_2, \xi_2), (\mathcal{H}_3, \xi_3)]\right) \\ & \searrow^{\rho_{\mathrm{ind},1}} & \downarrow \mathrm{Ad}(\Phi) \\ & & B\left(\bigstar_{\mathrm{ind}}[(\mathcal{H}_1, \xi_1), \bigstar_{\mathrm{ind}}[(\mathcal{H}_2, \xi_2), (\mathcal{H}_3, \xi_3)]]\right), \end{array}$$

and for $j = 2, 3$,

$$\begin{array}{ccc} B(\mathcal{H}_j) & \xrightarrow{\rho_{\mathrm{ind},j}} & B\left(\bigstar_{\mathrm{ind}}[(\mathcal{H}_1, \xi_1), (\mathcal{H}_2, \xi_2), (\mathcal{H}_3, \xi_3)]\right) \\ \rho_{\mathrm{ind},j} \downarrow & & \downarrow \mathrm{Ad}(\Phi) \\ B\left(\bigstar_{\mathrm{ind}}[(\mathcal{H}_2, \xi_2), (\mathcal{H}_3, \xi_3)]\right) & \xrightarrow{\rho_{\mathrm{ind},2}} & B\left(\bigstar_{\mathrm{ind}}[(\mathcal{H}_1, \xi_1), \bigstar_{\mathrm{ind}}[(\mathcal{H}_2, \xi_2), (\mathcal{H}_3, \xi_3)]]\right), \end{array}$$

The construction of the isomorphism Φ is a straighforward rearrangement of the summands in the product space, using the distributive and associative properties of tensor products. Let us describe this construction in the monotone case as an example. Denote

$$(\mathcal{K}, \zeta) = \bigstar_{\mathrm{mono}}[(\mathcal{H}_2, \xi_2), (\mathcal{H}_3, \xi_3)].$$

Then $\bigstar_{\text{mono}}[(\mathcal{H}_1,\xi_1),(\mathcal{K},\zeta)]$ is given by

$$\begin{aligned}
&\mathcal{B}\xi \oplus \mathcal{H}_1^\circ \oplus \mathcal{K}^\circ \oplus (\mathcal{K}^\circ \otimes_{\mathcal{B}} \mathcal{H}_1^\circ)\\
\cong&\mathcal{B}\xi \oplus \mathcal{H}_1^\circ \oplus [\mathcal{H}_2^\circ \oplus \mathcal{H}_3^\circ \oplus (\mathcal{H}_3^\circ \otimes_{\mathcal{B}} \mathcal{H}_2^\circ)] \oplus ([\mathcal{H}_2^\circ \oplus \mathcal{H}_3^\circ \oplus (\mathcal{H}_3^\circ \otimes_{\mathcal{B}} \mathcal{H}_2^\circ)] \otimes_{\mathcal{B}} \mathcal{H}_1^\circ)\\
\cong&\mathcal{B}\xi \oplus \mathcal{H}_1^\circ \oplus \mathcal{H}_2^\circ \oplus \mathcal{H}_3^\circ \oplus (\mathcal{H}_3^\circ \otimes_{\mathcal{B}} \mathcal{H}_2^\circ) \oplus (\mathcal{H}_2^\circ \otimes_{\mathcal{B}} \mathcal{H}_1^\circ) \oplus (\mathcal{H}_3^\circ \otimes_{\mathcal{B}} \mathcal{H}_1^\circ) \oplus (\mathcal{H}_3^\circ \otimes_{\mathcal{B}} \mathcal{H}_2^\circ \otimes_{\mathcal{B}} \mathcal{H}_1^\circ),
\end{aligned}$$

which is precisely $\bigstar_{\text{mono}}[(\mathcal{H}_1,\xi_1),(\mathcal{H}_2,\xi_2),(\mathcal{H}_3,\xi_3)]$. The commutativity of the diagram means that both ways of decomposing the product space will produce "the same" inclusion of $B(\mathcal{H}_j)$ into the bounded operators on the product space. This is a direct verification that we leave to the reader.

More generally, given $\mathcal{B}$-$\mathcal{B}$-correspondences $(\mathcal{H}_{i,j},\xi_{i,j})$ for $i = 1, \dots, k$ and $j = 1, \dots, n_i$, we have an isomorphism

$$\begin{aligned}
&\bigstar_{\text{ind}}[(\mathcal{H}_{1,1},\xi_{1,1}),\dots,(\mathcal{H}_{1,n_1},\xi_{1,n_1}),\dots\dots,(\mathcal{H}_{k,1},\xi_{k,1}),\dots,(\mathcal{H}_{k,n_k},\xi_{k,n_k})]\\
&\qquad\cong \bigstar_{\text{ind}}\big[\bigstar_{\text{ind}}[(\mathcal{H}_{1,1},\xi_{1,1}),\dots,(\mathcal{H}_{1,n_1},\xi_{1,n_1})],\dots,\bigstar_{\text{ind}}[(\mathcal{H}_{k,1},\xi_{k,1}),\dots,(\mathcal{H}_{k,n_k},\xi_{k,n_k})]\big]
\end{aligned}$$

with the associated commutative diagrams. A complete argument for this fact (and in much greater generality than only the boolean, free, and (anti-)monotone cases) can be found in [JL19, Theorem 5.2].

As a corollary of this result and Lemma 5.2.8, we have the following method for checking independence of subalgebras. We leave the argument to the reader; it can alternatively be proved directly from the moment conditions defining independence.

Lemma 5.4.1. *Let* ind $\in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\dagger\}$. *Let* $(\mathcal{A}, E)$ *be a* $\mathcal{B}$*-valued probability space, and* $\mathcal{A}_{i,j}$ *for* $i = 1,\dots,k$ *and* $j = 1,\dots,n_i$ *be* $\mathcal{B}$*-subalgebras (assumed to the unital in the free case). The following are equivalent:*

1. $\mathcal{A}_{1,1}, \dots, \mathcal{A}_{1,n_1}, \dots\dots, \mathcal{A}_{k,1}, \dots, \mathcal{A}_{k,n_k}$ *are independent.*
2. *The algebras* $\mathcal{A}_{1,1} \vee \dots \vee \mathcal{A}_{1,n_1}, \dots, \mathcal{A}_{k,1} \vee \dots \vee \mathcal{A}_{k,n_k}$ *are independent and for each* i*, the algebras* $\mathcal{A}_{i,1}, \dots, \mathcal{A}_{i,n_i}$ *are independent.*

Here $\mathcal{A}_{i,1} \vee \dots \vee \mathcal{A}_{i,n_i}$ *denotes the* $\mathcal{B}$*-subalgebra generated by* $\mathcal{A}_{i,1}, \dots, \mathcal{A}_{i,n_i}$*.*

5.5 Convolution and analytic transforms

Definition 5.5.1. Let ind $\in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\dagger\}$. We define the ind-convolution of two $\mathcal{B}$-valued laws μ and ν as the law of $X + Y$ when X and Y are ind-independent and the law of X is μ and the law of Y is ν. The convolution is denoted by $\mu \boxplus_{\text{ind}} \nu$, or alternatively

$$\begin{aligned}
&\mu \uplus \nu \quad &&\text{(boolean case)}\\
&\mu \boxplus \nu \quad &&\text{(free case)}\\
&\mu \rhd \nu \quad &&\text{(monotone case)}\\
&\mu \lhd \nu \quad &&\text{(anti-monotone case).}
\end{aligned}$$

In order to verify this definition makes sense, observe first that using the GNS construction (Theorem 2.6.5) and the product space construction, there always exist independent operators X and Y in a $\mathcal{B}$-valued probability space $(\mathcal{A}, E)$ such that the law of X is μ and the law of Y is ν. And second, the law of $X + Y$ is uniquely determined by μ and ν and the independence of X and Y by Lemma 5.2.8.

More generally, given laws $\mu_1, \dots, \mu_N$, we may construct operators $X_1, \dots, X_N$ which are independent with X_j having the law μ_j, using the GNS construction and the product space construction. Then it follows from the associativity considerations of the previous section that $X_1 + \dots + X_N$ has the law $\mu_1 \boxplus_{\text{ind}} (\mu_2 \boxplus_{\text{ind}} (\dots (\mu_N) \dots))$. Moreover, we also have $\mu_1 \boxplus_{\text{ind}} (\mu_2 \boxplus_{\text{ind}} \mu_3) = (\mu_1 \boxplus_{\text{ind}} \mu_2) \boxplus_{\text{ind}} \mu_3$, or in other words $\boxplus_{\text{ind}}$ is associative, and thus we may remove the parentheses when expressing an iterated ind-convolution.

Observation 5.5.2.

(1) The operations $\uplus$, $\boxplus$, $\rhd$, and $\lhd$ are associative.

(2) The operations $\boxplus$ and $\uplus$ are commutative.

(3) We have $\mu \rhd \nu = \nu \lhd \mu$.

Proof. The first claim follows from the preceding discussion of associativity. The second claim is true because the conditions defining free and boolean independence do not depend on the order of the subalgebras, while for the third claim, if we were to reverse the order of the indices in monotone independence, then we obtain anti-monotone independence. □

Our main task in this section is to develop analytic tools for computing the independent convolution of two laws. In the classical case, this role is played by characteristic function (Fourier transform) of a law given by $\mathcal{F}\mu(\xi) = \int e^{ix\xi}\, d\mu(x)$, since addition of independent random variables or classical convolution of laws corresponds to multiplication of the Fourier transforms. In the non-commutative setting, this role is played by various fully matricial functions related to the Cauchy-Stieltjes transform.

5.5.1 The boolean case

The results of this section can be found in [SW97] [Ber06, Theorem 2.2] for the scalar case and [PV13, §2 and §5.3] in the operator-valued case. The proof we give here is based on analogy with the proof from the free case in the next subsection (where we also explain the history and references).

Definition 5.5.3. For a $\mathcal{B}$-valued law μ, we define the *K-transform* as

$$K_\mu(z) := z - F_\mu(z).$$

Remark 5.5.4. We caution that some authors work instead with $B_\mu(z) = \tilde{K}_\mu(z)$ or slight variants of this definition. We showed in Theorem 4.5.3 that $K_\mu(z) = \mu(X)^{(n)} + G_\sigma(z)$ for some generalized law σ.

Theorem 5.5.5. *$K_{\mu\uplus\nu}(z) = K_\mu(z) + K_\nu(z)$ as fully matricial functions.*

Proof. Let X and Y be freely independent random variables in $(\mathcal{A}, E)$ which realize the laws μ and ν respectively. For z in $M_n(\mathcal{B})$ with $\|z\| < 1/\operatorname{rad}(\mu)$, define

$$U_X^{(n)}(z) = (1 - zX^{(n)})^{-1} - 1 = \sum_{k=1}^{\infty} (zX^{(n)})^k.$$

This is an $\mathcal{A}$-valued fully matricial function. To simplify the notation, we will suppress all the superscripts (n), so that X will stand for $X^{(n)}$, where n is the size of the matrix z. Note that

$$1 + E[U_X(z)] = E[(1 - zX)^{-1}] = \tilde{G}_\mu(z)z^{-1}$$

or in other words

$$(1 + E[U_X(z)])^{-1} = z\tilde{F}_\mu(z)$$

Note that $U_X(z)$ is in the closed span of $\mathcal{B}\langle X\rangle_0$. Define $U_Y(z)$ analogously. Then

$$1 - zX - zY = (1 + U_X(z))^{-1} + (1 + U_Y(z))^{-1} - 1$$

Therefore,

$$\begin{aligned}(1 - zX - zY)^{-1} &= [(1 + U_X(z))^{-1} + (1 + U_Y(z))^{-1} - 1]^{-1} \\ &= (1 + U_X(z))[1 - U_Y(z)U_X(z)]^{-1}(1 + U_Y(z)) \\ &= (1 + U_X(z)) \left(\sum_{k=0}^{\infty} (U_Y(z)U_X(z))^k \right) (1 + U_Y(z)).\end{aligned}$$

Next, we take the expectation. Because $U_X(z)$ and $U_Y(z)$ are in the closures of $M_n(\mathcal{B}\langle X\rangle_0)$ and $M_n(\mathcal{B}\langle Y\rangle_0)$ respectively and because X and Y are Boolean independent, we have

$$\begin{aligned}E[(1 - zX - zY)^{-1}] &= (1 + E[U_X(z)]) \left(\sum_{k=0}^{\infty} (E[U_Y(z)]E[U_X(z)])^k \right) (1 + E[U_Y(z)]) \\ &= [(1 + E[U_X(z)])^{-1} + (1 + E[U_Y(z)])^{-1} - 1]^{-1}\end{aligned}$$

Therefore,

$$\tilde{G}_{\mu\uplus\nu}(z)z^{-1} = [(1 + E[U_X(z)])^{-1} + (1 + E[U_Y(z)])^{-1} - 1]^{-1}$$

By taking reciprocals,

$$\begin{aligned}z\tilde{F}_{\mu\uplus\nu}(z) &= (1 + E[U_X(z)])^{-1} + (1 + E[U_Y(z)])^{-1} - 1 \\ &= z\tilde{F}_\mu(z) + z\tilde{F}_\nu(z) - 1,\end{aligned}$$

Because $z\tilde{F}_\mu(z) - 1 = z\tilde{K}_\mu(z)$ and the same holds for Y and $X + Y$, this means precisely that

$$z\tilde{K}_{\mu \uplus \nu}(z) = z\tilde{K}_\mu(z) + z\tilde{K}_\nu(z)$$

for z in a neighborhood of 0. By Corollary 3.9.7, we have $K_{\mu \uplus \nu} = K_\mu + K_\nu$ on the upper half plane. □

5.5.2 The free case

The following analytic transforms were defined by Voiculescu [Voi86]. In the operator-valued case, the definition was developed by Dykema [Dyk07, §6].

Definition 5.5.6. For a $\mathcal{B}$-valued law μ, we define $F_\mu(z) = G_\mu(z)^{-1}$ and

$$\Phi_\mu(z) := F_\mu^{-1}(z) - z,$$

where $F_\mu^{-1}(z)$ is the functional inverse and z is in the image of F_μ.

Remark 5.5.7. Many authors work with the R-transform $R_\mu(z) = \tilde{\Phi}_\mu(z) = \Phi_\mu(z^{-1})$. We showed in Lemma 4.5.8 that Φ_μ is defined for $\operatorname{Im} z \geq 2\delta$ whenever $\delta > \|\operatorname{Var}_\mu[1]\|^{1/2}$ and in Lemma 4.5.9 that $R_\mu(z)$ is defined in a fully matricial ball around zero.

The following result on the additivity of the R-transform was discovered in the scalar-valued case by Voiculescu [Voi86]. The original proof by Voiculescu used canonical realizations of a law μ by (non-self-adjoint) random variables on a Fock space, and this was adapted to the operator-valued setting by Dykema [Dyk07, §6]. This theorem can also be proved through the combinatorial apparatus of free cumulants due to Speicher [Spe94, Spe98]. The analytic proof presented here is due (in the scalar-valued setting) to Lehner [Leh01, Theorem 3.1].

Theorem 5.5.8. *For* $\operatorname{Im} z \geq 2\delta > 2\|\operatorname{Var}_\mu(1) + \operatorname{Var}_\nu(1)\|^{1/2}$*, we have*

$$\Phi_{\mu \boxplus \nu}(z) = \Phi_\mu(z) + \Phi_\nu(z).$$

Also, for z *in a fully matricial neighborhood of* 0*, we have*

$$R_{\mu \boxplus \nu}(z) = R_\mu(z) + R_\nu(z)$$

Proof. Let X and Y be freely independent random variables in $(\mathcal{A}, E)$ which realize the laws μ and ν.

We begin by analyzing $R_\mu(z)$ in a neighborhood of the origin. Now $z^{-1} + R_\mu(z)$ is the functional inverse of $G_\mu(z)$ in a neighborhood of 0 which means that

$$E[(z^{-1} + R_\mu(z) - X)^{-1}] = z.$$

Multiplying by z^{-1} on the right, we can write rewrite this as

$$E[(1 + zR_\mu(z) - zX)^{-1}] = 1,$$

or in other words, the $\mathcal{A}$-valued fully matricial function

$$U_X(z) = (1 + zR_\mu(z) - zX)^{-1} - 1$$

has expectation zero. (Here, as in the previous case, we suppress the superscripts (n) but $U_X(z)$ stands for $U_X^{(n)}(z)$ and X denotes $X^{(n)}$ where n is the size of the matrix z). The same holds for the analogously-defined function $U_Y(z)$. We want to show that $z^{-1} - R_\mu(z) - R_\nu(z)$ is the functional inverse of $G_{\mu \boxplus \nu}$, which means that

$$G_{\mu \boxplus \nu}(z^{-1} + R_\mu(z) + R_\nu(z)) = z,$$

which after multiplying by z^{-1} on the right is equivalent to

$$E[(1 + zR_\mu(z) + zR_\nu(z) - zX - zY)^{-1}] = 1.$$

We will rewrite the left hand side in terms of $U_X(z)$ and $U_Y(z)$ so that we can apply freeness together with the fact that $U_X(z)$ and $U_Y(z)$ have expectation zero. Note that

$$\begin{aligned}
&(1 + zR_\mu(z) + zR_\nu(z) - zX - zY)^{-1} \\
&\quad = [(1 + U_X(z))^{-1} + (1 + U_Y(z))^{-1} - 1]^{-1} \\
&\quad = (1 + U_X(z))[(1 + U_Y(z)) + (1 + U_X(z)) - (1 + U_Y(z))(1 + U_X(z))]^{-1}(1 + U_Y(z)) \\
&\quad = (1 + U_X(z))[1 - U_Y(z)U_X(z)]^{-1}(1 + U_Y(z)).
\end{aligned}$$

Now because $U_X(0) = 0 = U_Y(0)$, we know that if $\|z\|$ is sufficiently small, then we can expand $[1 - U_Y(z)U_X(z)]^{-1}$ as a geometric series, and thus for small z,

$$(1 - zR_\mu(z) - zR_\nu(z) - zX - zY)^{-1} = (1 + U_X(z)) \left(\sum_{k=0}^{\infty} (U_Y(z)U_X(z))^k \right) (1 + U_Y(z)).$$

Next, we take the expectation. Because $U_X(z)$ and $U_Y(z)$ have expectation zero and because X and Y are free, all the terms on the right hand side have zero expectation except the term 1 which comes from multiplying together the 1 from $1 + U_X(z)$, the 1 from the geometric series, and the 1 from $1 + U_Y(z)$. Therefore, as desired,

$$E[(1 - zR_\mu(z) - zR_\nu(z) - zX - zY)^{-1}] = 1.$$

This shows that

$$R_{\mu \boxplus \nu}(z) = R_\mu(z) + R_\nu(z)$$

holds in a neighborhood of zero.

This implies that $\Phi_{\mu \boxplus \nu} = \Phi_\mu + \Phi_\nu$ if $\operatorname{Im} z$ is sufficiently large, and hence by Corollary 3.9.7, we have $\Phi_{\mu \boxplus \nu} = \Phi_\mu + \Phi_\nu$ on $\mathbb{H}_{+,2\delta}(\mathcal{B})$, provided that this lies inside the common domain of $\Phi_{\mu \boxplus \nu}$, Φ_μ, and Φ_ν. Since $\operatorname{Var}_{\mu \boxplus \nu}(1) = \operatorname{Var}_\mu(1) + \operatorname{Var}_\nu(1)$ and all these elements are positive, we have $\|\operatorname{Var}_{\mu \boxplus \nu}(1)\| \geq \max(\|\operatorname{Var}_\mu(1)\|, \|\operatorname{Var}_\nu(1)\|)$, and hence it is sufficient that $\delta > \|\operatorname{Var}_{\mu \boxplus \nu}(1)\|^{1/2}$. $\square$

5.5.3 The (anti-)monotone case

The following result is due to [Mur00, Theorem 3.1] in the scalar-valued case and [Pop08, Theorems 3.2 and 3.7] in the operator-valued case, whose proof we follow here. Another proof in the scalar case is in [Ber05].

Theorem 5.5.9. *We have* $F_{\mu \rhd \nu}(z) = F_\mu(F_\nu(z))$ *and* $F_{\mu \lhd \nu}(z) = F_\nu(F_\mu(z))$ *as fully matricial functions.*

Proof. Let inv denote the fully matricial function $z \mapsto z^{-1}$ where defined. Since $F_\mu = \operatorname{inv} \circ \tilde{G}_\mu \circ \operatorname{inv}$ and inv is an involution, it suffices to show that $\tilde{G}_{\mu \rhd \nu} = \tilde{G}_\mu \circ \tilde{G}_\nu$.

Let X and Y be monotone independent random variables in $(\mathcal{A}, E)$ realizing the laws μ and ν. We know that

$$E[f_0(Y)g_1(X)f_1(Y)\dots g_n(X)f_n(Y)] = E[E[f_0(Y)]g_1(X)E[f_1(Y)]\dots g_n(X)E[f_n(Y)]]$$

whenever $f(Y) \in \mathcal{B}\langle Y \rangle_0$ and $f(X) \in \mathcal{B}\langle X \rangle_0$. However, this also holds trivially if $f_j(Y) \in \mathcal{B}$, and thus by linearity it holds when $f_j(Y) \in \mathcal{B}\langle Y \rangle$.

Now for $\|z\|$ sufficiently small, we have

$$\begin{aligned} \tilde{G}_{\mu \rhd \nu}(z) &= E[(1 - zX - zY)^{-1} z] \\ &= E[(1-(1-zY)^{-1}zX)^{-1}(1-zY)^{-1}z] \quad = E\left[\sum_{k=1}^\infty [(1-zY)^{-1}zX]^k (1-zY)^{-1}z\right]. \end{aligned}$$

Note that $(1 - zY)^{-1}$ is in the closure of $M_n(\mathcal{B}\langle Y \rangle)$ and $zX \in M_n(\mathcal{B}\langle X \rangle_0)$ and hence by monotone independence

$$\begin{aligned} \tilde{G}_{\mu \rhd \nu}(z) &= E\left[\sum_{k=1}^\infty [E[(1-zY)^{-1}z]X]^k E[(1-zY)^{-1}z]\right] \\ &= E\left[\sum_{k=1}^\infty [\tilde{G}_\nu(z)X]^k \tilde{G}_\nu(z)\right] \\ &= \tilde{G}_\mu \circ \tilde{G}_\nu(z). \end{aligned}$$

This equality extends to all z by Corollary 3.9.7. The anti-monotone case follows from the monotone case since $\mu \lhd \nu = \nu \rhd \mu$. □

CHAPTER 6

Tools: Norm estimates and subordination

6.1 The norm of an independent sum

The sum of "independent" operators on the product space $\mathcal{H}$ can be estimated as follows. The statement and proof are very close to the scalar-valued free case proved in [Voi86, Lemma 3.2]. The exact formulation of the proof here is taken from [JL19, Proposition 3.18].

Lemma 6.1.1. *Let $(\mathcal{H}_j, \xi_j)$ for $j = 1, \dots, N$ be $\mathcal{B}$-$\mathcal{B}$ correspondences with $\mathcal{B}$-central unit vectors. Let $\mathrm{ind} \in \{\mathrm{bool}, \mathrm{free}, \mathrm{mono}, \mathrm{mono}\,\dagger\}$. Let $(\mathcal{H}, \xi) = \bigstar_{\mathrm{ind}}[(\mathcal{H}_1, \xi_1), \dots, (\mathcal{H}_N, \xi_N)]$. For $j = 1, \dots, N$, let $a_j \in B(\mathcal{H}_j)$ with $\langle \xi_j, a_j \xi_j \rangle = 0$. Then we have*

$$\left\| \sum_{j=1}^N \rho_{\mathrm{ind},j}(a_j) \right\| \leq \left\| \sum_{j=1}^N \langle a_j \xi_j, a_j \xi_j \rangle \right\|^{1/2} + \left\| \sum_{j=1}^N \langle a_j^* \xi_j, a_j^* \xi_j \rangle \right\|^{1/2} + \max_{j=1,\dots,N} \|a_j\|$$
$$\leq 2 \left(\sum_{j=1}^N \|a_j\|^2 \right)^{1/2} + \max_{j=1,\dots,N} \|a_j\|.$$

Conversely,

$$\max_{j=1,\dots,N} \|a_j\| \leq \left\| \sum_{j=1}^N \rho_{\mathrm{ind},j}(a_j) \right\|.$$

Proof. Let $P_j \in \mathcal{L}(\mathcal{H}_j)$ be the projection onto ξ_j and let $Q_j = 1 - P_j$. Because $\langle \xi_j, a_j \xi_j \rangle = 0$, we have $P_j a_j P_j = 0$, and hence

$$a_j = Q_j a_j P_j + P_j a_j Q_j + Q_j a_j Q_j.$$

Thus, we will estimate

$$\left\| \sum_{j=1}^N \rho_{\mathrm{ind},j}(a_j) \right\| \leq \left\| \sum_{j=1}^N \rho_{\mathrm{ind},j}(Q_j a_j P_j) \right\| + \left\| \sum_{j=1}^N \rho_{\mathrm{ind},j}(P_j a_j Q_j) \right\| + \left\| \sum_{j=1}^N \rho_{\mathrm{ind},j}(Q_j a_j Q_j) \right\|.$$

For the first term, let us use the notation $\mathcal{M}_j$ defined in §5.3 and observe that $\rho_{\mathrm{ind},j}(Q_j)$ is the projection onto $\mathcal{H}_j^\circ \otimes \mathcal{M}_j$, that is, the direct sum of the tensor products where the first index is j. Thus, the ranges of $\rho_{\mathrm{ind},j}(Q_j)$ are orthogonal, and hence

$$\left(\sum_{j=1}^{N}\rho_{\mathrm{ind},j}(Q_j a_j P_j)\right)^*\left(\sum_{j=1}^{N}\rho_{\mathrm{ind},j}(Q_j a_j P_j)\right)=\sum_{j=1}^{N}\rho_{\mathrm{ind},j}(P_j a_j^* Q_j a_j P_j).$$

Now $P_j a_j P_j = 0$ implies that

$$(Q_j a_j P_j)^*(Q_j a_j P_j) = P_j a_j^* Q_j a_j P_j = P_j a_j^* a_j P_j = \langle a_j\xi_j, a_j\xi_j\rangle P_j,$$

where the last equality follows because P_j is the projection onto $\mathcal{B}\xi_j$. Thus,

$$\left\|\sum_{j=1}^{N}\rho_{\mathrm{ind},j}(Q_j a_j P_j)\right\| = \left\|\sum_{j=1}^{N}\langle a_j\xi_j, a_j\xi_j\rangle\rho_{\mathrm{ind},j}(P_j)\right\|^{1/2}$$

Let $b_j = \langle a_j\xi_j, a_j\xi_j\rangle$ and note that $b_j \geq 0$ in $\mathcal{B}$. Also, b_j and P_j commute in $B(\mathcal{H}_j)$. Therefore, we have $b_j\rho_{\mathrm{ind},j}(P_j) = \rho_{\mathrm{ind},j}(P_j)b_j$ since $\rho_{\mathrm{ind},j}$ is a $\mathcal{B}$-$\mathcal{B}$-bimodule map. This implies that in $B(\mathcal{H})$ we have

$$b_j\rho_{\mathrm{ind},j}(P_j) = b_j^{1/2}\rho_{\mathrm{ind},j}(P_j)b_j^{1/2} \leq b_j.$$

Thus,

$$\sum_{j=1}^{N}\langle a_j\xi_j, a_j\xi_j\rangle\rho_{\mathrm{ind},j}(P_j) \leq \sum_{j=1}^{n}\langle a_j\xi_j, a_j\xi_j\rangle.$$

and so

$$\left\|\sum_{j=1}^{N}\rho_{\mathrm{ind},j}(Q_j a_j P_j)\right\| \leq \left\|\sum_{j=1}^{n}\langle a_j\xi_j, a_j\xi_j\rangle\right\|^{1/2}.$$

Similarly,

$$\left\|\sum_{j=1}^{N}\rho_{\mathrm{ind},j}(P_j a_j Q_j)\right\| = \left\|\sum_{j=1}^{N}\rho_{\mathrm{ind},j}(Q_j a_j^* P_j)\right\| \leq \left\|\sum_{j=1}^{n}\langle a_j\xi_j, a_j\xi_j\rangle\right\|^{1/2}.$$

Finally, because the projections $\rho_{\mathrm{ind},j}(Q_j)$'s have orthogonal ranges, we have

$$\left\|\sum_{j=1}^{N}\rho_{\mathrm{ind},j}(Q_j a_j Q_j)\right\| = \max_j\|\rho_{\mathrm{ind},j}(Q_j a_j Q_j)\| \leq \max_j\|a_j\|.$$

Adding the estimates for the three terms together completes the proof of the first estimate.

For the converse estimate, note that $\mathcal{H}$ contains the direct summands $\mathcal{B}\xi$ and $\mathcal{H}_j^\circ$, which together form a copy of $\mathcal{H}_j$. Moreover, the compression of $\sum_{j=1}^{N}\rho_{\mathrm{ind},j}(a_j)$ to this subspace is a copy of a_j. Hence, as desired

$$\|a_j\| \leq \left\|\sum_{j=1}^{N}\rho_{\mathrm{ind},j}(a_j)\right\|. \qquad \square$$

This proposition implies in particular that if $\mu_1, \ldots, \mu_N$ are $\mathcal{B}$-valued laws for self-adjoint operators with mean zero, then

$$\operatorname{rad}(\mu_1 \boxplus_{\mathrm{ind}} \cdots \boxplus_{\mathrm{ind}} \mu_N) \leq 2 \left(\sum_{j=1}^N \operatorname{rad}(\mu_j)^2 \right)^{1/2} + \max_{j=1,\ldots,N} \operatorname{rad}(\mu_j).$$

More precisely, the term $\langle a_j \xi_j, a_j \xi_j \rangle$ on the right hand side of Lemma 6.1.1 is the norm of the variance of μ_j. Also, the term $\max \|a_j\|$ can actually be replaced by $\|Q_j a_j Q_j\|$. Proceeding as in the proof of Theorem 4.5.3, if $(0, \sigma_j)$ corresponds to μ_j by $F_{\mu_j}(z) = z - G_{\sigma_j}(z)$, then we can realize the law μ_j by an operator a_j on $\mathcal{B}\xi_j \oplus (\mathcal{B}\langle Y \rangle \otimes_\sigma \mathcal{B})$ such that $Q_j a_j Q_j = Y$, and hence $\|Q_j a_j Q_j\| = \operatorname{rad}(\sigma_j)$. Thus, we have

$$\operatorname{rad}(\mu_1 \boxplus_{\mathrm{ind}} \cdots \boxplus_{\mathrm{ind}} \mu_N) \leq 2 \left\| \sum_{j=1}^N \operatorname{Var}(\mu_j)[1] \right\|^{1/2} + \max_{j=1,\ldots,N} \operatorname{rad}(\sigma_j). \tag{6.1}$$

One would hope that conversely $\operatorname{rad}(\mu_j) \leq \operatorname{rad}(\mu_1 \boxplus_{\mathrm{ind}} \cdots \boxplus_{\mathrm{ind}} \mu_N)$ (still in the mean-zero case). However, this does not immediately follow since a priori we cannot rule out the possibility that $\|\rho_{\mathrm{ind},1}(a_1) + \cdots + \rho_{\mathrm{ind},N}(a_N)\|$ could be larger than $\operatorname{rad}(\mu_1 \boxplus_{\mathrm{ind}} \cdots \boxplus_{\mathrm{ind}} \mu_N)$.

This issue does not arise in the scalar-valued setting for algebras with a *faithful* state since then the norm of an operator is always the same as its spectral radius with respect to the state. However, for boolean and monotone independence, the states constructed on product Hilbert spaces are often not faithful on the algebra generated by the $\rho_{\mathrm{ind},j}(a_j)$'s. Nonetheless, the desired sharp radius estimates are not difficult to establish in the scalar-valued setting by complex-analytic methods.

For the operator-valued setting, we will prove slightly less sharp "reverse" radius estimates in §6.3. Our method will reduce the free case to the monotone case, and the monotone case to the boolean case using the theory of subordination for operator-valued independences (§6.2, which has significant interest in its own right.

6.2 Subordination

This section will show that $\mu_1 \boxplus \mu_2$ can also be written as $\mu_1 \rhd \nu$ for some law ν, and hence $F_{\mu_1 \boxplus \mu_2} = F_{\mu_1} \circ F_\nu$ and $G_{\mu_1 \boxplus \mu_2} = G_{\mu_1} \circ F_\nu$. This result, known as "analytic subordination", has a long history in free probability. In the scalar case, it was first proved by Voiculescu [Voi93, Proposition 4.4], although under some additional transversality-type assumptions. The subordination theory was further developed by Biane [Bia98, Theorem 3.1], who connected it with conditional expectations and Markov transition kernels. The operator-valued setting was addressed in the sense of formal power series by [Voi00] and [Voi02a]. In the sense of analytic functions, the operator-valued case was addressed by [BMS13] using analytic tools such as the Earle-Hamilton theorem, and a numerically viable approximation scheme was described.

We will instead follow the approach of Lenszewski [Len07] that uses operator models rather than complex analysis. We will construct an operator explicitly that realizes the law ν and show that $\mu_1 \boxplus \mu_2 = \mu_1 \rhd \nu$ by manipulating free and monotone product $\mathcal{B}$-$\mathcal{B}$-correspondences. Proposition 6.2.2 and its proof are implicit in Lenczewski's work in the scalar-valued case [Len07, §7]; the multivariable case was considered in [Nic09] (see Remark 4.11); and the operator-valued version is given in [Liu18, Proposition 7.2]. The present author and Weihua Liu have have generalized this approach to more types of convolution in [JL19, §6].

Similar to the free / monotone subordination result, there is another subordination result that relates anti-monotone and boolean convolution. Namely, given laws μ_1 and μ_2, there is another law ν such that $\mu_1 \lhd \mu_2 = \mu_1 \uplus \nu$. We will prove this one first as a "warm-up" because it is much simpler.

Proposition 6.2.1. *Let μ_1 and μ_2 be $\mathcal{B}$-valued laws. Then there is a $\mathcal{B}$-valued law ν such that $\mu_1 \lhd \mu_2 = \mu_1 \uplus \nu$.*

Proof. For $j = 1, 2$, let X_j be an operator on a $\mathcal{B}$-$\mathcal{B}$ correspondence $\mathcal{H}_j$ with $\mathcal{B}$-central unit vector ξ_j that realizes the law μ_j. Let $(\mathcal{H}, \xi)$ be the anti-monotone product of $(\mathcal{H}_1, \xi_1)$ and $(\mathcal{H}_2, \xi_2)$, that is,

$$\mathcal{H} = \mathcal{B}\xi \oplus \mathcal{H}_1^\circ \oplus \mathcal{H}_2^\circ \oplus \mathcal{H}_1^\circ \otimes_{\mathcal{B}} \mathcal{H}_2^\circ.$$

Let $\rho_{\text{mono}\,\dagger,j}$ be the map $B(\mathcal{H}_j) \to B(\mathcal{H})$ given by the anti-monotone product construction. Now let

$$\mathcal{H}_3 = \mathcal{B}\xi_3 \oplus \mathcal{H}_2^\circ \oplus \mathcal{H}_1^\circ \otimes_{\mathcal{B}} \mathcal{H}_2^\circ,$$

and note that $\mathcal{H}$ can alternatively be written as the boolean product of $(\mathcal{H}_1, \xi_1)$ and $(\mathcal{H}_3, \xi_3)$. For $j = 1, 3$, let $\rho_{\text{bool},j}$ be the inclusion map $B(\mathcal{H}_j) \to B(\mathcal{H})$ given by the boolean product.

We claim that

$$\rho_{\text{mono}\,\dagger,1}(X_1) + \rho_{\text{mono}\,\dagger,2}(X_2) = \rho_{\text{bool},1}(X_1) + \rho_{\text{bool},3}(X_3),$$

where X_3 is the operator given as follows: We may decompose $\mathcal{H}_3$ either as

$$(\mathcal{B}\xi_3 \oplus \mathcal{H}_2^\circ) \oplus (\mathcal{H}_1^\circ \otimes_{\mathcal{B}} \mathcal{H}_2^\circ) \cong \mathcal{H}_2 \oplus (\mathcal{H}_1^\circ \otimes_{\mathcal{B}} \mathcal{H}_2^\circ)$$

or as

$$\mathcal{B}\xi_3 \oplus (\mathcal{H}_2^\circ \oplus \mathcal{H}_1^\circ \otimes_{\mathcal{B}} \mathcal{H}_2^\circ) \cong \mathcal{B}\xi_3 \oplus (\mathcal{H}_1 \otimes_{\mathcal{B}} \mathcal{H}_2^\circ),$$

and then we define X_3 to be the operator $X_2 \oplus 0$ with respect to the first decomposition plus the operator $0 \oplus X_1 \otimes \text{id}$ with respect to the second decomposition.

The identity $\rho_{\text{mono}\,\dagger,1}(X_1) + \rho_{\text{mono}\,\dagger,2}(X_2) = \rho_{\text{bool},1}(X_1) + \rho_{\text{bool},3}(X_3)$ is straightforward to check. Here $\rho_{\text{mono}\,\dagger,1}(X_1)$ can be expressed as the direct sum of X_1 on $\mathcal{B}\xi \oplus \mathcal{H}_1^\circ$ and $X_1 \otimes \text{id}$ on $(\mathcal{B} \oplus \mathcal{H}_1^\circ) \otimes_{\mathcal{B}} \mathcal{H}_2^\circ$. The first piece is equal to $\rho_{\text{bool},1}(X_1)$, while the second piece of $\rho_{\text{mono}\,\dagger,1}(X_1)$ is one of the pieces of $\rho_{\text{bool},3}(X_3)$. The other piece of $\rho_{\text{bool},3}(X_3)$ is the operator X_2 acting on the direct summands $\mathcal{B}\xi \oplus \mathcal{H}_2^\circ$ of $\mathcal{H}$, which is the same as $\rho_{\text{mono}\,\dagger,2}(X_2)$.

Let ν be the law of X_3. Since $\rho_{\text{bool},1}(X_1)$ and $\rho_{\text{bool},3}(X_3)$ are boolean independent by construction, we have $\mu_1 \lhd \mu_2 = \mu_1 \uplus \nu$. $\square$

Proposition 6.2.2. *Let μ_1 and μ_2 be $\mathcal{B}$-valued laws. Then there is a $\mathcal{B}$-valued law ν such that $\mu_1 \boxplus \mu_2 = \mu_1 \rhd \nu$.*

Proof. For $j = 1, 2$, let X_j be an operator on a $\mathcal{B}$-$\mathcal{B}$ correspondence $\mathcal{H}_j$ with $\mathcal{B}$-central unit vector ξ_j that realizes the law μ_j. Let $(\mathcal{H}, \xi)$ be the free product of $(\mathcal{H}_1, \xi_1)$ and $(\mathcal{H}_2, \xi_2)$, that is,

$$\mathcal{H} = \mathcal{B}\xi \oplus \bigoplus_{n \geq 1} \bigoplus_{\substack{j_1, \dots, j_n \in \{1,2\} \\ j_k \neq j_{k+1}}} \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{H}_{j_n}^\circ.$$

Let $\rho_{\text{mono}\,\dagger,j}$ be the map $B(\mathcal{H}_j) \to B(\mathcal{H})$ given by the free product construction. Next, let

$$\mathcal{H}_3 = \mathcal{B}\xi_3 \oplus \bigoplus_{n \geq 1} \bigoplus_{\substack{j_1, \dots, j_n \in \{1,2\} \\ j_n = 2 \\ j_k \neq j_{k+1}}} \mathcal{H}_{j_1}^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{H}_{j_n}^\circ,$$

the direct sum of $\mathcal{B}\xi_3$ and all the tensor products with $\mathcal{H}_2^\circ$ as the last term. Now observe that

$$\mathcal{H} \cong \mathcal{B}\xi \oplus \mathcal{H}_1^\circ \oplus \mathcal{H}_3^\circ \oplus (\mathcal{H}_3^\circ \otimes_{\mathcal{B}} \mathcal{H}_1^\circ),$$

that is, $(\mathcal{H}, \xi)$ is isomorphic to the monotone product of $(\mathcal{H}_1, \xi_1)$ and $(\mathcal{H}_3, \xi_3)$. Let $\rho_{\text{mono},1}$ and $\rho_{\text{mono},3}$ be the corresponding maps $B(\mathcal{H}_1) \to B(\mathcal{H})$ and $B(\mathcal{H}_3) \to B(\mathcal{H})$.

We claim that

$$\rho_{\text{free},1}(X_1) + \rho_{\text{free},2}(X_2) = \rho_{\text{mono},1}(X_1) + \rho_{\text{mono},3}(X_3),$$

where X_3 is the operator given as follows: We may decompose $\mathcal{H}_3$ either as

$$(\mathcal{B} \oplus \mathcal{H}_2^\circ) \otimes_{\mathcal{B}} \left(\bigoplus_{n \geq 0} (\mathcal{H}_1^\circ \otimes_{\mathcal{B}} \mathcal{H}_2^\circ)^{\otimes_{\mathcal{B}} n} \right)$$

or as

$$\mathcal{B}\xi_3 \oplus (\mathcal{B} \oplus \mathcal{H}_1^\circ) \otimes_{\mathcal{B}} \left(\bigoplus_{n \geq 0} (\mathcal{H}_2^\circ \otimes_{\mathcal{B}} \mathcal{H}_1^\circ)^{\otimes_{\mathcal{B}} n} \otimes_{\mathcal{B}} \mathcal{H}_2^\circ \right)$$

and then we define X_3 to be the operator $X_2 \otimes \text{id}$ with respect to the first decomposition plus the operator $0 \oplus X_1 \otimes \text{id}$ with respect to the second decomposition.

We leave it to the reader to verify that $\rho_{\text{free},1}(X_1) + \rho_{\text{free},2}(X_2) = \rho_{\text{mono},1}(X_1) + \rho_{\text{mono},3}(X_3)$ because reading the explanation would not be any easier than working out the details by oneself. To conclude, let ν be the law of X_3. Since $\rho_{\text{mono},1}(X_1)$ and $\rho_{\text{mono},3}(X_3)$ are monotone independent by construction, we have $\mu_1 \boxplus \mu_2 = \mu_1 \rhd \nu$. $\square$

6.3 Reverse radius estimates

Now we use the results on analytic subordination from the previous section to get the reverse radius bounds that we could not prove before. In the proposition below, we conjecture that the constants $3 + 2\sqrt{2}$ and 2 can be replaced by 1, but we were not able to prove this in all cases. However, the values of the constants do not make any qualitative difference in our results, either here or in the rest of the paper.

Proposition 6.3.1. *Let* $\text{ind} \in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\dagger\}$*, and let* μ_1 *and* μ_2 *be* $\mathcal{B}$*-valued laws. Then*

$$\operatorname{rad}(\mu_j) \leq (3 + 2\sqrt{2})\Big(\max(\|\mu_1(X)\|, \|\mu_2(X)\|) + 2\operatorname{rad}(\mu_1 \boxplus_{\text{ind}} \mu_2)\Big).$$

Proof. First, consider the boolean case. Let $\mu = \mu_1 \uplus \mu_2$. Let (σ, b), (σ_1, b_1), and (σ_2, b_2) correspond to μ, μ_1, and μ_2 as in Theorem 4.5.3. Then $\sigma = \sigma_1 + \sigma_2$. Hence,

$$\operatorname{rad}(\sigma_1) \leq \operatorname{rad}(\sigma) \leq \operatorname{rad}(\mu)$$

and

$$\|\sigma_1(1)\|^{1/2} \leq \|\sigma(1)\|^{1/2} \leq \|\mu(X^2)\|^{1/2} \leq \operatorname{rad}(\mu).$$

Therefore,

$$\operatorname{rad}(\mu_1) \leq \|b_1\| + \operatorname{rad}(\sigma) + \|\sigma(1)\|^{1/2} \leq \|b_1\| + 2\operatorname{rad}(\mu).$$

Of course, the analogous bound holds for μ_2. This is already a better estimate than what we asserted above.

Next, consider the (anti-)monotone case. Let $\mu = \mu_1 \rhd \mu_2$. Note that $\mu = \mu_2 \uplus \nu$ where ν is the law given by Proposition 6.2.1. Therefore, $\operatorname{rad}(\mu_2) \leq \|b_2\| + 2\operatorname{rad}(\mu)$, where $b_2 = \mu_2(X)$. In order to get an estimate for $\operatorname{rad}(\mu_1)$, observe that $\tilde{G}_{\mu_1} = \tilde{G}_\mu \circ \tilde{G}_{\mu_2}^{-1}$. As we remarked in the proof of Lemma 4.5.9, it follows from the inverse function theorem that if $R \leq 1/\operatorname{rad}(\mu_2)$, then $\tilde{G}_\mu^{-1}$ maps $B(0, (3 - 2\sqrt{2})R)$ to $B(0, (1 - 1/\sqrt{2})R)$. This implies that $\tilde{G}_{\mu_1}$ is fully matricial and bounded on $B(0, (3 - 2\sqrt{2})R)$ provided that

$$R < \frac{1}{\|b_2\| + 2\operatorname{rad}(\mu)}, \qquad \left(1 - \frac{1}{\sqrt{2}}\right) R < \frac{1}{\operatorname{rad}(\mu)}.$$

The first condition is strictly stronger than the second. Therefore,

$$\begin{aligned}\operatorname{rad}(\mu_1) &\leq \frac{1}{3 - 2\sqrt{2}} \left(\|b_2\| + 2\operatorname{rad}(\mu)\right) \\ &= (3 + 2\sqrt{2}) \left(\|b_2\| + 2\operatorname{rad}(\mu)\right).\end{aligned}$$

This finishes the (anti-)monotone case.

The free case follows because $\mu_1 \boxplus \mu_2 = \mu_1 \rhd \nu_2 = \mu_2 \rhd \nu_1$ for some laws ν_1, ν_2 given by Proposition 6.2.2. □

CHAPTER 7

Results: Evolution equations for subordination families

7.1 Introduction

Our main goal in this chapter is to describe how the F-transforms F_{X_t} evolve over time when $(X_t)_{t\in[0,T]}$ is a process with independent increments for each of the four types of independence described in §5. The main result will be roughly speaking that if μ_t is the law of X_t, then F_{μ_t} satisfies the equation

$$\partial_t F_{\mu_t}^{(n)}(z) = \begin{cases} -[b(t) + G_{\sigma(\cdot,t)}(z)], & \text{boolean case,} \\ -DF_{\mu_t}^{(n)}(z)[b(t) + G_{\sigma(\cdot,t)}^{(n)}(F_{\mu_t}^{(n)}(z))], & \text{free case,} \\ -DF_{\mu_t}^{(n)}(z)[b(t) + G_{\sigma(\cdot,t)}^{(n)}(z)], & \text{monotone case,} \\ -[b(t) + G_{\sigma(\cdot,t)}^{(n)}(F_{\mu_t}^{(n)}(z))], & \text{anti-monotone case,} \end{cases} \tag{7.1}$$

where $G_{\sigma(\cdot,t)}$ is the Cauchy-Stieltjes transform of a generalized law $\sigma(\cdot,t)$ depending on t, and where $DF_{\mu_t}(z) = \Delta F_{\mu_t}(z,z)$.

The evolution of F_{X_t} (or equivalently of G_{X_t}) has been studied in many previous papers in special cases. The first case to be worked out for each type of independence was when $(X_t)_{t\in[0,T]}$ has independent and *stationary* increments (that is, $X_t - X_s \sim X_{t-s}$ in law), or equivalently the laws $(\mu_t)_{t\in[0,T]}$ form a convolution semigroup. Prior work on the differential equations associated to such semigroups is summarized in Table 7.1. See also the discussion of the Lévy-Hinčin formula in §1.3 and the subsequent discussion of semigroups in §9.1.

To address the differential equations for processes with non-stationary increments in the operator-valued setting, we must deal with the technicalities of differentiation for Banach-valued functions in order to even make sense of the equation. These difficulties do not arise in the scalar-valued setting because scalar-valued absolutely continuous functions are

	boolean	free	(anti-)monotone
scalar-valued	[SW97, Thm. 3.6]	[Voi86, Thm. 4.3] [Bia98]	[Has10a] [Has10b] [HS14] [AW14]
operator-valued / multivariable	[BN08] [BPV12] [PV13, §2]	[Spe98, §4.5 - 4.7] [BPV12] [PV13, §3]	[AW16] [Jek20]

Table 7.1: References on non-commutative convolution semigroups.

always differentiable almost everywhere. And in the operator-valued setting, if we have the additional symmetry of stationary increments, the differentiability of F_{X_t} in t can be established by direct arguments with inverse function theorems and/or iteration, as for instance in [AW16, Proposition 3.3]. However, in the non-stationary operator-valued setting (even with the assumption of bounded support), we inevitably run into the issue that not every absolutely continuous function from $[0,T]$ into a Banach space can be differentiated almost everywhere, even in the weak or weak-$*$-topology. But we will circumvent this problem by instead treating the time-derivatives as operator-valued distributions on $[0,T]$.

The present author studied the monotone case of $\mathcal{B}$-valued Lipschitz subordination families with bounded support in [Jek20] (the free and boolean cases being easier to understand by previously existing techniques), and this chapter uses many of the same material as in that paper. We first discuss some preliminary definitions and observations about processes with independent increments. Then we describe the properties of the derivatives of Lipschitz functions from $[0,T]$ into Banach spaces, and of Lipschitz families of fully matricial functions. Finally, using these tools, we show that the transforms F_{X_t} for a process X_t with independent increments (and some Lipschitz conditions in time) will satisfy the equation above for some σ.

7.1.1 Processes and subordination families

Definition 7.1.1. Let ind $\in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\,\dagger\}$. A *process with $\mathcal{B}$-valued* ind-*independent increments on* $[0,T]$ is a collection of non-commutative self-adjoint operators $(X_t)_{t\in[0,T]}$ in $\mathcal{B}$-valued probability space $(\mathcal{A}, E)$ such that for every $0 = t_0 < t_1 < \dots < t_N = T$, the operators X_{t_0}, $X_{t_1} - X_{t_0}$, ..., $X_{t_N} - X_{t_{N-1}}$ are ind-independent over $\mathcal{B}$.

Another viewpoint on the same idea is to look at the non-commutative law μ_t of X_t rather than the operator itself. This leads to the following definition.

Definition 7.1.2. Let ind $\in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\,\dagger\}$. An ind-*subordination family on* $[0,T]$ is a collection of non-commutative laws $(\mu_t)_{t\in[0,T]}$ such that for each $0 \le s \le t \le T$, there exists another non-commutative law $\mu_{s,t}$ such that $\mu_t = \mu_s \boxplus_{\text{ind}} \mu_{s,t}$.

We make the following observations:

- If $(X_t)_{t\in[0,T]}$ is a process with independent increments, and if μ_t is the law of X_t, then $(\mu_t)_{t\in[0,T]}$ is a subordination family because we can take $\mu_{s,t}$ to be the law of $X_t - X_s$.
- Suppose that $(\mu_t)_{t\in[0,T]}$ is an ind-subordination family. Then there is only one possible choice of $\mu_{s,t}$ satisfying $\mu_t = \mu_s \boxplus_{\text{ind}} \mu_{s,t}$. This is because the analytic transforms of $\mu_{s,t}$

must satisfy the equations

$$\begin{aligned} K_{\mu_t} &= K_{\mu_s} + K_{\mu_{s,t}}, && \text{boolean case,} \\ \Phi_{\mu_t} &= \Phi_{\mu_s} + \Phi_{\mu_{s,t}}, && \text{free case,} \\ F_{\mu_t} &= F_{\mu_s} \circ F_{\mu_{s,t}}, && \text{monotone case,} \\ F_{\mu_t} &= F_{\mu_{s,t}} \circ F_{\mu_s}, && \text{anti-monotone case,} \end{aligned}$$

and $\mu_{s,t}$ is uniquely determined by knowing $K_{\mu_{s,t}}$, $\Phi_{\mu_{s,t}}$, or $F_{\mu_{s,t}}$ in a neighborhood of ∞.

- Again, let $(\mu_t)_{t\in[0,T]}$ be a subordination family. Using associativity of convolution, we have for $s \le t \le u$ that $\mu_u = \mu_s \boxplus_{\text{ind}} (\mu_{s,t} \boxplus_{\text{ind}} \mu_{t,u})$, and therefore it follows that $\mu_{s,u} = \mu_{s,t} \boxplus_{\text{ind}} \mu_{t,u}$ by the previous claim about uniqueness of $\mu_{s,u}$.
- A desirable property for a subordination chain would be that the $\operatorname{rad}(\mu_t)$ is uniformly bounded for $t \in [0,T]$. However, this is automatic; it follows from Proposition 6.3.1 that
$$\sup_{t\in[0,T]} \operatorname{rad}(\mu_t) \le (3+2\sqrt{2}) \left(2\operatorname{rad}(\mu_T) + \sup_{t\in[0,T]} \|\mu_t(X)\| \right).$$

Remark 7.1.3. It is not difficult to show that any subordination family arises from a process with independent increments. Indeed, if we consider finitely many times $0 = t_0 < \cdots < t_N$, then we can construct independent variables Y_{t_0}, Y_{t_0,t_1}, $\dots$, Y_{t_{N-1},t_N} and set $Y_{t_j} = Y_{t_0} + Y_{t_0,t_1} + \cdots + Y_{t_{N-1},t_N}$. Then Y_{t_0}, $\dots$, Y_{t_N} are a family of variables indexed by $\{t_0,\dots,t_N\}$ with independent increments. We can do this for any finite family of times. It remains to show that all the finite-time marginals can be realized simultaneously by the same process. One can reduce this claim to the case where $\mathcal{B}$ is a von Neumann algebra. Then by using compactness in the pointwise WOT of the space of laws of processes satisfying $\|X_t\| \le C$, one can conclude that there is a family $(X_t)_{t\in[0,T]}$ that realizes each of these finite-time marginals simultaneously. In the last argument, the only challenge is to get a uniform bound on the operator norm Y_t over all partitions $\{t_0,\dots,t_N\}$ that contain t, in order to obtain the existence of *bounded* operators $(X_t)_{t\in[0,T]}$. This can be done by a careful use of our operator-norm bounds Lemma 6.1.1. However, we will not carry out this argument in detail because we will discuss a more enlightening systematic construction of processes with independent increments in the next chapter, under some continuity assumptions.

7.1.2 Setup and conditions for differentiation

Consider a subordination family $(\mu_t)_{t\in[0,T]}$. Under what reasonably general conditions can we expect to be able to differentiate F_{μ_t} with respect to t? We know that $\sup_t \operatorname{rad}(\mu_t) < +\infty$, so it would be natural for $\tilde{G}_{\mu_t}$ to also be differentiable with respect to t in a neighborhood of zero, which implies that each of the moments of μ_t should be differentiable. Thus, we should at least guarantee that the mean $\mu_t(X)$ and the variance at 1 given by $\text{Var}(\mu_t)[1] = \mu_t((X - \mu_t(X))^2)$ are differentiable with respect to t, and in fact this will turn out to be sufficient.

Next, under what conditions can we differentiate the maps $t \mapsto \mu_t(X)$ and $t \mapsto \mu_t(X^2)$? We should require that they are *absolutely continuous* as maps from $[0, T]$ into the Banach space $\mathcal{B}$. That is, for every $\epsilon > 0$, there exists $\delta > 0$ such that if $\{[a_i, b_i)\}_{i=1}^N$ are disjoint intervals in $[0, T]$ with $\sum_i (b_i - a_i) < \delta$, then $\sum_i \|\mu_{b_i}(X) - \mu_{a_i}(X)\| < \epsilon$, and the same for $\mu_t((X - \mu_t(X))^2)$ rather than $\mu_t(X)$. Now if we let $\phi(t)$ and $\psi(t)$ be the total variation of $t \mapsto \mu_t(X)$ and $t \mapsto \mu_t((X - \mu_t(X))^2)$ respectively, then we can reparametrize time using the inverse function of $f(t) = \phi(t) + \psi(t) + t$. Letting $\nu_{f^{-1}(t)}$, then we have $\|\nu_t(X) - \nu_s(X)\| \leq C|t - s|$ for some constant C, and the same is true for $\nu_t((X - \nu_t(X))^2)$.

Therefore, if we want to study subordination families where $\mu_t(X)$ and $\mu_t((X - \mu_t(X)))$ are absolutely continuous, then without loss of generality, we can restrict our attention to the case where they are Lipschitz in t. Of course, for many concrete examples of subordination families, the mean and variance might be of the form $b_0 + tb$ for some $b_0, b \in \mathcal{B}$, and the Lipschitz assumption obviously holds in such cases. Thus, we make the following definition.

Definition 7.1.4. Let $(\mu_t)_{t \in [0,T]}$ be a subordination family with respect to boolean, free, monotone, or anti-monotone independence. We say that $(\mu_t)_{t \in [0,T]}$ is *Lipschitz* if $t \mapsto \mu_t(X)$ and $t \mapsto \mu_t((X - \mu_t(X))^2)$ are Lipschitz on $[0, T]$.

However, even in the Lipschitz case, we run into technical issues with differentiation. Our solution will ultimately be to avoid pointwise differentiation altogether using a distributional theory presented in the next section. As motivation, we will first explain why pointwise differentiation is hopeless in the level of generality we are aiming for, where we do not assume F_{μ_t} is C^1 in t and where $\mathcal{B}$ is allowed to be a general C*-algebra.

A Lipschitz function from $[0, T]$ into a Banach space $\mathcal{X}$ may not be differentiable almost everywhere with respect to the norm on $\mathcal{X}$ or even with respect to the weak topology, or the weak-$*$ topology if $\mathcal{X}$ happens to be a dual space. Known results about differentiating Banach valued functions (see e.g. [Kom67, Appendix]) rely on either separability or reflexitvity of $\mathcal{X}$, which is something we cannot assume in an operator algebras setting. Indeed, infinite-dimensional C*-algebras are never reflexive, and furthermore, infinite-dimensional von Neumann algebras are never separable in the norm topology.

Pointwise differentiation will certainly not be possible in the norm topology. If $\mathcal{A}$ is a von Neumann algebra acting on a separable Hilbert space, then differentiation in the strong operator topology (SOT) may be possible (thanks to the theory of differentiation of Hilbert-valued functions). However, in order to use the chain rule and similar manipulations for SOT differentiation (which we will have to do here since we must consider composition of time-dependent functions), we would have to make the additional assumption that the Fréchet derivatives of the maps we are composing are SOT-continuous, which means making additional SOT continuity assumptions about the laws μ_t that are not automatic.

Furthermore, suppose that we *can* for a fixed z, differentiate $F(z,t)$ almost everywhere with respect to t; then it would still be problematic to carry out such differentiation with the same exceptional null set of times for all values of z ranging over an open set in a non-separable Banach space. One might try to solve this problem by assuming that $\mathcal{A}$ is separable in SOT and that our analytic functions are continuous in SOT. However, even this is not sufficient because we cannot enforce SOT equicontinuity of $(F(z,t+\delta) - F(z,t))/\delta$ as $\delta \to 0$.

Another possible idea would be to assume that $\mathcal{A}$ is tracial von Neumann algebra and that for each function $F(z) = z - G_\sigma(z)$ that we are dealing with, the state $\tau \circ \sigma$ is tracial on $\mathcal{A}\langle X \rangle$. The problems with the SOT approach sketched above would be solved by using explicit estimates in L^2 norm to guarantee SOT-equicontinuity of $(F(z,t+\delta) - F(z,t))/\delta$ for different values of δ, as well as SOT equicontinuity of $a \mapsto DF_t(z)[a]$ for different values of t. However, traciality of σ seems like an artificial and restrictive condition. If $F_\mu(z) = z - G_\sigma(z)$, it is unclear (at least to the author) whether traciality of σ and traciality of μ are related.

For free independence, we at least know that if μ and ν can be realized by variables in a tracial von Neumann algebra, then so can $\mu \boxplus \nu$. And perhaps tracial von Neumann algebras are a good place to start developing the theory of non-commutative laws for unbounded operator-valued random variables. However, that is not the goal of the present work.

7.2 Distributional derivatives

Since differentiation of F_{μ_t} in a pointwise sense is not feasible, we will consider the time-derivative in a distributional sense. Thus, this section examines distributional derivatives of Lipschitz functions $\gamma : [0,T] \to \mathcal{X}$, where $\mathcal{X}$ is some Banach space. The good news is that this distributional derivative is not too badly behaved. It makes sense to pair with functions in $L^1[0,T]$, and it "almost as good as an L^∞ function" in that we can still in many ways manipulate it as if it were a pointwise defined function. More precisely, we will describe how to perform various "pointwise" operations with elements of $\mathcal{L}(L^1[0,T], \mathcal{X})$, including nonlinear operations involving composition.

As motivation, recall that if $\gamma : [0,T] \to \mathbb{C}$ is Lipschitz, then the distributional derivative $\dot{\gamma} : C_c^\infty(0,T) \to \mathbb{C}$ is represented by a function in $L^\infty(0,T) = L^1(0,T)^*$. For functions into a Banach space, there are several analogues of L^∞ functions, including the *Bochner* L^∞ *space* $L^\infty_{\text{Boch}}([0,T], \mathcal{X})$, defined as follows: Consider countably-valued simple functions $\phi(t) = \sum_{j=1}^\infty x_j \chi_{E_j}(t)$, where $(E_j)_{j \in \mathbb{N}}$ are disjoint measurable subsets of $[0,T]$ with positive measure and $\|\phi\|_{L^\infty_{\text{Boch}}[0,T]} := \sup_j \|x_j\|$ is finite; then $L^\infty_{\text{Boch}}[0,T]$ is the completion of this space with respect to $\|\cdot\|_{L^\infty_{\text{Boch}}[0,T]}$.

If $\gamma : [0,T] \to \mathcal{X}$ is Lipschitz, then the distributional derivative $\dot{\gamma} : C_c^\infty(0,T) \to \mathcal{X}$ is *not* necessarily represented by a function in $L^\infty_{\text{Boch}}([0,T], \mathcal{X})$. However, we claim that $\dot{\gamma}$ does extend to a bounded map $L^1[0,T] \to \mathcal{X}$. In the following, we denote by $\mathcal{L}(L^1[0,T], \mathcal{X})$ the space of bounded linear maps $L^1[0,T] \to \mathcal{X}$.

Observation 7.2.1. *If $\gamma : [0,T] \to \mathcal{X}$ is Lipschitz, then there exists a unique $\dot\gamma \in \mathcal{L}(L^1[0,T],\mathcal{X})$ satisfying*

$$\dot\gamma[\chi_{[a,b]}] = \gamma(b) - \gamma(a). \tag{7.2}$$

Conversely, if $\rho \in \mathcal{L}(L^1[0,T],\mathcal{X})$, then the function

$$\gamma(t) = \rho[\chi_{[0,t]}] \tag{7.3}$$

is Lipschitz and satisfies $\dot\gamma = \rho$. Also, $\|\dot\gamma\|_{\mathcal{L}(L^1,\mathcal{X})}$ equals the Lipschitz seminorm of γ.

Proof. Suppose $\gamma : [0,T] \to \mathcal{X}$ is C-Lipschitz. The action of $\dot\gamma$ on step functions is defined by $\dot\gamma[\chi_{[a,b]}] = \gamma(b) - \gamma(a)$. For any step function ϕ, we have $\|\dot\gamma[\phi]\| \le C\|\phi\|_{L^1[0,T]}$, hence the $\dot\gamma$ extends to bounded linear map $L^1[0,T] \to \mathcal{X}$. The other claims are left as exercises. □

The following fact will be handy for proving identities and estimates involving distributional derivatives.

Lemma 7.2.2. *If $\rho \in \mathcal{L}(L^1[0,T],\mathcal{X})$, then*

$$\|\rho\|_{\mathcal{L}(L^1,\mathcal{X})} = \sup_{0\le a<b\le T} \frac{\|\rho[\chi_{a,b}]\|}{b-a} = \lim_{\epsilon\to 0} \sup_{0<b-a\le\epsilon} \frac{\|\rho[\chi_{a,b}]\|}{b-a}. \tag{7.4}$$

As a consequence, if ρ and $\tilde\rho$ are bounded maps $L^1[0,T] \to \mathcal{X}$ and $\rho[\chi_{[a,b]}] = \tilde\rho[\chi_{[a,b]}] + o(|b-a|)$, then $\rho = \tilde\rho$.

Proof. The nontrivial part of the proof is to show that

$$\|\rho\|_{\mathcal{L}(L^1,\mathcal{X})} \le \liminf_{\epsilon\to 0} \sup_{0<b-a\le\epsilon} \frac{\|\rho[\chi_{a,b}]\|}{b-a}. \tag{7.5}$$

If C is the right hand side, then it is sufficient to show that $\|\rho[\phi]\| \le C\|\phi\|_{L^1[0,T]}$ when ϕ is continuous. This can be proved by approximating ϕ uniformly by a sequence of step functions, such that mesh size of the partition also approaches zero. □

Remark 7.2.3. Note that by the previous lemma and some basic results on L^∞_{Boch}, there is an isometric inclusion $\iota : L^\infty_{\text{Boch}}([0,T],\mathcal{X}) \to \mathcal{L}(L^1[0,T],\mathcal{X})$ given by

$$\iota(\rho) : \phi \mapsto \int_0^T \rho(t)\phi(t)\,dt,$$

for $\rho \in L^\infty_{\text{Boch}}([0,T],\mathcal{X})$, so in the sequel we will regard $L^\infty_{\text{Boch}}([0,T],\mathcal{X})$ as a subspace of $\mathcal{L}(L^1[0,T],\mathcal{X})$.

If we had a bounded function $R : [0,T] \times [0,T] \to \mathcal{X}$ denoted $R(s,t)$, then of course we could define the diagonal restriction $R(t,t)$. We claim that under appropriate hypotheses, this operation still makes sense when $R(s,\cdot)$ is an element of $\mathcal{L}(L^1[0,T],\mathcal{X})$ rather than a bounded function $[0,T] \to \mathcal{X}$. For this to be rigorous, we must view R as a map $[0,T] \to \mathcal{L}(L^1[0,T],\mathcal{X})$.

Lemma 7.2.4 (Diagonal restriction)**.** *There exists a unique linear map*

$$\mathrm{diag} : L^\infty_{\mathrm{Boch}}([0,T], \mathcal{L}(L^1[0,T], \mathcal{X})) \to \mathcal{L}(L^1[0,T], \mathcal{X})$$

such that

(1) If $R(s) = \sum_{j=1}^\infty \chi_{E_j}(s) \cdot \rho_j$ *where the sets* E_j *are disjoint measurable sets and*

$$\sup_j \|\rho_j\|_{\mathcal{L}(L^1[0,T],\mathcal{X})} < \infty,$$

and if $\phi \in L^1[0,T]$*, we have*

$$(\mathrm{diag}\, R)[\phi] = \sum_{j=1}^\infty \rho_j[\chi_{E_j}\phi]. \tag{7.6}$$

(2) We have

$$\|\mathrm{diag}\, R\|_{\mathcal{L}(L^1[0,T],\mathcal{X})} \leq \|R\|_{L^\infty_{\mathrm{Boch}}([0,T],\mathcal{L}(L^1[0,T],\mathcal{X}))}. \tag{7.7}$$

Furthermore, this map diag *satisfies the estimate*

$$\|(\mathrm{diag}\, R)[\phi]\| \leq \int_0^T |\phi(t)| \|R(t,\cdot)\|_{\mathcal{L}(L^1[0,T],\mathcal{X})}\, dt. \tag{7.8}$$

Proof. For a simple function R, we can define $\mathrm{diag}\, R$ unambiguously by (7.6), that is, it is independent of the decomposition of the simple function. We check that (7.8) and hence (7.7) hold for simple functions. Then (7.7) implies that $\mathrm{diag}\, R$ has a unique extension to $L^\infty_{\mathrm{Boch}}([0,T], \mathcal{L}(L^1[0,T], \mathcal{X}))$. The inequality (7.8) extends to $L^\infty_{\mathrm{Boch}}([0,T], \mathcal{L}(L^1[0,T], \mathcal{X}))$ because both sides are continuous in the Bochner L^∞ norm. □

In the rest of the paper, we will often use more suggestive notation which treats the elements of $\mathcal{L}(L^1[0,T], \mathcal{X})$ like pointwise defined functions. Although using function notation for distributions has some drawbacks, the ultimate benefit will be a more intuitive statement of identities such Lemma 7.3.4 below, and more generally a compact notation for constructing and transforming such distributions.

Notation 7.2.5. For a function $\rho \in \mathcal{L}(L^1[0,T], \mathcal{X})$, we will use the notation $\rho(t)$ where t is formal or "dummy" variable. For $\phi \in L^1[0,T]$, we define

$$\int_0^T \phi(t)\rho(t)\, dt := \rho[\phi] \tag{7.9}$$

as well as

$$\int_a^b \rho(t)\, dt := \rho[\chi_{[a,b]}]. \tag{7.10}$$

To obviate potential confusion, when we apply ρ as a linear map to a function ϕ in $L^1[0,T]$, we will use square brackets and not write the dummy variable t. For instance, the application of ρ to the identity function t on $[0,T]$ would be denoted by $\rho[\mathrm{id}_{[0,T]}]$ or $\int_0^T \rho(t)t\,dt$ and *not* by $\rho(t)$ or $\rho[t]$. Similarly, $\rho(2t)$ would denote the element of $\mathcal{L}(L^1[0,T/2],\mathcal{X})$ defined by

$$\int_0^{T/2} \rho(2t)\phi(t)\,dt := \frac{1}{2}\int_0^T \rho(t)\phi(2t)\,dt$$

but on the other hand $\int \rho(t)\cdot 2t\,dt$ would denote the application of ρ as a linear map to the function $2t$ on $[0,T]$.

Notation 7.2.6. If R is in $L^\infty_{\mathrm{Boch}}([0,T],\mathcal{L}(L^1[0,T],\mathcal{X}))$, then we will write R formally as a function of two variables (s,t), where the s corresponds to the first "$[0,T]$" and the t corresponds to the second "$[0,T]$" in "$L^\infty_{\mathrm{Boch}}([0,T],\mathcal{L}(L^1[0,T],\mathcal{X}))$"; in other words, the distributional dependence occurs in the second variable t. We will denote $(\operatorname{diag} R)(t)$ by $R(t,t)$.

Thus, for example, if $R(s,t) = \sum_{j=1}^\infty \chi_{E_j}(s)\rho_j(t)$, then (7.6) becomes

$$\int_0^T \phi(t)R(t,t)\,dt = \sum_{j=1}^\infty \int_{E_j} \phi(t)\rho_j(t)\,dt, \tag{7.11}$$

and hence in a formal sense

$$R(t,t) = \sum_{j=1}^\infty \chi_{E_j}(t)\rho_j(t). \tag{7.12}$$

Also, (7.8) becomes

$$\left\|\int_0^T \phi(t)R(t,t)\,dt\right\| \le \int_0^T |\phi(t)|\|R(t,\cdot)\|_{\mathcal{L}(L^1[0,T],\mathcal{X})}\,dt. \tag{7.13}$$

We will mainly use two special cases of the diagonal restriction.

Definition 7.2.7. Suppose that $\rho \in \mathcal{L}(L^1[0,T],\mathcal{X})$ and $A \in L^\infty_{\mathrm{Boch}}([0,T],\mathcal{L}(\mathcal{X},\mathcal{Y}))$. Then we define $(A\rho)(t) = A(t)\rho(t)$ in $\mathcal{L}(L^1[0,T],\mathcal{Y})$ as the diagonal restriction $R(t,t)$ of the function $R(s,t) = A(s)\rho(t)$, which is in $L^\infty_{\mathrm{Boch}}([0,T],\mathcal{L}(L^1[0,T],\mathcal{Y}))$.

Observation 7.2.8.

(1) The product $A\cdot\rho$ defined above is bilinear in A and ρ.

(2) If I is a subinterval of $[0,T]$, then we have $A|_I \cdot \rho|_I = (A\rho)|_I$.

(3) $\|A\cdot\rho\|_{\mathcal{L}(L^1[0,T],\mathcal{Y})} \le \|A\|_{L^\infty_{\mathrm{Boch}}([0,T],\mathcal{L}(\mathcal{X},\mathcal{Y}))}\|\rho\|_{\mathcal{L}(L^1[0,T],\mathcal{X})}$.

Definition 7.2.9. Suppose that W is a metric space, $F(w,t)$ is a continuous map $W \to \mathcal{L}(L^1[0,T],\mathcal{X})$, and $w : [0,T] \to W$ is continuous. Then $R(s,t) = F(w(s),t)$ is a continuous map $[0,T] \to \mathcal{L}(L^1[0,T],\mathcal{X})$. We define $F(w(t),t)$ to be the diagonal restriction of R.

Observation 7.2.10. *Suppose that F is uniformly continuous as a map $W \to \mathcal{L}(L^1[0,T],\mathcal{X})$ with modulus of continuity ω_F, and let d_∞ be the supremum metric on $C([0,T],W)$. Then for $w, \tilde{w} \in C([0,T],W)$, we have*

$$\begin{aligned}\|F(w(t),t) - F(\tilde{w}(t),t)\|_{\mathcal{L}(L^1[0,T],\mathcal{X})_{dt}} &\leq \|F\circ w - F\circ \tilde{w}\|_{L^\infty_{\mathrm{Boch}}([0,T],\mathcal{L}(L^1[0,T],\mathcal{X}))} \\ &\leq \omega_F(d_\infty(w,\tilde{w})).\end{aligned}$$

7.3 Locally Lipschitz families of fully matricial Functions

7.3.1 Definition and properties

Next, we consider functions $F(z,t)$ that are fully matricial in z and locally Lipschitz in t, which of course includes the families of F-transforms in the main theorem for the chapter.

Definition 7.3.1 (Locally Lipschitz Family)**.** Let Ω_1 and Ω_2 be fully matricial domains over C*-algebras $\mathcal{B}_1$ and $\mathcal{B}_2$ respectively and let $T > 0$. A collection of functions $(F^{(n)})_{n\in\mathbb{N}}$ where $F^{(n)} : \Omega_1^{(n)} \times [0,T] \to \Omega_2^{(n)}$ is called a *locally Lipschitz family of fully matricial functions* if $F(\cdot,t) := (F^{(n)}(\cdot,t))_{n\in\mathbb{N}}$ is fully matricial for each t, and for each $z_0 \in \Omega_1$ there exist $r > 0$ and $C > 0$ such that $B(z,r) \subseteq \Omega_1$ and

$$\|F(z,s) - F(z,t)\| \leq C|s-t| \text{ for all } s,t \in [0,T] \text{ for all } z \in B(z,r). \tag{7.14}$$

Here the word "locally" refers to the variable z but "Lipschitz" refers to the variable t, fully matricial functions being automatically locally Lipschitz in the space variable by Corollary 3.5.9. Moreover, if the codomain Ω_2 is not specified, we assume that it is $M_\bullet(\mathcal{B}_2)$. For each $z \in \Omega_1^{(n)}$, the function $F^{(n)}(z,\cdot)$ is Lipschitz, and so $\partial_t F^{(n)}(z,\cdot) \in \mathcal{L}(L^1[0,T],M_n(\mathcal{B}))$. This time derivative is the following type of object.

Definition 7.3.2 (Distributional family)**.** Let Ω_1 be a fully matricial domain over $\mathcal{B}_1$, let $\mathcal{B}_2$ be a C*-algebra, and let $T > 0$. A *distributional family of fully matricial functions* $\Omega_1 \to M_\bullet(\mathcal{B}_2)$ is a collection of maps $f = (f^{(n)})_{n\in\mathbb{N}}$ where $f^{(n)} : \Omega_1^{(n)} \times L^1[0,T] \to M_n(\mathcal{B}_2)$, such that

(1) For each $z \in \Omega_1^{(n)}$, the map $f^{(n)}(z,\cdot)$ is in $\mathcal{L}(L^1[0,T],M_n(\mathcal{B}_2))$.

(2) For each $z_0 \in \Omega_1^{(n)}$, there exist $r > 0$ and $C > 0$ such that

$$\|f(z,\cdot)\|_{\mathcal{L}(L^1[0,T],M_{nk}(\mathcal{B}_2))} \leq C \text{ for all } z \in B^{(nk)}(z^{(k)},r) \text{ for all } k \in \mathbb{N}.$$

(3) For each $\phi \in L^1[0,T]$, the collection of functions $H(z) = \int_0^T f(\cdot,t)\phi(t)\,dt$ given by $H^{(n)}(z) = \int_0^T f^{(n)}(z,t)\phi(t)\,dt$ is fully matricial.

Lemma 7.3.3 (Differentiation and Integration)**.**

(1) If F is a locally Lipschitz family of fully matricial functions $\Omega_1 \to M_\bullet(\mathcal{B}_2)$, then $\partial_t F$ is a distributional family of fully matricial functions $\Omega_1 \to M_\bullet(\mathcal{B}_2)$.

(2) Conversely, if f is a distributional family of fully matricial functions, then we can define a locally Lipschitz family of fully matricial functions by

$$F^{(n)}(z,t) = \int_0^t f^{(n)}(z,s)\,ds.$$

Proof. (1) For each $z \in \Omega_1^{(n)}$, since $F^{(n)}(z,t)$ is Lipschitz, we have $f^{(n)}(z,\cdot) := \partial_t F^{(n)}(z,\cdot) \in \mathcal{L}(L^1[0,T], M_n(\mathcal{B}))$. Now pick $z_0 \in \Omega_1^{(n)}$, and let C and r be as in (7.14), then we have for $0 \leq s \leq t \leq T$ and $z \in B^{(nk)}(z_0^{(k)}, r)$ that

$$\left\| \int_s^t f^{(nk)}(z,u)\,du \right\| \leq C|s-t|,$$

and by Lemma 7.2.2, we get that $\|f^{(nk)}(z,\cdot)\|_{\mathcal{L}(L^1[0,T],M_{nk}(\mathcal{B}_2))} \leq C$. Next, we must show that for each $\phi \in L^1[0,T]$, the collection of functions $H(z) = \int_0^T f(z,t)\phi(t)\,dt$ is fully matricial. Clearly this is true when $\phi = \chi_{[s,t]}$, hence it holds when ϕ is a step function, and then it holds for all $\phi \in L^1[0,T]$ by approximation (given our uniform a priori bounds on $\|f^{(n)}(z,\cdot)\|_{\mathcal{L}(L^1[0,T],M_n(\mathcal{B}_2))}$).

The verification of (2) is similarly straightforward and is left to the reader (see Observation 7.2.1). □

If $F : \Omega_1 \times [0,T] \to M_\bullet(\mathcal{B}_2)$ is a locally Lipschitz family, then $\Delta^k F$ also satisfies a local Lipschitz condition. Given $z_0 \in M_{n_0}(\mathcal{B}_1)$, ..., $z_k \in M_{n_k}(\mathcal{B}_1)$, we set $Z = (z_0, \dots, z_k)$ and choose a C and r as in (7.14) for this Z. Then by Lemma 3.4.2,

$$\|\Delta^k F(z_0,\dots,z_k,s) - \Delta^k F(z_0,\dots,z_k,t)\|_\# \leq \frac{C|s-t|}{r^k}.$$

Therefore, $\partial_t \Delta^k F(z_0,\dots,z_k,\cdot)$ is well-defined as a bounded linear map from $L^1[0,T]$ to completely bounded multilinear maps $M_{n_0,n_1}(\mathcal{B}_1) \times \cdots \times M_{n_{k-1},n_k}(\mathcal{B}_1) \to M_{n_0,n_k}(\mathcal{B}_2)$, with its norm bounded by $C|s-t|/r^k$. We also have that for $\phi \in L^1[0,T]$

$$\Delta^k \left(\int_0^T \partial_t F(\cdot,t)\phi(t)\,dt \right)(z_0,\dots,z_k) = \int_0^T \partial_t \Delta^k F(z_0,\dots,z_k,t)\phi(t)\,dt,$$

which is proved by checking it for $\phi = \chi_{[s,t]}$ and then approximating an arbitrary ϕ in $L^1[0,T]$ by step functions. This is the correct interpretation in our context of equality of mixed partials "$\partial_t \Delta^k F = \Delta^k \partial_t F$."

7.3.2 Chain rule

Furthermore, we have the following version of the chain rule for computing $\partial_t[F(G(z,t),t)]$. Although we do not directly cite the chain rule in proving our main theorem, the proof of the chain rule is a prototype for several of our arguments. The fact that the chain rule holds also validates our approach of using distributional differentiation for locally Lipschitz families. In the statement below, we use the notation $DF(z) := \Delta F(z,z)$ when F is a fully matricial function.

Lemma 7.3.4 (Chain Rule)**.** *Let Ω_1, Ω_2, and Ω_3 be fully matricial domains over C^*-algebras $\mathcal{B}_1$, $\mathcal{B}_2$, and $\mathcal{B}_3$ respectively. Let $F : \Omega_1 \times [0,T] \to \Omega_2$ and $G : \Omega_2 \times [0,T] \to \Omega_3$ be locally Lipschitz families of fully matricial functions. Then $F(G(z,t),t)$ is also a locally Lipschitz family of fully matricial functions. Moreover,*

$$\partial_t[F^{(n)}(G^{(n)}(z,t),t))] = DF^{(n)}(G^{(n)}(z,t),t)[\partial_t G^{(n)}(z,t)] + \partial_t F^{(n)}(G^{(n)}(z,t),t). \tag{7.15}$$

Here $DF^{(n)}(G^{(n)}(z,t),t)[\partial_t G^{(n)}(z,t)]$ is given by Definition 7.2.7 with $\rho(t) = \partial_t F^{(n)}(z,t)$ and $A(t) = DF^{(n)}(G^{(n)}(z,t),t)$. The other term $\partial_t F^{(n)}(G^{(n)}(z,t),t)$ is given by Definition 7.2.9 by taking W to be an appropriate open subset of $\Omega_2^{(n)}$ and setting $w(t) = G(z,t)$.

Proof. Because a fully matricial function is assumed to be locally bounded and because of the local Lipschitz estimate (7.14), we see that for each $w_0 \in \Omega_1^{(n)}$, there exists $r > 0$ and $M > 0$ such that

$$\|F(w,t)\| \leq M \text{ for all } t \in [0,T] \text{ and } w \in B(w_0,r).$$

Then because of Corollary 3.5.9, we have

$$\|F^{(nk)}(w,t) - F^{(nk)}(w',t)\| \leq \frac{2M}{r}\|w - w'\| \text{ for all } z,z' \in B^{(nk)}(w_0^{(k)}, r/2) \text{ and } t \in [0,T].$$

Together with the local Lipschitz estimate (7.14), this implies that for each $w_0 \in \Omega_1^{(n)}$, there exists some $r(w_0) > 0$, $M(w_0) > 0$, and $C(w_0) > 0$ such that

$$\|F^{(kn)}(w,t) - F^{(nk)}(w',t')\| \leq M(w_0)\|w - w'\| + C(w_0)|t - t'|$$

for all $w, w' \in B^{(nk)}(w_0^{(k)}, r(w_0))$ and $t \in [0,T]$, for all $k \in \mathbb{N}$.

Now fix $z_0 \in \Omega_1^{(n)}$. Since G is locally Lipschitz, pick $\delta > 0$ and $\gamma > 0$ such that

$$\|G(z,t) - G(z,t')\| \leq \gamma|t - t'| \text{ for } z \in B(z_0,\delta).$$

Again using that G is fully matricial in z and locally Lipschitz in t, we conclude that for each $t_0 \in [0,T]$, there exists $\epsilon(t_0) > 0$ and $\delta(t_0) \in (0,\delta)$ such that if $|t - t_0| < \epsilon(t_0)$ and $z \in$

$B(z_0, \delta(t_0))$, then $G(z,t) \in B(G(z_0,t_0), r(G(z_0,t_0)))$. Hence for all $t, t' \in (t_0 - \epsilon(t_0), t_0 + \epsilon(t_0))$ and $z \in B(z_0, \delta(t_0))$, we have $G(z,t), G(z,t') \in B(G(z_0,t_0), r(G(z_0,t_0)))$, which implies that

$$\begin{aligned}\|F(G(z,t),t) - F(G(z,t'),t)\| &\leq M(G(z_0,t_0))\|G(z,t) - G(z,t')\| + C(G(z_0,t_0))|t-t'| \\ &\leq M(G(z_0,t_0))\gamma|t-t'| + C(G(z_0,t_0))|t-t'|.\end{aligned}$$

Because $[0,T]$ is compact, it can be covered by finitely many of the intervals $(t_0 - \epsilon(t_0), t_0 + \epsilon(t_0))$, say those indexed by t_0 in a finite set S. Let

$$\begin{aligned}\delta^* &= \min_{t_0 \in S} \delta(t_0) \\ C^* &= \max_{t_0 \in S} \left(M(G(z_0,t_0))\gamma + C(G(z_0,t_0)) \right).\end{aligned}$$

By the Lebesgue number lemma, choose $\epsilon^* > 0$ such that every interval of length ϵ^* is contained in one of the intervals $(t_0 - \epsilon(t_0), t_0 + \epsilon(t_0))$ for $t_0 \in S$. Then if $t, t' \in [0,T]$ with $|t - t'| < \epsilon^*$, then they are both contained in one interval $(t_0 - \epsilon(t_0), t_0 + \epsilon(t_0))$, and hence for all $z \in B(z_0, \delta^*)$,

$$\|F(G(z,t)) - F(G(z,t'))\| \leq C^*|t-t'|.$$

Since this holds for $|t-t'|$ sufficiently small, it holds for all $t, t' \in [0,T]$, and hence $F(G(z,t),t)$ is a locally Lipschitz family.

Now let us prove the chain rule identity (7.15). As mentioned in the statement of the lemma, $DF^{(n)}(G^{(n)}(z,t),t)[\partial_t G^{(n)}(z,t)]$ is given by definition 7.2.7; here we use the fact that $DF^{(n)}(G^{(n)}(z,t),t)$ is locally Lipschitz in t which follows by a straighforward argument, and hence that $DF^{(n)}(G^{(n)}(z,t),t)$ is a Bochner L^∞ function from $[0,T]$ to completely bounded linear transformations $M_n(\mathcal{B}_2) \to M_n(\mathcal{B}_3)$. The other term $\partial_t F^{(n)}(G^{(n)}(z,t),t)$ is given by definition 7.2.9. Once we know that both sides of (7.15) are well-defined in $\mathcal{L}(L^1[0,T], M_n(\mathcal{B}_3))$, we can show that they are equal by proving that if we integrate both sides on a small interval $[s,t]$, then they agree up to a higher order error (Lemma 7.2.2).

So consider an interval $[s,t] \subseteq [0,T]$ and $z_0 \in \Omega_1^{(n)}$. Unsurprisingly, the first step is to write that

$$\begin{aligned}&F(G(z_0,t),t) - F(G(z_0,s),s) \\ &\qquad = [F(G(z_0,t),t) - F(G(z_0,s),t)] + [F(G(z_0,s),t) - F(G(z_0,s),s)]. \quad (7.16)\end{aligned}$$

The first term on the right-hand side can be rewritten as

$$\begin{aligned}\Delta F(G(z_0,t), G(z_0,s))[G(z_0,t) - G(z_0,s)] &= \Delta F(G(z_0,t), G(z_0,s)) \left[\int_s^t \partial_u G(z_0,u)\, du \right] \\ &= \int_s^t \Delta F(G(z_0,t), G(z_0,s))[\partial_u G(z_0,u)]\, du.\end{aligned}$$

Using the local Lipschitz estimates for F and G in both z and t, we know that for $u \in [s,t]$,

$$
\begin{aligned}
&\Delta F(G(z_0,t), G(z_0,s)) \\
&\quad = \Delta F(G(z_0,u), G(z_0,u)) + O(\|G(z_0,t) - G(z_0,u)\|) + O(\|G(z_0,s) - G(z_0,u)\|) \\
&\quad = DF(G(z_0,u)) + O(|s-t|),
\end{aligned}
$$

where the estimate works uniformly for $s, t \in [0,T]$ and the error is measured in $\|\cdot\|_{\#}$. Then applying Observation 7.2.8, we get

$$
\int_s^t \Delta F(G(z_0,t), G(z_0,s))[\partial_u G(z_0,u)]\, du = \int_s^t DF(G(z_0,u))[\partial_u G(z_0,u)]\, du + O(|s-t|^2).
$$

Therefore,

$$
F(G(z_0,t),t) - F(G(z_0,s),t) = \int_s^t DF(G(z_0,u))[\partial_u G(z_0,u)]\, du + O(|s-t|^2). \tag{7.17}
$$

Now we turn our attention to the second term on the right hand side of (7.16). Clearly,

$$
F(G(z_0,s),t) - F(G(z_0,s),s) = \int_s^t \partial_u F(G(z_0,s),u)\, du.
$$

For $u \in [s,t]$, we have $G(z_0,s) = G(z_0,u) + O(|s-t|)$ using the local Lipschitz properties of G. So then using Observation 7.2.10 on the interval $[s,t]$, we get that

$$
F(G(z_0,s),t) - F(G(z_0,s),s) = \int_s^t \partial_u F(G(z_0,u),u)\, du + O(|s-t|^2). \tag{7.18}
$$

Overall, by substituting (7.17) and (7.18) into (7.16), we have

$$
\int_s^t \partial_u[F(G(z,u),u)]\, du = \int_s^t DF(G(z_0,u))[\partial_u G(z_0,u)]\, du + \int_s^t \partial_u F(G(z_0,u),u)\, du + O(|s-t|^2).
$$

By Lemma 7.2.2, the chain rule identity holds. □

Remark 7.3.5. One can show that the two terms on the right-hand side of the chain rule, namely $DF^{(n)}(G(z,t),t)[\partial_t G(z,t)]$ and $\partial_t F(G(z,t),t)$ are distributional families of fully matricial functions. The proof uses the same simple-minded techniques we have used so far in this section. To show that the integral against a function ϕ in $L^1[0,T]$ is fully matricial, one uses step function approximations the same as in the definitions of these two terms (Definitions 7.2.7 and 7.2.9). And to obtain bounds on the norm in $\mathcal{L}(L^1[0,T], M_n(\mathcal{B}))$ uniformly for z in a neighborhood of z_0, one uses the local Lipschitz estimates and the compactness argument in the proof of Lemma 7.3.4 where necessary.

Remark 7.3.6. More generally, suppose that $f(z,t)$ is a distributional family of fully matricial functions and $F(z,t)$ is a locally Lipschitz family such that the image of F is contained in the space-domain of f. Then $g(z,t) := f(F(z,t),t)$ is a distributional family of fully matricial functions. Also, given another locally Lipschitz family $G(z,t)$, we have $g(G(z,t),t) = f(H(z,t),t)$ where $H(z,t) = F(G(z,t),t)$. In other words, the composition $f(F(G(z,t),t),t)$ is well-defined regardless of the order in which we perform the composition operations. The proof is a routine usage of step-function approximation techniques.

7.3.3 Distributional families of generalized laws

In (7.1), the left-hand side $\partial_t F(z,t)$ will be a distributional family of fully matricial functions. On the right-hand side, the term $G_{\sigma(\cdot,t)}$ will also be a distributional family of fully matricial functions. So $\sigma(\cdot,t)$ needs to be a family of generalized laws that depends on t in a distributional sense. Let us now explain the precise definitions and properties of such a σ.

Definition 7.3.7. A *distributional family of generalized laws on* $[0,T]$ is a map σ

$$\mathcal{B}\langle Y\rangle \times L^1[0,T] \to \mathcal{B}, \qquad (p(Y),\phi) \mapsto \int_0^T \sigma(p(Y),t)\phi(t)\,dt$$

with the following properties:

(1) For each $p(Y) \in \mathcal{B}\langle Y\rangle$, the map $\sigma(p(Y),\cdot)$ is in $\mathcal{L}(L^1[0,T],\mathcal{B})$.

(2) For each $\phi \geq 0$, the map $\int_0^T \sigma(\cdot,t)\phi(t)\,dt$ is a generalized law.

Lemma 7.3.8. *Let σ be a distributional family of generalized laws on $[0,T]$. Then we have for $p(Y) \in \mathcal{B}\langle Y\rangle$ and $\phi \in L^1[0,T]$ that*

$$\left\| \int_0^T \sigma(b_0 Y b_1 \dots Y b_k, t)\phi(t)\,dt \right\| \leq \|\sigma(1,\cdot)\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} \operatorname{rad}\left(\int_0^T \sigma(\cdot,t)\,dt \right)^k \|b_0\| \dots \|b_k\| \|\phi\|_{L^1[0,T]}. \tag{7.19}$$

In particular, for $\phi \geq 0$, we have

$$\operatorname{rad}\left(\int_0^T \sigma(\cdot,t)\phi(t)\,dt \right) \leq \operatorname{rad}\left(\int_0^T \sigma(\cdot,t)\,dt \right), \tag{7.20}$$

and we also have

$$\|\sigma(b_0 Y b_1 \dots Y b_k,\cdot)\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} \leq \|\sigma(1,\cdot)\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} \operatorname{rad}\left(\int_0^T \sigma(\cdot,t)\,dt \right)^k \|b_0\| \dots \|b_k\|. \tag{7.21}$$

Proof. First, consider an interval $[s,t] \subseteq [0,T]$. Note that

$$\int_0^T \sigma(\cdot,u)\,du = \int_0^s \sigma(\cdot,u)\,du + \int_s^t \sigma(\cdot,u)\,du + \int_t^T \sigma(\cdot,u)\,du,$$

and therefore by Lemma 2.6.8,

$$\operatorname{rad}\left(\int_s^t \sigma(\cdot,u)\,du \right) \leq \operatorname{rad}\left(\int_0^T \sigma(\cdot,u)\,du \right).$$

Therefore, by (2.1), we have

$$\left\| \int_s^t \sigma(b_0 Y b_1 \dots Y b_k, u)\, du \right\| \leq \left\| \int_s^t \sigma(1,u)\, du \right\| \operatorname{rad} \left(\int_s^t \sigma(\cdot, u)\, du \right)^k \|b_0\| \dots \|b_k\|$$
$$\leq \|\sigma(1,\cdot)\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} \operatorname{rad} \left(\int_0^T \sigma(\cdot, t)\, dt \right)^k \|b_0\| \dots \|b_k\| \|\chi_{[s,t]}\|_{L^1}.$$

If we consider a step function ϕ, that is, a linear combination of indicator functions of disjoint intervals, then (7.19) holds by the previous estimate and the triangle inequality. Since step functions are dense in $L^1[0,T]$, the relation (7.19) extends to all $\phi \in L^1[0,T]$. Finally, (7.19) easily implies (7.20) and (7.21). □

Next, given a distributional family of generalized laws σ, we will define the Cauchy-Stieltjes transform G_σ which will be a distributional family of fully matricial functions. Given $\phi \in L^1[0,T]$, let us write $\phi = \phi_1 - \phi_2$, where ϕ_1 and ϕ_2 are nonnegative L^1 functions. Then we set

$$\int_0^T G_{\sigma(\cdot,t)}^{(n)}(z)\phi(t)\, dt = G^{(n)}_{\int_0^T \sigma(\cdot,t)\phi_1(t)\, dt}(z) - G^{(n)}_{\int_0^T \sigma(\cdot,t)\phi_2(t)\, dt}(z).$$

The right-hand side makes sense because $\int_0^T \sigma(\cdot,t)\phi_j(t)\, dt$ is a generalized law; moreover, the right-hand side is independent of the choice of decomposition of ϕ into ϕ_1 and ϕ_2 because of the fact that $G_{\tau_1+\tau_2} = G_{\tau_1} + G_{\tau_2}$ for generalized laws τ_1 and τ_2.

Moreover, if $\phi = \phi_1 - \phi_2$, we have the estimate that for $\operatorname{Im} z \geq \epsilon$,

$$\left\| \int_0^T G_{\sigma(\cdot,t)}^{(n)}(z)\phi_j(t)\, dt \right\| \leq \frac{1}{\epsilon} \|\sigma(1,\cdot)\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} \|\phi_j\|_{L^1[0,T]},$$

which follows from Lemma 4.2.7. So if we take ϕ_1 and ϕ_2 to be respectively the positive and negative parts of ϕ and use the triangle inequality, we get the estimate

$$\left\| \int_0^T G_{\sigma(\cdot,t)}^{(n)}(z)\phi(t)\, dt \right\| \leq \frac{1}{\epsilon} \|\sigma(1,\cdot)\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} \|\phi\|_{L^1[0,T]}.$$

This means that $G_\sigma^{(n)}(z)$ is an element of $\mathcal{L}(L^1[0,T], M_n(\mathcal{B}))$ and

$$\left\|G_\sigma^{(n)}(z)\right\|_{\mathcal{L}(L^1[0,T],M_n(\mathcal{B}))} \leq \frac{1}{\epsilon} \|\sigma(1,\cdot)\|_{\mathcal{L}(L^1[0,T],\mathcal{B})}. \tag{7.22}$$

7.4 Differential equations associated to processes with independent increments

Theorem 7.4.1. *Let* ind $\in$ {bool, free, mono, mono †}*. Let* $(\mu_t)_{t\in[0,T]}$ *be a Lipschitz* ind-*subordination family. Then* $(F_{\mu_t})_{t\in[0,T]}$ *is a locally Lipschitz family of fully matricial functions, and there exists a unique* $b \in \mathcal{L}(L^1[0,T],\mathcal{B})$ *and a unique distributional family of* *generalized laws* $\sigma(\cdot,t)$ *such that*

$$\partial_t F_{\mu_t}^{(n)}(z) = \begin{cases} -[b(t) + G_{\sigma(\cdot,t)}(z)], & \textit{boolean case}, \\ -DF_{\mu_t}^{(n)}(z)[b(t) + G_{\sigma(\cdot,t)}^{(n)}(F_{\mu_t}^{(n)}(z))], & \textit{free case}, \\ -DF_{\mu_t}^{(n)}(z)[b(t) + G_{\sigma(\cdot,t)}^{(n)}(z)], & \textit{monotone case}, \\ -[b(t) + G_{\sigma(\cdot,t)}^{(n)}(F_{\mu_t}^{(n)}(z))], & \textit{anti-monotone case}, \end{cases} \tag{7.23}$$

7.4.1 Construction of σ

First, let $b \in \mathcal{L}(L^1[0,T],\mathcal{B})$ be the distributional time-derivative of $t \mapsto \mu_t(X)$, which makes sense because $\mu_t(X)$ is Lipschitz. The distributional law σ will be defined by a limiting procedure using approximations by step functions associated to partitions, very much in the spirit of Riemann integration.

Let $\mathcal{P} = \{t_0,\dots,t_N\}$ be a partition of $[0,T]$ (where we use the word "partition" in the sense of Riemann integration and follow the convention that $0 = t_0 < t_1 < \dots < t_N = T$). We denote $\operatorname{mesh}(\mathcal{P}) = \max_j (t_j - t_{j-1})$. For $0 \leq s \leq t \leq T$, let $\mu_{s,t}$ be the unique law such that $\mu_t = \mu_s \boxplus_{\text{ind}} \mu_{s,t}$. Let $\tau_{s,t}$ be given by Theorem 4.5.3 by the relation

$$F_{\mu_{s,t}}^{(n)}(z) = z - \mu_{s,t}(X)^{(n)} - G_{\tau_{s,t}}^{(n)}(z).$$

Then define $\sigma_{\mathcal{P}} : \mathcal{B}\langle Y \rangle \times [0,T] \to \mathcal{B}$ by

$$\sigma_{\mathcal{P}}(p(Y),t) = \sum_{j=1}^N \frac{1}{t_j - t_{j-1}} \tau_{t_{j-1},t_j}(p(Y)) \chi_{[t_{j-1},t_j)}(t).$$

Of course, $\sigma_{\mathcal{P}}$ defines a map $\mathcal{B}\langle Y \rangle \times L^1[0,T] \to \mathcal{B}$ by

$$(p(Y),\phi) \mapsto \int_0^T \sigma_{\mathcal{P}}(p(Y),t)\,dt = \sum_{j=1}^N \frac{1}{t_j - t_{j-1}} \tau_{t_{j-1},t_j}(p(Y)) \int_{t_{j-1}}^{t_j} \phi(t)\,dt.$$

It is clear that if $\phi \geq 0$, then $\int_0^T \sigma_{\mathcal{P}}(\cdot,t)\phi(t)\,dt$ is a positive linear combination of generalized laws, and hence is a generalized law. Thus, $\sigma_{\mathcal{P}}$ may be viewed as a distributional family of generalized laws. We claim that $\sigma_{\mathcal{P}}$ converges as $\operatorname{mesh}(\mathcal{P}) \to 0$ to a distributional family of generalized laws σ in the sense that

$$\lim_{\operatorname{mesh}(\mathcal{P}) \to 0} \int_0^T \sigma_{\mathcal{P}}(p(Y),t)\phi(t)\,dt = \int_0^T \sigma(p(Y),t)\phi(t)\,dt \tag{7.24}$$

for all $p \in \mathcal{B}\langle Y \rangle$ and $\phi \in L^1[0,T]$. As a first step, we observe the following a priori bounds for $\sigma_{\mathcal{P}}$.

Lemma 7.4.2. *Let* $R = \sup_{s,t} \operatorname{rad}(\mu_{s,t})$, *and let* C *be the Lipschitz norm of* $t \mapsto \mu_t((X - \mu_t(X))^2)$. *Then* $\operatorname{rad}(\sigma_{\mathcal{P}}) \leq R$ *and* $\|\sigma_{\mathcal{P}}(1,\cdot)\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} \leq C$. *Moreover, for a monomial* $p(Y) = b_0 Y b_1 \dots Y b_k$, *we have* $\|\sigma_{\mathcal{P}}(p(Y),\cdot)\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} \leq R^k \|b_0\| \dots \|b_k\| C$.

Proof. By Theorem 4.5.3, specifically (4.3), we have $\operatorname{rad}(\tau_{t_{j-1},t_j}) \leq \operatorname{rad}(\mu_{t_{j-1},t_j}) \leq R$, and it is easy to see that $\operatorname{rad}(\sigma_{\mathcal{P}}) = \max_j \operatorname{rad}(\tau_{t_{j-1},t_j})$. For the second bound, note by the same theorem and by our assumption that the subodination family is Lipschitz,

$$\|\tau_{t_{j-1},t_j}(1)\| = \|\operatorname{Var}(\mu_{t_{j-1},t_j})[1]\| = \|\operatorname{Var}(\mu_{t_j}))[1] - \operatorname{Var}(\mu_{t_{j-1}})[1]\| \leq C|t_j - t_{j-1}|.$$

Hence,

$$\left\| \int_0^T \sigma_{\mathcal{P}}(1,t)\phi(t)\,dt \right\| \leq \sum_{j=1}^N \frac{1}{t_j - t_{j-1}} C(t_j - t_{j-1}) \int_{t_{j-1}}^{t_j} |\phi(t)|\,dt = C\|\phi\|_{L^1[0,T]}.$$

To prove the third estimate, it suffices by Lemma 7.2.2 to show that $\|\int_s^t \sigma_{\mathcal{P}}(p(Y),t)\,dt\| \leq R^k\|b_0\| \dots \|b_k\| C|t-s|$ when $[s,t] \subseteq [0,T]$. But this follows from the estimate (2.1) applied to the generalized law $\int_s^t \sigma_{\mathcal{P}}(\cdot,t)\,dt$. □

In order to establish the existence of σ satisfying (7.24), it is sufficient to show the existence of the limit $\lim_{\operatorname{mesh}(\mathcal{P}) \to 0} \int_0^T \sigma_{\mathcal{P}}(p(Y),t)\phi(t)\,dt$ for every $p(Y) \in \mathcal{B}\langle Y \rangle$ and $\phi \in L^1[0,T]$. The fact that the limiting object σ is a distributional family of generalized laws will then follow automatically. Indeed, using our a priori bounds, we would have that $\int_0^T \sigma(p(Y),t)\phi(t)\,dt \leq R^k\|b_0\| \dots \|b_k\| L$ when $p(Y) = b_0 Y b_1 \dots Y b_k$. Also, for $\phi \geq 0$, the object $\int_0^T \sigma(\cdot,t)\phi(t)\,dt$ would be completely positive since $\sigma_{\mathcal{P}}$ is completely positive. So σ would be a distributional family of generalized laws with $\operatorname{rad}(\sigma) \leq R$ and $\|\sigma(1,\cdot)\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} \leq L$.

Moreover, to show that the limit $\lim_{\operatorname{mesh}(\mathcal{P}) \to 0} \int_0^T \sigma_{\mathcal{P}}(p(Y),t)\phi(t)\,dt$ exists, it suffices to check this for a family of functions ϕ whose span is dense in $L^1[0,T]$ because of our a priori bounds. Thus, we may restrict our attention to nonnegative continuous ϕ. Now $\int_0^T \sigma_{\mathcal{P}}(\cdot,t)\phi(t)\,dt$ is a generalized law with radius bounded by R and $\|\int \sigma(1,t)\phi(t)\,dt\| \leq L\|\phi\|_{L^1[0,T]}$. The convergence of this family of laws as $\operatorname{mesh}(\mathcal{P}) \to 0$ is equivalent to convergence of the Cauchy-Stieltjes transforms

$$G^{(n)}_{\int_0^T \sigma_{\mathcal{P}}(\cdot,t)\phi(t)\,dt}(z) = \int_0^T G^{(n)}_{\sigma_{\mathcal{P}}(\cdot,t)}(z)\phi(t)\,dt.$$

So it suffices to show that for z in the upper half-plane and $\phi \geq 0$ continuous,

$$\lim_{\operatorname{mesh}(\mathcal{P}) \to 0} \int_0^T G^{(n)}_{\sigma_{\mathcal{P}}(\cdot,t)}(z)\phi(t)\,dt \text{ exists in } \mathcal{B}.$$

To estimate the error when we pass from one partition to another partition which refines the first, we will use the following lemma.

Lemma 7.4.3. *Let $[s,t] \subseteq [0,T]$, and let $\mathcal{Q} = \{t_0, \dots, t_N\}$ be a partition of $[s,t]$. Then for each $z \in \mathbb{H}_+^{(n)}(\mathcal{B})$, we have*

$$G^{(n)}_{\tau_{s,t}}(z) = \sum_{j=1}^N G^{(n)}_{\tau_{t_{j-1},t_j}}(z) + O(|t-s|^2),$$

where the error estimate depends on z but holds uniformly for all (s,t) with $|s-t|$ sufficiently small and is independent of the partition $\mathcal{P}$. The precise estimates for each type of independence are written below in (7.25)*,* (7.27)*, and* (7.28)*.*

Proof. Clearly, it is necessary to split the proof into cases for each type of independence. In the boolean case, $\mu_{s,t}$ is the boolean convolution of $\mu_{t_0,t_1}, \dots, \mu_{t_{N-1},t_N}$ and hence

$$K_{\mu_{s,t}}^{(n)}(z) = \sum_{j=1}^{N} K_{\mu_{t_{j-1},t_j}}(z),$$

where

$$K_{\mu_{s,t}}^{(n)}(z) = z - F_{\mu_{s,t}}^{(n)}(z) = \mu_{s,t}(X)^{(n)} + G_{\tau_{s,t}}^{(n)}(z).$$

We also have $\mu_{s,t}(X) = \sum_{j=1}^{N} \mu_{t_{j-1},t_j}(X)$, and hence

$$G_{\tau_{s,t}}^{(n)}(z) = \sum_{j=1}^{N} G_{\tau_{t_{j-1},t_j}}^{(n)}(z). \tag{7.25}$$

Next, consider the free case. By free convolution, we have

$$\Phi_{\mu_{s,t}}^{(n)}(z) = \sum_{j=1}^{N} \Phi_{\mu_{t_{j-1},t_j}}^{(n)}(z),$$

where

$$\Phi_{\mu_{s,t}}^{(n)}(z) = (F_{\mu_{s,t}}^{(n)})^{-1}(z) - z$$

which is defined whenever $\operatorname{Im} z \geq 2\delta$ and $\delta > \|\tau_{s,t}(1)\|^{1/2}$. Now we claim that for such z, we acutally have $\Phi_{\mu_{s,t}}^{(n)}(z) = \mu_{s,t}(X)^{(n)} + G_{\tau_{s,t}}^{(n)}(z) + O(|s-t|^2)$ (and of course the same holds with (s,t) replaced by (t_{j-1},t_j)). Indeed, by (4.9) and our assumption that $(\mu_t)_{t\in[0,T]}$ is a Lipschitz subordination family, we get

$$\begin{aligned} \left\|\Phi_{\mu_{s,t}}^{(n)}(z) - \mu_{s,t}(X)^{(n)} - G_{\tau_{s,t}}^{(n)}(z)\right\| &\leq \frac{\|\tau_{s,t}(1)\|}{\delta \cdot 2\delta}\left(\|\mu_{s,t}(X)\| + \frac{\|\tau_{s,t}(1)\|}{\delta}\right) \\ &\leq \frac{C}{2\delta^2}(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + C/\delta)(t-s)^2. \end{aligned} \tag{7.26}$$

Therefore,

$$\left\|G_{\tau_{s,t}}^{(n)}(z) - \sum_{j=1}^{N} G_{\tau_{t_{j-1},t_j}}^{(n)}(z)\right\| \leq \frac{C}{2\delta^2}(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + C/\delta)\left((t-s)^2 + \sum_{j=1}^{N}(t_j - t_{j-1})^2\right).$$

Since of course $\sum_{j=1}^{N}(t_j - t_{j-1})^2 \leq (t-s)^2$, we get

$$\left\|G_{\tau_{s,t}}^{(n)}(z) - \sum_{j=1}^{N} G_{\tau_{t_{j-1},t_j}}^{(n)}(z)\right\| \leq \frac{C}{\delta^2}(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + C/\delta)(t-s)^2 \tag{7.27}$$

whenever $\operatorname{Im} z \geq 2\delta \geq 2C^{1/2}(t-s)^{1/2}$.

Finally, consider the monotone case. Note that

$$\begin{aligned} z - F_{\mu_{s,t}}^{(n)}(z) &= \sum_{j=1}^{N} \left(F_{\mu_{s,t_{j-1}}}^{(n)}(z) - F_{\mu_{s,t_j}}^{(n)}(z) \right) \\ &= \sum_{j=1}^{N} \left(F_{\mu_{s,t_{j-1}}}^{(n)}(z) - F_{\mu_{t_{j-1},t_j}}^{(n)} \circ F_{\mu_{s,t_{j-1}}}^{(n)}(z) \right) \\ &= \sum_{j=1}^{N} \left(\operatorname{id} - F_{\mu_{t_{j-1},t_j}}^{(n)} \right) \circ F_{\mu_{s,t_{j-1}}}^{(n)}(z) \\ &= \sum_{j=1}^{N} \left(\mu_{t_{j-1},t_j}(X) + G_{\tau_{t_{j-1},t_j}}^{(n)} \circ F_{\mu_{s,t_{j-1}}}^{(n)}(z) \right), \end{aligned}$$

so that

$$G_{\tau_{s,t}}^{(n)}(z) = \sum_{j=1}^{N} G_{\tau_{t_{j-1},t_j}}^{(n)} \circ F_{\mu_{s,t_{j-1}}}^{(n)}(z).$$

But recall that for $\operatorname{Im} z \geq \delta > 0$, we have

$$\left\| F_{\mu_{s,t_{j-1}}}^{(n)}(z) - z \right\| \leq \|\mu_{s,t_{j-1}}(X)\| + \frac{\|\tau_{s,t_{j-1}}(1)\|}{\delta}$$

and hence because $\operatorname{Im} F_{\mu_{s,t}}^{(n)}(z) \geq \delta$ also, we have

$$\begin{aligned} \left\| G_{\tau_{t_{j-1},t_j}}^{(n)} \circ F_{\mu_{s,t_{j-1}}}^{(n)}(z) - G_{\tau_{t_{j-1},t_j}}^{(n)}(z) \right\| &\leq \frac{\|\tau_{t_{j-1},t_j}(1)\|}{\delta^2} \left(\|\mu_{s,t_{j-1}}(X)\| + \frac{\|\tau_{s,t_{j-1}}(1)\|}{\delta} \right) \\ &\leq \frac{C}{\delta^2} \left(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + \frac{C}{\delta} \right) (t_j - t_{j-1})(t_{j-1} - s). \end{aligned}$$

Therefore,

$$\begin{aligned} \left\| G_{\tau_{s,t}}^{(n)}(z) - \sum_{j=1}^{N} G_{\tau_{t_{j-1},t_j}}^{(n)}(z) \right\| &\leq \sum_{j=1}^{N} \left\| G_{\tau_{t_{j-1},t_j}}^{(n)} \circ F_{\mu_{s,t_{j-1}}}^{(n)}(z) - G_{\tau_{t_{j-1},t_j}}^{(n)}(z) \right\| \\ &\leq \frac{C}{\delta^2} \left(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + \frac{C}{\delta} \right) \sum_{j=1}^{N} (t_j - t_{j-1})(t_{j-1} - s). \end{aligned}$$

The summation on the right hand side is a lower Riemann sum for the function $x \mapsto (x-s)$ on $[s,t]$, which integrates to $(t-s)^2/2$. Therefore,

$$\left\| G_{\tau_{s,t}}^{(n)}(z) - \sum_{j=1}^{N} G_{\tau_{t_{j-1},t_j}}^{(n)}(z) \right\| \leq \frac{C}{2\delta^2} \left(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + \frac{C}{\delta} \right) (t-s)^2. \tag{7.28}$$

The identical estimate holds in the anti-monotone case, and the proof is the same except that the order of the time indices is reversed. □

Lemma 7.4.4. *If $\phi : [0,T] \to [0,\infty)$ is continuous and $z \in \mathbb{H}_+^{(n)}(\mathcal{B})$, then*

$$\lim_{\operatorname{mesh}(\mathcal{P}) \to 0} \int_0^T G_{\sigma_{\mathcal{P}}(\cdot,t)}^{(n)}(z)\phi(t)\,dt \text{ exists.}$$

Hence, there exists a distributional family of generalized laws σ with

$$\lim_{\operatorname{mesh}(\mathcal{P}) \to 0} \int_0^T \sigma_{\mathcal{P}}(p(Y),t)\phi(t)\,dt = \int_0^T \sigma(p(Y),t)\phi(t)\,dt \text{ for } \phi \in L^1[0,T] \text{ and } p \in \mathcal{B}\langle Y \rangle.$$

Moreover, σ is the unique distributional family of generalized laws such that for each $z \in \mathbb{H}_+^{(n)}(\mathcal{B})$, we have

$$\lim_{|s-t| \to 0} \frac{1}{|s-t|} \left(G_{\tau_{s,t}}^{(n)}(z) - \int_s^t G_{\sigma(\cdot,u)}^{(n)}(z)\,du \right) = 0.$$

Proof. Fix $\phi : [0,T] \to \mathbb{C}$ continuous. Let $\mathcal{P} = \{t_1, \dots, t_N\}$ be a partition of $[0,T]$, and let $\mathcal{Q}$ be a refinement of $\mathcal{P}$. Let $\phi_{\mathcal{P}}$ be the step function

$$\phi_{\mathcal{P}} = \sum_{j=1}^N \chi_{[t_{j-1},t_j)} \frac{1}{t_j - t_{j-1}} \int_{t_{j-1}}^{t_j} \phi.$$

Let $\mathcal{Q}_j = \mathcal{Q} \cap [t_{j-1}, t_j]$. Since ϕ is constant on $[t_{j-1}, t_j)$, we can apply the previous lemma to $\mathcal{Q}_j$ to conclude that

$$\begin{aligned}\int_{t_{j-1}}^{t_j} G_{\sigma_{\mathcal{Q}}(\cdot,t)}^{(n)}(z)\phi_{\mathcal{P}}(t)\,dt - G_{\tau_{t_{j-1},t_j}}^{(n)}(z) \int_{t_{j-1}}^{t_j} \phi_{\mathcal{P}}(t)\,dt &= O(\|\phi\|_{L^\infty[0,T]} |t_j - t_{j-1}|^2) \\ &= O(\|\phi\|_{L^\infty[0,T]} (t_j - t_{j-1}) \operatorname{mesh}(\mathcal{P})).\end{aligned}$$

Summing over j leads to

$$\int_0^T G_{\sigma_{\mathcal{Q}}(\cdot,t)}^{(n)}(z)\phi_{\mathcal{P}}(t)\,dt - \int_0^T G_{\sigma_{\mathcal{P}}(\cdot,t)}^{(n)}(z)\phi_{\mathcal{P}}(t)\,dt = O(\|\phi\|_{L^\infty[0,T]} T \operatorname{mesh}(\mathcal{P})).$$

Moreover, from the definition of $\sigma_{\mathcal{P}}$, it is immediate that

$$\int_0^T G_{\sigma_{\mathcal{P}}(\cdot,t)}^{(n)}(z)\phi_{\mathcal{P}}(t)\,dt = \int_0^T G_{\sigma_{\mathcal{P}}(\cdot,t)}^{(n)}(z)\phi(t)\,dt.$$

Finally, since $\|\phi_{\mathcal{P}} - \phi\|_{L^1[0,T]} \leq T\omega_\phi(\operatorname{mesh}(\mathcal{P}))$, we have

$$\int_0^T G_{\sigma_{\mathcal{Q}}(\cdot,t)}^{(n)}(z)\phi_{\mathcal{P}}(t)\,dt - \int_0^T G_{\sigma_{\mathcal{Q}}(\cdot,t)}^{(n)}(z)\phi(t)\,dt = O(T\omega_\phi(\operatorname{mesh}(\mathcal{P}))).$$

Thus, overall

$$\int_0^T G^{(n)}_{\sigma_{\mathcal{P}}(\cdot,t)}(z)\phi(t)\,dt - \int_0^T G^{(n)}_{\sigma_{\mathcal{Q}}(\cdot,t)}(z)\phi(t)\,dt = O(T\|\phi\|_{L^\infty[0,T]}\operatorname{mesh}(\mathcal{P})) + O(T\omega_\phi(\operatorname{mesh}(\mathcal{P}))).$$

This shows that the sequence of integrals is Cauchy as $\operatorname{mesh}(\mathcal{P}) \to 0$, and therefore, the limit exists.

From the discussion above Lemma 7.4.3, this implies the existence of a distributional family of generalized laws σ such that $\int_0^T \sigma_{\mathcal{P}}(p(Y),t)\phi(t)\,dt \to \int_0^T \sigma(p(Y),t)\phi(t)\,dt$ for $\phi \in L^1[0,T]$.

Now we prove the final claim of the lemma. From the preceding lemma, for every partition $\mathcal{Q}$ of $[s,t]$, we have

$$G^{(n)}_{\tau_{s,t}}(z) - \int_s^t G^{(n)}_{\sigma_{\mathcal{Q}}(\cdot,u)}(z)\,du = O(|s-t|^2).$$

In particular, for every partition $\mathcal{P}$ of $[0,T]$, if we let $\mathcal{P}' = \mathcal{P} \cup \{s,t\}$, then we have

$$G^{(n)}_{\tau_{s,t}}(z) - \int_s^t G^{(n)}_{\sigma_{\mathcal{P}'}(\cdot,u)}(z)\,du = O(|s-t|^2).$$

Now as $\operatorname{mesh}(\mathcal{P}) \to 0$, we also have $\operatorname{mesh}(\mathcal{P}') \to 0$, and hence $\sigma_{\mathcal{P}'}$ converges to σ. So

$$G^{(n)}_{\tau_{s,t}}(z) - \int_s^t G^{(n)}_{\sigma(\cdot,u)}(z)\,du = O(|s-t|^2),$$

which is $o(|s-t|)$. Conversely, if $\overline{\sigma}$ is another law satisfying

$$G^{(n)}_{\tau_{s,t}}(z) - \int_s^t G^{(n)}_{\overline{\sigma}(\cdot,u)}(z)\,du = o(|s-t|),$$

then we have

$$\int_s^t G^{(n)}_{\sigma(\cdot,u)}(z)\,du - \int_s^t G^{(n)}_{\overline{\sigma}(\cdot,u)}(z)\,du = o(|s-t|).$$

But by Lemma 7.2.2, this implies that $G^{(n)}_{\sigma(\cdot,\cdot)} = G_{\overline{\sigma}}(\cdot,\cdot)$ in $\mathcal{L}(L^1[0,T],\mathcal{B})$ for each z. Hence, $\sigma = \overline{\sigma}$. □

7.4.2 Differential equations and estimates

The remainder of the proof of Theorem 7.4.1 is to check that F_{μ_t} and $G_{\sigma(\cdot,t)}$ satisfy the asserted differential equations. This will be done for each type of independence separately, and we will also state precise estimates for the convergence of $\sigma_{\mathcal{P}}$ to σ as $\operatorname{mesh}(\mathcal{P}) \to 0$.

Proposition 7.4.5. *Let $(\mu_t)_{t\in[0,T]}$ be a Lipschitz boolean subordination family, and continue all the notation from above.*

(1) For $0 \leq s \leq t \leq T$,

$$K_{\mu_{s,t}}^{(n)}(z) = \int_s^t b(u)^{(n)}\,du + \int_s^t G_{\sigma(\cdot,u)}^{(n)}(z)\,du.$$

(2) $(F_{\mu_t})_{t\in[0,T]}$ is a locally Lipschitz family of fully matricial functions and

$$\partial_t F_{\mu_t}^{(n)}(z) = -[b(t)^{(n)} + G_{\sigma(\cdot,t)}^{(n)}(z)].$$

(3) b and σ are uniquely determined by this differential equation.

(4) If $\phi : [0,T] \to \mathbb{C}$ is continuous with modulus of continuity ω_ϕ, then

$$\left\| \int_0^T G_{\sigma_{\mathcal{P}}(\cdot,t)}^{(n)}(z)\phi(t)\,dt - \int_0^T G_{\sigma(\cdot,t)}^{(n)}(z)\phi(t)\,dt \right\| \leq \frac{CT}{\delta}\omega_\phi(\operatorname{mesh}(\mathcal{P})),$$

whenever $\operatorname{Im} z \geq \delta$.

Proof. (1) Fix $s \leq t$ in $[0,T]$. Let $\mathcal{P}$ be a partition of $[0,T]$ and let $\mathcal{P}' = \mathcal{P} \cup \{s,t\}$. By applying (7.25) to $\mathcal{Q} := \mathcal{P}' \cap [s,t]$, we obtain

$$G_{\tau_{s,t}}^{(n)}(z) = \int_s^t G_{\sigma_{\mathcal{P}'}(\cdot,u)}^{(n)}(z)\,du.$$

As $\operatorname{mesh}(\mathcal{P}) \to 0$, we have $\operatorname{mesh}(\mathcal{P}') \to 0$ and hence

$$G_{\tau_{s,t}}^{(n)}(z) = \int_s^t G_{\sigma(\cdot,u)}^{(n)}(z)\,du.$$

Also have $K_{\mu_{s,t}}^{(n)}(z) = \mu_{s,t}(X)^{(n)} + G_{\tau_{s,t}}^{(n)}(z)$ and $\mu_{s,t}(X) = \int_s^t b(u)\,du$, which completes the proof of (1).

(2) It follows from (1) that

$$F_{\mu_{s,t}}^{(n)}(z) = z - \int_s^t b(u)^{(n)}\,du - \int_s^t G_{\sigma(\cdot,u)}^{(n)}(z)\,du$$

and hence

$$F_{\mu_t}^{(n)}(z) = z - K_{\mu_0}^{(n)}(z) - \int_0^t [b(u)^{(n)} + G_{\sigma(\cdot,u)}^{(n)}(z)\,du.$$

But since σ is a distributional family of generalized laws, the integral of $b(u)^{(N)} + G_{\sigma(\cdot,u)}^{(n)}$ is a locally Lipschitz family of fully matricial functions. We obtain $F_{\mu_t}^{(n)}$ by adding the fixed function $z - K_{\mu_0}^{(n)}(z)$, so that is also a locally Lipschitz family. And by differentiating the integral equation, we obtain the differential equation (2).

(3) This is clear because b and σ are uniquely determined by $b(t) + G_{\sigma(\cdot,t)}(z)$.

(4) In the proof of Lemma 7.4.4, we explained how to estimate $\int G^{(n)}_{\sigma_{\mathcal{P}}(\cdot,u)}(z)\phi(u)\,du$ minus $\int G^{(n)}_{\sigma_{\mathcal{Q}}(\cdot,u)}(z)\phi(u)\,du$ for a partition $\mathcal{P}$, a refinement $\mathcal{Q}$, and a continuous $\phi : [0,T] \to \mathbb{C}$. If we carry out the estimates at each step explicitly and substitute (7.25), then we obtain the estimate (3). We leave the details to the reader. □

Proposition 7.4.6. *Let $(\mu_t)_{t\in[0,T]}$ be a Lipschitz free subordination family, and continue all the notation from above.*

(1) For $0 \le s \le t \le T$,

$$\Phi^{(n)}_{\mu_{s,t}}(z) = \int_s^t b(u)^{(n)}\,du + \int_s^t G^{(n)}_{\sigma(\cdot,u)}(z)\,du.$$

(2) $(F_{\mu_t})_{t\in[0,T]}$ is a locally Lipschitz family of fully matricial functions and

$$\partial_t F^{(n)}_{\mu_t}(z) = -DF^{(n)}_{\mu_t}(z)\left[b(t)^{(n)} + G^{(n)}_{\sigma(\cdot,t)}(F^{(n)}_{\mu_t}(z))\right].$$

(3) b and σ are uniquely determined by this differential equation.

(4) If $\phi : [0,T] \to \mathbb{C}$ is continuous with modulus of continuity ω_ϕ, then

$$\left\| \int_0^T G^{(n)}_{\sigma_{\mathcal{P}}(\cdot,t)}(z)\phi(t)\,dt - \int_0^T G^{(n)}_{\sigma(\cdot,t)}(z)\phi(t)\,dt \right\| \\ \le \frac{C}{2\delta^2}\left(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + \frac{C}{\delta}\right)\operatorname{mesh}(\mathcal{P}) + \frac{CT}{\delta}\omega_\phi(\operatorname{mesh}(\mathcal{P})),$$

whenever $\operatorname{Im} z \ge 2\delta$ and $\delta \ge C^{1/2}\operatorname{mesh}(\mathcal{P})^{1/2}$.

Proof. (1) Fix $s \le t$. Let $\mathcal{P}$ be a partition of $[0,T]$, let $\mathcal{P}' = \mathcal{P} \cup \{s,t\}$, and $\mathbb{Q} = \mathcal{P}' \cap [s,t]$. Let us write $\mathcal{Q} = \{t_0, \dots, t_N\}$. Then by (7.26), we have

$$\left\|\Phi^{(n)}_{\mu_{t_{j-1},t_j}}(z) - \mu_{t_{j-1},t_j}(X)^{(n)} - G^{(n)}_{\tau_{t_{j-1},t_j}}(z)\right\| \le \frac{C}{2\delta^2}\left(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + \frac{C}{\delta}\right)(t_j - t_{j-1})^2$$

whenever $\operatorname{Im} z \ge 2\delta$ and $\delta > C^{1/2}(t_j - t_{j-1})^{1/2}$. Therefore, summing over j, we get

$$\left\|\Phi^{(n)}_{\mu_{s,t}}(z) - \mu_{s,t}(X)^{(n)} - \int_s^t G^{(n)}_{\sigma_{\mathcal{P}'}(\cdot,u)}(z)\,du\right\| \le \frac{C}{2\delta^2}\left(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + \frac{C}{\delta}\right)(t-s)\operatorname{mesh}(\mathcal{P}').$$

Taking the limit as $\operatorname{mesh}(\mathcal{P}') \to 0$, we obtain $\Phi^{(n)}_{\mu_{s,t}}(z) = \mu_{s,t}(X)^{(n)} + \int_s^t G^{(n)}_{\sigma(\cdot,u)}(z)\,du$ as desired.

(2) In order to obtain our local Lipschitz estimates, fix $s \leq t$ and we will bound $F_{\mu_t}^{(n)}(z) - F_{\mu_s}^{(n)}(z)$. Note that for $\operatorname{Im} z$ sufficiently large

$$(F_{\mu_s}^{(n)})^{-1}(z) - (F_{\mu_t}^{(n)})^{-1}(z) = -\Phi_{\mu_s}^{(n)}(z) + \Phi_{\mu_t}^{(n)}(z) = \Phi_{\mu_{s,t}}^{(n)}(z),$$

so that $(F_{\mu_t}^{(n)})^{-1}(z) = (F_{\mu_s}^{(n)})^{-1}(z) + \Phi_{\mu_{s,t}}^{(n)}(z)$. Thus,

$$\begin{aligned}(F_{\mu_t}^{(n)})^{-1} \circ F_{\mu_s}^{(n)}(z) &= (F_{\mu_s}^{(n)})^{-1} \circ F_{\mu_s}^{(n)}(z) + \Phi_{\mu_{s,t}}^{(n)} \circ F_{\mu_s}^{(n)}(z) \\ &= z + \Phi_{\mu_{s,t}}^{(n)} \circ F_{\mu_s}^{(n)}(z).\end{aligned}$$

It follows that

$$\begin{aligned}F_{\mu_s}^{(n)}(z) &= F_{\mu_t}^{(n)} \circ (F_{\mu_t}^{(n)})^{-1} \circ F_{\mu_s}^{(n)}(z) \\ &= F_{\mu_t}^{(n)} \left(z + \Phi_{\mu_{s,t}}^{(n)} \circ F_{\mu_s}^{(n)}(z) \right),\end{aligned}$$

and hence

$$\begin{aligned}F_{\mu_t}^{(n)}(z) - F_{\mu_s}^{(n)}(z) &= F_{\mu_t}^{(n)} - F_{\mu_t}^{(n)} \left(z + \Phi_{\mu_{s,t}}^{(n)} \circ F_{\mu_s}^{(n)}(z) \right) \\ &= -\Delta F_{\mu_t}^{(n)}(z, \Phi_{\mu_{s,t}}^{(n)} \circ F_{\mu_s}^{(n)}(z)) \left[\Phi_{\mu_{s,t}}^{(n)} \circ F_{\mu_s}^{(n)}(z) \right].\end{aligned}$$

Since $\Phi_{\mu_{s,t}}^{(n)}$ is given by the formula (1), it extends to a function on the entire upper half-plane. Moreover, as long as $\delta > C^{1/2}(t-s)^{1/2}$ and $\operatorname{Im} z \geq 2\delta$, we have

$$\left\| \int_s^t G_{\sigma(\cdot,u)}^{(n)}(z) \, du \right\| \leq \frac{C(t-s)}{2\delta} < \frac{\delta}{2}.$$

Thus,

$$\operatorname{Im}(z + \Phi_{\mu_{s,t}}^{(n)}(z)) = \operatorname{Im}(z) + \operatorname{Im} \int_s^t G_{\sigma(\cdot,u)}^{(n)}(z) \, du \geq 2\delta - \frac{\delta}{2} = \frac{3}{\delta} 2.$$

By analytic continuation, the relation

$$F_{\mu_t}^{(n)}(z) - F_{\mu_s}^{(n)}(z) = -\Delta F_{\mu_t}^{(n)}(z, z + \Phi_{\mu_{s,t}}^{(n)} \circ F_{\mu_s}^{(n)}(z)) \left[\Phi_{\mu_{s,t}}^{(n)} \circ F_{\mu_s}^{(n)}(z) \right], \tag{7.29}$$

extends to $\operatorname{Im} z \geq 2\delta$. Let τ_t be the law such that $F_{\mu_t}^{(n)}(z) = z - \mu_t(X)^{(n)} - G_{\tau_t}^{(n)}(z)$. Note that for z and w in the upper half plane,

$$\begin{aligned}\left\| \Delta F_{\mu_t}^{(n)}(z, w) \right\|_{\#} &= \left\| \operatorname{id} - \Delta G_{\tau_t}^{(n)}(z, w) \right\|_{\#} \\ &\leq 1 + \frac{\|\operatorname{Var}_{\mu_t}[1]\|}{\epsilon_1 \epsilon_2}\end{aligned}$$

whenever $\operatorname{Im} z \geq \epsilon_1$ and $\operatorname{Im} w \geq \epsilon_2$. Setting $w = z + \Phi^{(n)}_{\mu_{s,t}}(z)$ and noting that $\operatorname{Im} w \geq (3/2)\delta$ and $\operatorname{Im} z \geq 2\delta$, we get

$$\left\| \Delta F^{(n)}_{\mu_t}(z, z + \Phi^{(n)}_{\mu_{s,t}} \circ F^{(n)}_{\mu_s}(z)) \right\| \leq 1 + \frac{\|\mathrm{Var}_{\mu_t}(1)\|}{3\delta^2}.$$

Thus,

$$\begin{aligned}\left\| F^{(n)}_{\mu_t}(z) - F^{(n)}_{\mu_s}(z) \right\| &\leq \left(1 + \frac{\|\mathrm{Var}_{\mu_t}(1)\|}{3\delta^2}\right) \left\| \Phi^{(n)}_{\mu_{s,t}} \circ F^{(n)}_{\mu_s}(z) \right\| \\ &\leq \left(1 + \frac{\|\mathrm{Var}_{\mu_t}(1)\|}{3\delta^2}\right) \left(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + \frac{C}{2\delta} \right) (t - s).\end{aligned}$$

Note that $\|\mathrm{Var}_{\mu_t}[1]\| \leq \|\mathrm{Var}_{\mu_T}[1]\|$ is uniformly bounded. If we fix δ and z with $\operatorname{Im} z \geq 2\delta$, then the above Lipschitz estimate holds for $|s - t|$ sufficiently small, which implies that it actually holds for all $s \leq t$ in $[0, T]$ that

$$\left\| F^{(n)}_{\mu_t}(z) - F^{(n)}_{\mu_s}(z) \right\| \leq \left(1 + \frac{\|\mathrm{Var}_{\mu_T}(1)\|}{3\delta^2}\right) \left(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + \frac{C}{2\delta} \right) (t - s).$$

Thus, we have a uniform Lipschitz estimate for $z \geq 2\delta$ for each $\delta > 0$, so $(F_{\mu_t})_{t \in [0,T]}$ is a locally Lipschitz family.

Now to establish the differential equation, observe that since $F^{(n)}_{\mu_t}$ is a locally Lipschitz family, if we fix $z \in \mathbb{H}^{(n)}_+(\mathcal{B})$, then $\partial_t F^{(n)}_{\mu_t}(z)$ and $-DF^{(n)}_{\mu_t}(z)[b(t) + G^{(n)}_{\sigma(\cdot,t)} \circ F^{(n)}_{\mu_t}(z)]$ are well-defined elements of $\mathcal{L}(L^1[0,T], M_n(\mathcal{B}))$. Thus, to show that they are equal, it suffices to show that for a fixed z, for all $s \leq t$ in $[0, T]$,

$$\int_s^t \partial_u F^{(n)}_{\mu_u}(z)\, du = \int_s^t -DF^{(n)}_{\mu_u}(z)[b(u) + G^{(n)}_{\sigma(\cdot,u)} \circ F^{(n)}_{\mu_u}(z)]\, du + o(|s - t|).$$

The left hand side is simply $F^{(n)}_{\mu_t}(z) - F^{(n)}_{\mu_s}(z)$, which we already know by (7.29) is equal to $-\Delta F^{(n)}_{\mu_t}(z, z + \Phi^{(n)}_{\mu_{s,t}} \circ F^{(n)}_{\mu_s}(z)) \left[\Phi^{(n)}_{\mu_{s,t}} \circ F^{(n)}_{\mu_s}(z) \right]$, which we will now approximate by

$$\int_s^t -DF^{(n)}_{\mu_u}(z)[b(u) + G^{(n)}_{\sigma(\cdot,u)} \circ F^{(n)}_{\mu_u}(z)]\, du$$

by swapping out one piece at a time. Now if $u \in [s, t]$, then

$$\begin{aligned}-\Delta F_{\mu_t}(z, z + \Phi^{(n)}_{\mu_{s,t}} \circ F^{(n)}_{\mu_s}(z)) &= -\Delta F^{(n)}_{\mu_t}(z, z) + O(t - s) \\ &= -DF^{(n)}_{\mu_u}(z) + O(t - s).\end{aligned}$$

Next, since $F^{(n)}_{\mu_s}(z) = F^{(n)}_{\mu_u}(z) + O(t - s)$, we have

$$\left\| u \mapsto G^{(n)}_{\sigma(\cdot,u)} \circ F^{(n)}_{\mu_s} - G^{(n)}_{\sigma(\cdot,u)} \circ F^{(n)}_{\mu_u} \right\|_{\mathcal{L}(L^1[0,T], M_n(\mathcal{B}))} = O(t - s).$$

Therefore, overall we have

$$\int_s^t \Delta F_{\mu_t}^{(n)}(z, z + \Phi_{\mu_{s,t}}^{(n)} \circ F_{\mu_s}^{(n)}(z)) \left[b(u)^{(n)} + G_{\sigma(\cdot,u)}^{(n)} \circ F_{\mu_s}^{(n)}(z)\right] du$$
$$= \int_s^t DF_{\mu_u}^{(n)}(z) \left[b(u)^{(n)} + G_{\sigma(\cdot,u)}^{(n)} \circ F_{\mu_u}^{(n)}(z)\right] du + O((t-s)^2),$$

which means that

$$F_{\mu_t}^{(n)}(z) - F_{\mu_s}^{(n)}(z) = \int_s^t DF_{\mu_u}^{(n)}(z) \left[b(u)^{(n)} + G_{\sigma(\cdot,u)}^{(n)} \circ F_{\mu_u}^{(n)}(z)\right] du,$$

which proves (2).

(3) Observe that $F_{\mu_t}^{-1}(z) = z + \Phi_{\mu_t}(z) = z + \int_0^t [b(s) + G_{\sigma(\cdot,s)}(z)]\, ds$ on its domain, and hence $F_{\mu_t}^{-1}(z)$ is a locally Lipschitz family of fully matricial functions on $\bigcup_{\delta > 2C^{1/2}T^{1/2}} \{\mathrm{Im}(z) \geq \delta\}$. Moreover, on this domain

$$(DF_{\mu_t}(z))^{-1} = D(F_{\mu_t})^{-1}(F_{\mu_t}(z)).$$

Therefore, $b + G_\sigma(z)$ is uniquely determined on this domain by

$$b(t) + G_{\sigma(\cdot,t)}(z) = -D(F_{\mu_t})^{-1}(F_{\mu_t}(z))[(\partial_t F_{\mu_t})(F_{\mu_t}^{-1}(z)]$$

(here we rely on Remark 7.3.6). By analytic continuation, $b(t) + G_{\sigma(\cdot,t)}(z)$ is determined on the entire upper half-plane, and hence b and σ are determined.

(4) We know from (1) that $\int_s^t G_{\sigma(\cdot,u)}^{(n)}(z)\, du = \Phi_{\mu_{s,t}}^{(n)}(z) - \mu_{s,t}(X)^{(n)}$. Thus, our earlier error estimate (7.26) becomes

$$\left\| \int_s^t G_{\sigma(\cdot,u)}^{(n)}(z)\, du - G_{\tau_{s,t}}^{(n)}(z) \right\| \leq \frac{C}{2\delta^2} \left(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + \frac{C}{\delta} \right) (t-s)^2$$

when $\mathrm{Im}\, z \geq 2\delta$ and $\delta > C^{1/2}(t-s)^{1/2}$. Putting this estimate into the proof of Lemma 7.4.4 will produce the estimate (3); we leave the details to the reader. □

Proposition 7.4.7. *Let $(\mu_t)_{t\in[0,T]}$ be a Lipschitz monotone subordination family, and continue all the notation from above.*

(1) $(F_{\mu_t})_{t\in[0,T]}$ is a locally Lipschitz family of fully matricial functions and

$$\partial_t F_{\mu_t}^{(n)}(z) = -DF_{\mu_t}^{(n)}(z) \left[b(t)^{(n)} + G_{\sigma(\cdot,t)}^{(n)}(z)\right].$$

(2) b and σ are uniquely determined by this equation.

(3) If $\phi : [0,T] \to \mathbb{C}$ is continuous with modulus of continuity ω_ϕ, then

$$\left\| \int_0^T G^{(n)}_{\sigma_{\mathcal{P}}(\cdot,t)}(z)\phi(t)\,dt - \int_0^T G^{(n)}_{\sigma(\cdot,t)}(z)\phi(t)\,dt \right\| \leq \frac{C}{2\delta^2}\left(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + \frac{C}{\delta} \right) \operatorname{mesh}(\mathcal{P}) + \frac{CT}{\delta}\omega_\phi(\operatorname{mesh}(\mathcal{P})),$$

whenever $\operatorname{Im} z \geq \delta$.

Proof. (1) First, we show that $(F_{\mu_t})_{t\in[0,T]}$ is a locally Lipschitz family of fully matricial functions. Note that

$$\begin{aligned} F^{(n)}_{\mu_t}(z) - F^{(n)}_{\mu_s}(z) &= F^{(n)}_{\mu_s} \circ F^{(n)}_{\mu_{s,t}}(z) - F^{(n)}_{\mu_s} \\ &= \Delta F^{(n)}_{\mu_s}(F^{(n)}_{\mu_{s,t}}(z), z)[F^{(n)}_{\mu_{s,t}}(z) - z] \\ &= -\Delta F^{(n)}_{\mu_s}(F^{(n)}_{\mu_{s,t}}(z), z)[\mu_{s,t}(X)^{(n)} + G^{(n)}_{\tau_{s,t}}(z)]. \end{aligned}$$

Now since $\operatorname{Im} z \geq \delta$ and $\operatorname{Im} z F^{(n)}_{\mu_{s,t}}(z) \geq \delta$, we have

$$\left\| \Delta^{(n)}_{\mu_s}(F^{(n)}_{\mu_{s,t}}(z), z) \right\|_{\#} \leq 1 + \frac{\|\operatorname{Var}_{\mu_s}[1]\|}{\delta^2}.$$

Then plugging in our usual a priori bounds for $\mu_{s,t}(X)^{(n)}$ and $G^{(n)}_{\tau_{s,t}}(z)$, we get

$$\left\| F^{(n)}_{\mu_t}(z) - F^{(n)}_{\mu_s}(z) \right\| \leq \left(1 + \frac{\|\operatorname{Var}_{\mu_s}[1]\|}{\delta^2} \right) \left(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + \frac{C}{\delta} \right)(t-s).$$

Thus, $(F_{\mu_t})_{t\in[0,T]}$ is a locally Lipschitz family.

Next, we prove the differential equation. Since $(F_{\mu_t})_{t\in[0,T]}$ is a locally Lipschitz family, it suffices to show that for each fixed z, and for all $s \leq t$ in $[0,T]$,

$$F^{(n)}_{\mu_t}(z) - F^{(n)}_{\mu_s}(z) = -\int_s^t DF^{(n)}_{\mu_u}(z)\left[b(u) + G^{(n)}_{\sigma(\cdot,u)}(z) \right] du + o(t-s).$$

The left hand side is equal to $-\Delta F^{(n)}_{\mu_s}(F^{(n)}_{\mu_{s,t}}(z), z)[\mu_{s,t}(X)^{(n)} + G^{(n)}_{\tau_{s,t}}(z)]$. It follows from Lemma 7.4.3 that

$$G^{(n)}_{\tau_{s,t}}(z) = \int_s^t G^{(n)}_{\sigma(\cdot,u)}(z)\,du + O((t-s)^2),$$

so that

$$\begin{aligned} F^{(n)}_{\mu_t}(z) - F^{(n)}_{\mu_s}(z) &= -\Delta F^{(n)}_{\mu_s}(F^{(n)}_{\mu_{s,t}}(z), z) \int_s^t \left[b(u)^{(n)} + G^{(n)}_{\sigma(\cdot,u)}(z) \right] du + O((t-s)^2) \\ &= -\int_s^t \Delta F^{(n)}_{\mu_s}(F^{(n)}_{\mu_{s,t}}(z), z)\left[b(u)^{(n)} + G^{(n)}_{\sigma(\cdot,u)}(z) \right] du + O((t-s)^2). \end{aligned}$$

It is straightforward to check that $\Delta F_{\mu_s}^{(n)}(F_{\mu_{s,t}}^{(n)}(z), z) = DF_{\mu_u}^{(n)}(z) + O(t-s)$ for all $u \in [s,t]$. Then plugging this into our earlier equation shows that

$$F_{\mu_t}^{(n)}(z) - F_{\mu_s}^{(n)}(z) = -\int_s^t DF_{\mu_u}^{(n)}(z) \left[b(u)^{(n)} + G_{\sigma(\cdot,u)}^{(n)}(z) \right] du + O((t-s)^2),$$

which finishes the proof of the differential equation.

(2) Similar to the free case, note that $(DF_{\mu_t}(z))^{-1}$ is a continuous function of t for $\operatorname{Im} z$ sufficiently large and hence $b(t) + G_{\sigma(\cdot,t)}$ is uniquely determined by the differential equation.

(3) This follows from substituting the explicit estimate (7.28) in the proof of Lemma 7.4.4 and computing all the errors explicitly. We leave the details to the reader. □

Proposition 7.4.8. *Let $(\mu_t)_{t\in[0,T]}$ be a Lipschitz anti-monotone subordination family, and continue all the notation from above.*

(1) $(F_{\mu_t})_{t\in[0,T]}$ is a locally Lipschitz family of fully matricial functions and

$$\partial_t F_{\mu_t}^{(n)}(z) = -[b(t)^{(n)} + G_{\sigma(\cdot,t)}^{(n)}(F_{\mu_t}^{(n)}(z))].$$

(2) b and σ are uniquely determined by this equation.

(3) The same estimate as in Proposition 7.4.7 (3) holds.

Proof. (1) First, to show that $(F_{\mu_t})_{t\in[0,T]}$ is a locally Lipschitz family, suppose that $s \leq t$.

$$\begin{aligned} F_{\mu_t}^{(n)}(z) - F_{\mu_s}^{(n)}(z) &= F_{\mu_{s,t}}^{(n)} \circ F_{\mu_s}^{(n)}(z) - F_{\mu_s}^{(n)}(z) \\ &= -\mu_{s,t}(X)^{(n)} - G_{\tau_{s,t}}^{(n)} \circ F_{\mu_s}^{(n)}(z). \end{aligned}$$

Hence,

$$\left\| F_{\mu_t}^{(n)}(z) - F_{\mu_s}^{(n)}(z) \right\| \leq \left(\|b\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + \frac{C}{\delta} \right)(t-s).$$

To check the differential equation, we use similar reasoning to the free and monotone cases to justify the string of approximate equalities

$$\begin{aligned} F_{\mu_t}^{(n)}(z) - F_{\mu_s}^{(n)}(z) &= -\mu_{s,t}(X)^{(n)} - G_{\tau_{s,t}}^{(n)} \circ F_{\mu_s}^{(n)}(z) \\ &= -\int_s^t \left[b(u)^{(n)} + G_{\sigma(\cdot,u)}^{(n)} \circ F_{\mu_s}^{(n)}(z) \right] du + O((t-s)^2) \\ &= -\int_s^t \left[b(u)^{(n)} + G_{\sigma(\cdot,u)}^{(n)} \circ F_{\mu_u}^{(n)}(z) \right] du + O((t-s)^2). \end{aligned}$$

(2) Recall that $F_{\mu_t}^{-1}$ maps $\{\operatorname{Im} z \geq 2\delta\}$ into $\{\operatorname{Im} z \geq \delta\}$ for $\delta > C^{1/2}T^{1/2}$. Moreover, $(F_{\mu_t}^{-1})_{t\in[0,T]}$ is a locally Lipschitz family on this domain; this follows by observing that for $s \leq t$,

$$F_{\mu_t}^{-1}(z) - F_{\mu_s}^{-1}(z) = (F_{\mu_{s,t}}^{-1} - \mathrm{id}) \circ F_{\mu_s}^{-1}(z),$$

and then applying Lemma (4.8) to $F_{\mu_{s,t}}$. Since $F_{\mu_t}^{-1}$ is a locally Lipschitz family, the equation

$$b(t) + G_{\sigma(\cdot,t)}(z) = [\partial_t F_{\mu_t}](F_{\mu_t}^{-1}(z))$$

makes sense and is true for z with $\operatorname{Im} z > 2\delta > 2C^{1/2}T^{1/2}$ (again relying on Remark 7.3.6). This uniquely determines b and σ by analytic continuation as in the free case.

(3) This is the same as the monotone case. □

CHAPTER 8

Results: Fock space realization of subordination families

8.1 Introduction

In the last chapter, we showed that for any Lipschitz process $(X_t)_{t\in[0,T]}$ with independent increments, the F-transforms F_{X_t} satisfied a certain differential equation with respect to the Cauchy-Stieltjes transforms G_σ, where σ is a distributional family of generalized laws. Our goal in this chapter is to prove conversely that for every such σ, there is a process with independent increments that satisfies the equation.

It is possible to construct the transforms F_{X_t} purely analytically by solving the differential equation; see e.g. [Bau05, Theorem 5.5] and [Jek20, §5.3] for Loewner chains (the monotone case) or [Wil17, Corollary 5.2] for the free case. However, we will take a different approach here. Starting with a distributional family of generalized laws σ, we will directly construct operators $(X_t)_{t\in[0,T]}$ whose F-transforms satisfy the differential equation (see Theorem 8.5.1). These operators X_t act on a certain C*-correspondence known as a Fock space.

While the term "Fock space" does not a have a precise definition, it refers to a direct sum of tensor powers of some basic building block, possibly with a few tweaks unique to a given situation. Besides being used in physics, Fock spaces provide a canonical way to construct operators realizing a certain non-commutative law and prove facts about non-commutative independence (see e.g. [Voi86]). In particular, they provide a canonical way to construct Brownian motion and more generally Lévy processes for free, Boolean, and monotone independence. Prior work on free, boolean, and monotone Fock spaces is listed in Table 8.1.

We aim to generalize prior work from semigroups to subordination families. The Fock space realization for monotone subordination families was given by the author in [Jek20, §6],

	boolean	free	(anti-)monotone
scalar-valued		[Voi86] [GSS92]	[Mur97] [Lu97]
operator-valued	[PV13, §2]	[Spe98, §4.7] [PV13, §3.2]	[Jek20, §6]
	[ABF13, §7]	[ABF13, Rem. 7.7]	

Table 8.1: References on boolean, free, and monotone Fock spaces.

while the analogous realizations for free and Boolean subordination families can be deduced easily from previously known results. We will present the constructions for the four types of independence in parallel.

The basic building block for our Fock space is obtained directly from the distributional family of generalized laws σ by a version of the GNS construction.

Definition 8.1.1. Let $\mathcal{B}$ be a unital C*-algebra, and let $\sigma : \mathcal{B}\langle Y \rangle \times L^1[0,T] \to \mathcal{B}$ be a distributional family of generalized laws. We construct a right Hilbert $\mathcal{B}$-module $\int_\oplus \mathcal{B}\langle Y \rangle \otimes_{\sigma(\cdot,t)} \mathcal{B}\, dt$ as follows.

Let $\mathcal{L} \subseteq L^\infty[0,T]$ be the algebra of simple functions. Define a pre-inner product on $\mathcal{L} \otimes_{\text{alg}} \mathcal{B}\langle Y \rangle \otimes_{\text{alg}} \mathcal{B}$ by

$$\langle f_1 \otimes p_1(Y) \otimes b_1, f_2 \otimes p_2(Y) \otimes b_2 \rangle = \int_0^T \overline{f_1}(t) f_2(t) b_1^* \sigma(p_1(Y)^* p_2(Y), t) b_2 \, dt.$$

To check nonnegativity of this pre-inner product, suppose that $\zeta = \sum_{j=1}^m f_j \otimes p_j(Y) \otimes b_j$. Choose a measurable partition $\{E_k\}_{k=1}^n$ of $[0,T]$ such that $f_j|_{E_k} = \alpha_{j,k} \in \mathbb{C}$. Then

$$\langle \zeta, \zeta \rangle = \sum_{k=1}^n \left\langle \sum_{j=1}^m \alpha_{j,k} p_j(Y) \otimes b_j, \sum_{j=1}^m \alpha_{j,k} p_j(Y) \otimes b_j \right\rangle_{\int_{E_k} \sigma},$$

where the last inner product is the one coming from $\mathcal{B}\langle Y \rangle \otimes_{\int_{E_k} \sigma} \mathcal{B}$ and where $\int_{E_k} \sigma = \int \sigma(\cdot, t) \chi_{E_k}(t)\, dt$. We already showed in the proof of Theorem 2.6.6 that $\langle \cdot, \cdot \rangle_{\int_{E_k} \sigma}$ is nonnegative. Thus, our pre-inner-product on $\mathcal{L} \otimes_{\text{alg}} \mathcal{B}\langle Y \rangle \otimes_{\text{alg}} \mathcal{B}$ is nonnegative. We define $\int_\oplus \mathcal{B}\langle Y \rangle \otimes_{\sigma(\cdot,t)} \mathcal{B}\, dt$ as a right Hilbert $\mathcal{B}$-module to be the separation-completion with respect to this pre-inner-product.

Lemma 8.1.2. *Left multiplication by Y defines a bounded operator on $\int_\oplus \mathcal{B}\langle Y \rangle \otimes_{\sigma(\cdot,t)} \mathcal{B}\, dt$ with norm bounded by* $\operatorname{rad}(\sigma)$. *Similarly, left multiplication by $b \in \mathcal{B}$ defines a bounded operator on $\int_\oplus \mathcal{B}\langle Y \rangle \otimes_{\sigma(\cdot,t)} dt$, making the latter a $\mathcal{B}$-$\mathcal{B}$ correspondence. Moreover, for measurable $E \subseteq [0,T]$ and for $p(Y) \in \mathcal{B}\langle Y \rangle$, we have*

$$\langle \xi_E, p(Y) \xi_E \rangle = \int_E \sigma(p(Y), t)\, dt,$$

where $\xi_E = \chi_E \otimes 1 \otimes 1$ in $\int_\oplus \mathcal{B}\langle Y \rangle \otimes_{\sigma(\cdot,t)} \mathcal{B}\, dt$.

Proof. Suppose that $\zeta = \sum_{j=1}^m f_j \otimes p_j(Y) \otimes b_j$ in the algebraic tensor product. Choose a partition $\{E_k\}_{k=1}^m$ of $[0,T]$ such that $f_j|_{E_k} = \alpha_{j,k} \in \mathbb{C}$. Let $\zeta_k = \sum_j \alpha_{j,k} p_j(Y) \otimes b_j$. Then we have

$$\begin{aligned}\langle Y\zeta, Y\zeta\rangle &= \sum_{k=1}^{n} \langle Y\zeta_k, Y\zeta_k\rangle_{\int_{E_k}\sigma} \\ &\leq \sum_{k=1}^{n} \operatorname{rad}\left(\int_{E_k}\sigma\right)^2 \langle \zeta_k, \zeta_k\rangle_{\int_{E_k}\sigma} \\ &\leq \operatorname{rad}(\sigma)^2 \sum_{k=1}^{n} \langle \zeta_k, \zeta_k\rangle_{\int_{E_k}\sigma} \\ &= \operatorname{rad}(\sigma)^2 \langle \zeta, \zeta\rangle.\end{aligned}$$

Hence, $\|Y\zeta\| \leq \operatorname{rad}(\sigma)\|\zeta\|$, so multiplication by Y passes to a bounded operator on the separation-completion. The same reasoning works for left multiplication by $b \in \mathcal{B}$. The relation $\langle \xi_E, p(Y)\xi_E\rangle = \int_E \sigma(p(Y), t)\, dt$ is a direct computation. □

In the above construction of $\int_\oplus \mathcal{B}\langle Y\rangle \otimes_{\sigma(\cdot,t)} \mathcal{B}\, dt$, the "time coordinate" is not represented in the $\mathcal{B}$-$\mathcal{B}$-correspondence structure or in the operator Y, only in the different vectors ξ_E. But another way that the time coordinate naturally appears is that there is an action of the algebra $\mathcal{L}$ of simple functions in $L^\infty[0,T]$ on $\int_\oplus \mathcal{B}\langle Y\rangle \otimes_{\sigma(\cdot,t)} \mathcal{B}\, dt$ by left multiplication, and in fact a left multiplication action by $L^\infty[0,T]$ as we will demonstrate in Lemma 8.2.6 below. It will be useful for our purposes in this chapter and the next to better understand this left $L^\infty[0,T]$-action and how it interacts with tensor products.

8.2 Normal $L^\infty(\Omega,\mu)$-actions on C*-correspondences

Definition 8.2.1. Let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ be unital C*-algebras. Let $\mathcal{H}$ be an $\mathcal{A}$-$\mathcal{B}$-correspondence. Then a *left $\mathcal{C}$-action* on $\mathcal{H}$ is a unital $*$-homomorphism $\pi : \mathcal{C} \to B(\mathcal{H})$ such that $\pi(c)a\xi = a\pi(c)\xi$ for $a \in \mathcal{A}$ and $c \in \mathcal{C}$. We will henceforth write $c\xi$ rather than $\pi(c)\xi$ when the meaning is clear, similar to the notation used for correspondences.

Remark 8.2.2. Those readers familiar with C*-algebra theory will note that this is equivalent to an $(\mathcal{A}\otimes_{\max}\mathcal{C})$-$\mathcal{B}$-correspondence structure on $\mathcal{H}$, where $\mathcal{A}\otimes_{\max}\mathcal{C}$ is the maximal C*-tensor product.

Definition 8.2.3. Let $\mathcal{H}$ be an $\mathcal{A}$-$\mathcal{B}$-correspondence, and let (Ω,μ) be a finite measure space (which implicitly includes a choice of σ-algebra). A left $L^\infty(\Omega,\mu)$-action on $\mathcal{H}$ is said to be *normal* if for any sequence $f_n \in L^\infty(\Omega,\mu)$ such that $\|f_n\|_\infty \leq 1$ and $f_n \to f$ in measure, we have $f_n\xi \to f\xi$ for each $\xi \in \mathcal{H}$.

Remark 8.2.4. In the case $\mathcal{B} = \mathbb{C}$, this definition is equivalent to the $*$-homomorphism $L^\infty(\Omega,\mu) \to B(\mathcal{H})$ being continuous in the strong operator topology on the unit ball, which is the standard definition of "normal" in the theory of von Neumann algebras.

Normal left $L^\infty(\Omega,\mu)$-actions can be characterized more concretely in terms of projections as follows.

Lemma 8.2.5. *Let $\mathcal{H}$ be an $\mathcal{A}$-$\mathcal{B}$-correspondence, let (Ω,μ) be a finite measure space with the associated σ-algebra $\mathcal{F}$. Let $\mathcal{E}$ be an algebra of sets that generates $\mathcal{F}$ as a σ-algebra. Suppose that for each $E \in \mathcal{E}$, there is a projection P_E in $B(\mathcal{H})$ that commutes with the left action of $\mathcal{A}$, such that*

(1) $P_E + P_{\Omega\setminus E} = 1$.

(2) $P_{E_1}P_{E_2} = P_{E_1\cap E_2}$ for $E_1, E_2 \in \mathcal{E}$.

(3) For every $\xi \in \mathcal{H}$ and every $\epsilon > 0$, there exists $\delta > 0$ such that for all $E \in \mathcal{E}$, if $\mu(E) < \delta$, then $\|P_E\xi\| < \epsilon$.

Then there is a unique normal left $L^\infty(\Omega,\mu)$-action π on $\mathcal{H}$ such that $\pi(\chi_E) = P_E$ for all $E \in \mathcal{E}$. Conversely, if π is a normal $L^\infty(\Omega,\mu)$-action on $\mathcal{H}$, then $P_E := \pi(\chi_E)$ satisfies (1) - (3) above. Furthermore, for (3) to hold for all ξ, it suffices that it holds on a set of vectors whose span is dense in $\mathcal{H}$.

Proof. First, assume we have a family of projections $\{P_E\}_{E\in\mathcal{E}}$ satisfying (1) - (3). The conditions (1) and (2) imply that

$$\begin{aligned} P_{E_1} + P_{E_2} &= 1 + P_{E_1}P_{E_2} - (1 - P_{E_1})(1 - P_{E_2}) \\ &= 1 + P_{E_1\cap E_2} - P_{\Omega\setminus E_1}P_{\Omega\setminus E_2} \\ &= 1 - P_{\Omega\setminus(E_1\cup E_2)} + P_{E_1\cap E_2} = P_{E_1\cup E_2} + P_{E_1\cap E_2}. \end{aligned}$$

In particular, it follows that if $E_1, \dots, E_n$ are disjoint, then $P_{E_1\cup\cdots\cup E_n} = P_{E_1} + \cdots + P_{E_n}$. Note also that (3) implies that if $\mu(E) = 0$, then $P_E = 0$, and in particular $P_\varnothing = 0$.

Let $\mathcal{L} \subseteq L^\infty(\Omega,\mu)$ be the algebra of functions generated by $\{\chi_E\}_{E\in\mathcal{E}}$, which we may think of as "step functions" relative to the algebra of sets $\mathcal{E}$. Note that $\mathcal{L}$ is the same as the vector span of $\{\chi_E\}_{E\in\mathcal{E}}$ since E is closed under intersections. We can define a linear map $\pi : \mathcal{L} \to B(\mathcal{H})$ by $\pi(\sum_{j=1}^n \alpha_j\chi_{E_j}) = \sum_{j=1}^n \alpha_j P_{E_j}$ when the E_j's are disjoint measurable sets. Indeed, the output is independent of the choice of decomposition for the step function because we showed that the projections are additive under disjoint unions. Moreover, the output is unchanged if we modify E_j by a null set. The additivity relation also implies linearity of π. Since $P_{E_1}P_{E_2} = P_{E_1\cap E_2}$ and $P_E = P_E^*$, we see that π is $*$-homomorphism.

It is also clear that $\|\pi(f)\| \leq \|f\|$; indeed, of $f = \sum_{j=1}^k \alpha\chi_{E_j}$ with the E_j's disjoint with positive measure, then the projections P_{E_j} have orthogonal ranges, and hence $\|\sum_{j=1}^k \alpha_j P_{E_j}\| \leq \max_j |\alpha_j| = \|f\|$.

Next, we claim that for each ξ and $\epsilon > 0$, there exist $\epsilon' > 0$ and $\delta > 0$ such that if $f_1, f_2 \in \mathcal{L}$ with $\|f_1\| \leq 1$ and $\|f_2\| \leq 1$ and if $\mu(\{|f_1 - f_2| > \epsilon'\}) < \delta$, then $\|\pi(f_1)\xi - \pi(f_2)\xi\| < \epsilon$. To prove this, fix ξ and ϵ. Let ϵ' be chosen small enough that $\epsilon'\|\xi\| < \epsilon/2$. By condition (4), there exists δ such that $\mu(E) < \delta$ implies $\|P_E \xi\| < \epsilon/4$. Now suppose that f_1 and f_2 satisfy $\|f_1\| \leq 1$ and $\|f_2\| \leq 1$ and $\mu(\{|f_1 - f_2| > \epsilon'\}) < \delta$. Let $E = \{|f_1 - f_2| > \epsilon'\}$. Then we have

$$\begin{aligned}\|\pi(f_1)\xi - \pi(f_2)\xi\| &= \|\pi[(f_1 - f_2)\chi_E]\xi\| + \|\pi[(f_1 - f_2)(1 - \chi_E)]\xi\| \\ &\leq \|f_1 - f_2\| \|P_E \xi\| + \|(f_1 - f_2)(1 - \chi_E)\| \|\xi\| \\ &\leq 2 \cdot \frac{\epsilon}{4} + \epsilon' \|\xi\| < \epsilon.\end{aligned}$$

This proves our claim. Since $\mathcal{L}$ is dense in $L^\infty(\Omega, \mu)$ with respect to convergence in measure, the claim implies that there is a unique extension of π to a normal $*$-homomorphism $L^\infty(\Omega, \mu) \to B(\mathcal{H})$. Moreover, since $\pi(f)$ commutes with the left action of $\mathcal{A}$ for each $f \in \mathcal{L}$, the same is true for $f \in L^\infty(\Omega, \mu)$.

To prove the converse claim, suppose that π is a normal left $L^\infty(\Omega, \mu)$ action on $\mathcal{H}$. If $P_E = \pi(\chi_E)$, then the relations (1) and (2) are immediate from the corresponding algebraic relations in $L^\infty(\Omega, \mu)$. If we assume for contradiction that (3) does not hold, then there exists ξ and $\epsilon > 0$ and a sequence of sets $E_n \in \mathcal{E}$ with $\mu(E_n) \to 0$ but $\|P_{E_n} \xi\| \geq \epsilon$. But then $\chi_{E_n} \to 0$ in measure, so that contradicts normality.

Finally, we prove the last claim that to verify (3), it suffices to test a set of vectors whose span is dense in $\mathcal{H}$. Let S be the set of vectors ξ such that the condition described in (3) holds. It is straightforward to check that S is closed under addition and scalar multiplication. It remains to check that S is a closed set. Let $\xi \in \overline{S}$ and pick $\epsilon > 0$. There exists $\zeta \in S$ with $\|\xi - \zeta\| < \epsilon/2$. There also exists $\delta > 0$ such that $\mu(E) < \delta$ implies that $\|P_E \zeta\| < \epsilon/2$ and hence $\|P_E \xi\| \leq \|P_E(\xi - \zeta)\| + \|P_E \zeta\| < \epsilon$. □

The first application is to construct a normal left $L^\infty[0, T]$-action on the C*-correspondences from distributional families of generalized laws.

Lemma 8.2.6. *Let $\sigma : \mathcal{B}\langle Y \rangle \times L^1[0, T] \to \mathcal{B}$ be a distributional family of generalized laws. Let $\mathcal{A}$ be the C*-algebra generated by $\mathcal{B}$ and Y acting by left multiplication on $\int \mathcal{B}\langle Y \rangle \otimes_{\sigma(\cdot, t)} \mathcal{B}\, dt$, so that the latter is an $\mathcal{A}$-$\mathcal{B}$-correspondence. The left multiplication action of the algebra $\mathcal{L}$ of simple functions upon $\mathcal{L} \otimes_{\text{alg}} \mathcal{B}\langle Y \rangle \otimes_{\text{alg}} \mathcal{B}$ extends to a unique normal left $L^\infty[0, T]$-action on $\int \mathcal{B}\langle Y \rangle \otimes_{\sigma(\cdot, t)} \mathcal{B}\, dt$, where $[0, T]$ is equipped with the Borel σ-algebra and Lebesgue measure. Moreover, letting $\xi = 1 \otimes 1 \otimes 1$, we have for $p(Y) \in \mathcal{B}\langle Y \rangle$ and $E \subseteq [0, T]$ Borel that*

$$\int_E \sigma(p(Y), t)\, dt = \langle \xi, p(Y) \chi_E \xi \rangle = \langle \xi, p(\chi_E Y) \xi \rangle,$$

and if $p(Y) \in \mathcal{B}\langle Y \rangle_0$, then this also equals $\langle \xi, p(\chi_E Y) \xi \rangle$.

Proof. Suppose that $\zeta \in \mathcal{L} \otimes_{\text{alg}} \mathcal{B}\langle Y \rangle \otimes_{\text{alg}} \mathcal{B}$, and let $E \subseteq [0, T]$ be Borel. Then $\zeta = \chi_E \zeta + \chi_{[0,T] \setminus E} \zeta$. Moreover, one can check that $\chi_E \zeta$ and $\chi_{\Omega \setminus E} \zeta$ are orthogonal with respect to the inner product defined by σ. Thus, $\|\chi_E \zeta\| \leq \|\zeta\|$ and $\|\chi_{\Omega \setminus E} \zeta\|$. This shows that the operator P_E given by multiplication by χ_E is a well-defined operator on the separation-completion. Also, it is immediate to check that $P_E^* = P_E = P_E^2$ and $P_E + P_{[0,T] \setminus E} = 1$ and

$P_{E_1} P_{E_2} = P_{E_1 \cap E_2}$ from evaluating the operators on elements of $\mathcal{L} \otimes_{\text{alg}} \mathcal{B}\langle Y \rangle \otimes_{\text{alg}} \mathcal{B}$. In other words, the projections $\{P_E\}$ satisfy (1) and (2) of Lemma 8.2.5. They also clearly commute with the left action of $\mathcal{A}$.

Now we check (3) on a set of vectors whose span is dense, namely on simple tensors $\zeta = f \otimes p(Y) \otimes b$ where $p(Y) = b_0 Y b_1 \dots Y b_k$ is a monomial. Let $E \subseteq [0, T]$ be Borel. Then we have

$$\langle P_E \zeta, P_E \zeta \rangle = \int b^* \sigma(p(Y)^* p(Y), t) b \chi_E(t) |f(t)|^2 \, dt \leq \|b\|^2 \|b_0\|^2 \dots \|b_k\|^2 \operatorname{rad}(\sigma)^k \int_E |f(t)|^2 \, dt.$$

It is well known that for each $f \in L^2[0, T]$, for any $\epsilon > 0$, there exists $\delta > 0$ such that $|E| < \delta$ implies $\int_E |f|^2 < \epsilon$. This implies that (3) holds for the vector ζ. Therefore, by Lemma 8.2.5, the left multiplication action by simple functions extends to a unique normal left $L^\infty[0, T]$-action.

The formula $\int_E \sigma(p(Y), t) \, dt = \langle \xi, p(Y) \chi_E \xi \rangle$ is immediate from Lemma 8.1.2 since χ_E is a projection and commutes with $p(Y)$. Also, $p(Y) \chi_E = p(\chi_E Y) \chi_E$ for the same reason, and if $p(Y) \in \mathcal{B}\langle Y \rangle_0$, then $p(\chi_E Y) \chi_E = p(\chi_E Y)$. □

Normal $L^\infty(\Omega, \mu)$-actions behave very nicely under tensor products, which will be useful for our construction of the Fock space, especially in the (anti-)monotone case.

Lemma 8.2.7. *Let $\mathcal{A}_0$, ..., $\mathcal{A}_k$ be unital* C**-algebras, let (Ω_1, μ_1), ..., (Ω_k, μ_k) be finite measure spaces, and for $j = 1$, ..., n, let $\mathcal{H}_j$ be an $\mathcal{A}_{j-1}$-$\mathcal{A}_j$-correspondence with a normal left $L^\infty(\Omega_j, \mu_j)$-action. Then the $\mathcal{A}_0$-$\mathcal{A}_k$-correspondence $\mathcal{H} := \mathcal{H}_1 \otimes_{\mathcal{A}_1} \cdots \otimes_{\mathcal{A}_k} \mathcal{H}_k$ has a unique normal left $L^\infty(\Omega_1 \times \dots \times \Omega_k, \mu_1 \otimes \dots \otimes \mu_k)$-action satisfying*

$$(f_1 \otimes \dots \otimes f_k)(\xi_1 \otimes \dots \otimes \xi_k) = f_1 \xi_1 \otimes \dots \otimes f_k \xi_k. \tag{8.1}$$

for $f_j \in L^\infty(\Omega_j, \mu_j)$ and $\xi_j \in \mathcal{H}_j$ for $j = 1, \dots, k$.

Proof. To simplify notation, let us only consider the case $k = 2$. The proof of the general case is the same, or alternatively the general case can be deduced from the $k = 2$ case by induction; we leave the details as an exercise.

Let $\mathcal{E}$ be the algebra of sets generated by the rectangles $F \times G$ where $F \subseteq \Omega_1$ and $G \subseteq \Omega_2$ are measurable. We want to define projections P_E for $E \in \mathcal{E}$ as in the previous lemma. Fix E. We can choose partitions F_1, ..., F_k and G_1, ..., G_ℓ for Ω_1 and Ω_2 respectively such that E is a disjoint union of rectangles of the form $F_i \times G_j$. Since χ_{F_1}, ..., χ_{F_k} are projections on $\mathcal{H}_1$ with orthogonal ranges, we have $\mathcal{H}_1 \cong \bigoplus_{i=1}^k \chi_{F_i} \mathcal{H}_1$ as $\mathcal{A}_0$-$\mathcal{A}_1$-correspondences, and similarly $\mathcal{H}_2 \cong \bigoplus_{j=1}^\ell \chi_{G_j} \mathcal{H}_2$, which implies

$$\mathcal{H}_1 \otimes_{\mathcal{A}_1} \mathcal{H}_2 \cong \bigoplus_{i,j} (\chi_{F_i} \mathcal{H}_1) \otimes_{\mathcal{A}_1} (\chi_{G_j} \mathcal{H}_2).$$

We define $P_{F_i \times G_j}$ to be the projection on the (i,j) summand, and define P_E to be the sum of the $P_{F_i \times G_j}$'s over (i,j) such that $F_i \times G_j \subseteq E$. One can check that the projection P_E thus constructed is independent of the choice of partition and only depends on E, and thus notation we just used for P_E and $P_{F_i \times G_j}$ is consistent.

One can also check that the relations (1) and (2) of Lemma 8.2.5 are satisfied for $E \in \mathcal{E}$. Of course, the idea for (2) is to choose a partition that works for two sets E_1 and E_2 simultaneously and then write everything in terms of the minimal projections associated to this partition. We leave the details to the reader.

Now we check condition (3) of Lemma 8.2.5 for $(\Omega, \mu) = (\Omega_1 \times \Omega_2, \mu_1 \otimes \mu_2)$. As shown in that lemma, it suffices to check it for ξ in a set whose span is dense, and thus we may restrict our attention to simple tensors $\xi = \xi_1 \otimes \xi_2$ with $\|\xi_j\| \leq 1$. Fix $\epsilon > 0$. By applying Lemma 8.2.5 (3) to each $\mathcal{H}_1$ and $\mathcal{H}_2$, we see that there exist $\delta_1, \delta_2 > 0$ such that for measurable $F \subseteq \Omega_1$ and $G \subseteq \Omega_2$, we have

$$\mu_1(F) < \delta_1 \implies \|P_F \xi_1\| < \frac{\epsilon}{\sqrt{2}}, \quad \mu_2(G) < \delta_2 \implies \|P_G \xi_2\| < \frac{\epsilon}{\sqrt{2}}.$$

Let $E \in \mathcal{E}$ with $\mu(E) < \delta_1 \delta_2$. For $\omega_2 \in \Omega_2$, let $E^{\omega_2} = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in E\}$, and then we have

$$\mu(E) = \int_{\Omega_2} \mu_1(E^{\omega_2}) \, d\mu_2(\omega_2).$$

Since this is less than $\delta_1 \delta_2$, we have by the Markov inequality that

$$\mu_2(\{\omega_2 : \mu_1(E^{\omega_2}) \geq \delta_1\}) < \frac{\delta_1 \delta_2}{\delta_1} = \delta_2.$$

Let $G = \{\omega_2 : \mu_1(E^{\omega_2}) \geq \delta_1\}$. Since $E \in \mathcal{E}$, the map $\omega_2 \mapsto E^{\omega_2}$ takes only finitely many values, so there exists a partition of $\Omega_2 \setminus G$ into measurable sets $G_1, \dots, G_k$ and there exist measurable sets $F_1, \dots, F_k$ in Ω_1 such that

$$E \setminus (\Omega_1 \otimes G) = \bigsqcup_{i=1}^k F_i \times G_i,$$

and by definition of G, we have $\mu_1(F_i) < \delta_2$. Then we have

$$\begin{aligned} P_E(\xi_1 \otimes \xi_2) &= P_E P_{\Omega_1 \otimes G}(\xi_1 \otimes \xi_2) + \sum_{i=1}^k P_{F_i \otimes G_i}(\xi_1 \otimes \xi_2) \\ &= P_E(\xi_1 \otimes \chi_G \xi_2) + \sum_{i=1}^k \chi_{F_i} \xi_1 \otimes \chi_{G_i} \xi_2. \end{aligned}$$

Note that the terms on the right hand side are mutually orthogonal, hence we will apply the Pythagorean identity. The first term can be estimated by Lemma 2.3.2 as

$$\|P_E(\xi_1 \otimes \chi_G \xi_2)\| \leq \|\xi_1 \otimes \chi_G \xi_2\| \leq \|\xi_1\| \|\chi_G \xi_2\| < \frac{\epsilon}{2}$$

since $\mu_2(G) < \delta_2$. Furthermore, we have

$$\begin{aligned}\langle \chi_{F_i}\xi_1 \otimes \chi_{G_i}\xi_2, \chi_{F_i}\xi_1 \otimes \chi_{G_i}\xi_2 \rangle &\leq \|\chi_{F_i}\xi_1\|^2 \langle \chi_{G_i}\xi_2, \chi_{G_i}\xi_2 \rangle \\ &\leq \frac{\epsilon^2}{2} \langle \chi_{G_i}\xi_2, \chi_{G_i}\xi_2 \rangle.\end{aligned}$$

Hence, we have

$$\left\langle \sum_{i=1}^{k} \chi_{F_i}\xi_1 \otimes \chi_{G_i}\xi_2, \sum_{i=1}^{k} \chi_{F_i}\xi_1 \otimes \chi_{G_i}\xi_2 \right\rangle \leq \frac{\epsilon^2}{2} \sum_{i=1}^{k} \langle \chi_{G_i}\xi_2, \chi_{G_i}\xi_2 \rangle = \frac{\epsilon^2}{2} \langle \chi_{\Omega_2 \setminus G}\xi_2, \chi_{\Omega_2 \setminus G}\xi_2 \rangle,$$

so that

$$\left\| \sum_{i=1}^{k} \chi_{F_i}\xi_1 \otimes \chi_{G_i}\xi_2 \right\| \leq \frac{\epsilon}{\sqrt{2}} \|\chi_{\Omega_2 \setminus G}\xi_2\| \leq \frac{\epsilon}{\sqrt{2}} \|\xi_2\| \leq \frac{\epsilon}{\sqrt{2}}.$$

And overall,

$$\|P_E(\xi_1 \otimes \chi_G\xi_2)\|^2 \leq \|\xi_1 \otimes \chi_G\xi_2\|^2 + \left\| \sum_{i=1}^{k} \chi_{F_i}\xi_1 \otimes \chi_{G_i}\xi_2 \right\|^2 < \frac{\epsilon^2}{2} + \frac{\epsilon^2}{2} = \epsilon^2.$$

Thus, we have checked (3) of Lemma 8.2.5 for $\xi = \xi_1 \otimes \xi_2$.

Therefore, there exists a normal left $L^\infty(\Omega, \mu)$-action π with $\pi(\chi_E) = P_E$ for $E \in \mathcal{E}$. It is easy to check that π satisfies

$$\pi(f_1 \otimes f_2)(\xi_1 \otimes \xi_2) = f_1\xi_1 \otimes f_2\xi_2$$

when f_1 and f_2 are simple functions. This extends to all f_1 and f_2 by taking limits with respect to convergence in measure (where we also rely on Lemma 2.3.2 to take the limit on the right hand side). The normal $L^\infty(\Omega, \mu)$-action satisfying $\pi(f_1 \otimes f_2)(\xi_1 \otimes \xi_2) = f_1\xi_1 \otimes f_2\xi_2$ is unique because this relation determines $\pi(\chi_E)$ for $E \in \mathcal{E}$, which in turn uniquely determines π by Lemma 8.2.5. □

8.3 The Fock space

Let $\mathcal{H}$ be a $\mathcal{B}$-$\mathcal{B}$-correspondence with a normal $L^\infty[0, T]$-action. We define

$$\begin{aligned}E_{\text{mono}}(k, T) &= \{(t_1, \ldots, t_k) \in [0, T]^k : t_1 > t_2 > \cdots > t_k\} \\ E_{\text{mono}\,\dagger}(k, T) &= \{(t_1, \ldots, t_k) \in [0, T]^k : t_1 < t_2 < \cdots < t_k\}.\end{aligned}$$

Then we define the *Fock spaces*

$$\begin{aligned}
\mathcal{F}_{\text{bool}}(\mathcal{H}) &= \mathcal{B} \oplus \mathcal{H} \\
\mathcal{F}_{\text{free}}(\mathcal{H}) &= \mathcal{B} \oplus \bigoplus_{k=1}^{\infty} \underbrace{\mathcal{H} \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{H}}_{k} \\
\mathcal{F}_{\text{mono}}(\mathcal{H}) &= \mathcal{B} \oplus \bigoplus_{k=1}^{\infty} \chi_{E_{\text{mono}}(k,T)}[\underbrace{\mathcal{H} \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{H}}_{k}] \\
\mathcal{F}_{\text{mono}\,\dagger}(\mathcal{H}) &= \mathcal{B} \oplus \bigoplus_{k=1}^{\infty} \chi_{E_{\text{mono}\,\dagger}(n,T)}[\underbrace{\mathcal{H} \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{H}}_{k}].
\end{aligned}$$

Here the multiplication by $\chi_{E_{\text{ind}}(k,T)}$ for the monotone and anti-monotone cases is performed through the normal $L^\infty([0,T]^k)$-action on $\mathcal{H}^{\otimes_{\mathcal{B}} k}$ defined by Lemma 8.2.7; it is easy to check that the image of multiplication by $\chi_{E_{\text{ind}}(k,T)}$ on $\mathcal{H}^{\otimes_{\mathcal{B}} k}$ is a $\mathcal{B}$-$\mathcal{B}$-correspondence, and in fact a direct summand of the whole tensor product.

To simplify notation in the future, we will write

$$\mathcal{H}^{\otimes_{\mathcal{B}} k} = \begin{cases} \mathcal{B}, & k = 0, \\ \underbrace{\mathcal{H} \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{H}}_{n}, & n > 0 \end{cases}$$

as a $\mathcal{B}$-$\mathcal{B}$-correspondence. Then we also define

$$\begin{aligned}
E_{\text{bool}}(k,T) &= \begin{cases} [0,T], & k = 1 \\ \varnothing, & k > 1 \end{cases} \\
E_{\text{free}}(k,T) &= [0,T]^k.
\end{aligned}$$

Thus, the definition of the Fock spaces can be written more compactly as

$$\mathcal{F}_{\text{ind}} = \bigoplus_{k=0}^{\infty} \chi_{E_{\text{ind}}(k,T)} \mathcal{H}^{\otimes_{\mathcal{B}} n},$$

where we make the convention that $\chi_{E_{\text{ind}}(0,T)} = 1$. Furthermore, for each type of independence, we can define an operator P_{ind} on $\mathcal{F}_{\text{free}}$ as the direct sum of the operators of multiplication by $\chi_{E_{\text{ind}}(0,T)}$ on $\mathcal{H}_{\otimes_{\mathcal{B}} k}$ for $k \geq 0$. Then we have

$$\mathcal{F}_{\text{ind}} = P_{\text{ind}}[\mathcal{F}_{\text{free}}].$$

8.3.1 Creation and annihilation operators

For $\zeta \in \mathcal{H}$, we define the *creation operator* $\ell_{\text{ind}}(\zeta)$ on $\mathcal{F}_{\text{ind}}$ by

$$\begin{aligned}
\ell_{\text{ind}}(\zeta)[b] &= \chi_{E_{\text{ind}}(1,T)} \zeta \text{ for } b \in \mathcal{B} \\
\ell_{\text{ind}}(\zeta)[\chi_{E_{\text{ind}}(k,T)}(\zeta_1 \otimes \cdots \otimes \zeta_k)] &= \chi_{E_{\text{ind}}(k+1,T)}[\zeta \otimes \zeta_1 \otimes \cdots \otimes \zeta_k],
\end{aligned}$$

and the *annihilation operator* $\ell_{\text{ind}}(\zeta)^*$ will be its adjoint. In order to verify that $\ell_{\text{ind}}(\zeta)$ is well-defined, bounded, and adjointable, it is easiest to start with the free case because the $\chi_{E_{\text{ind}}(n,T)}$ terms equal 1. If we take η and η' in the k-fold algebraic tensor product of $\mathcal{H}$, then we have

$$\langle \ell_{\text{free}}(\zeta)\eta, \ell_{\text{free}}(\zeta)\eta' \rangle = \langle \eta, \langle \zeta, \zeta \rangle \eta' \rangle.$$

In fact, the same holds for η and η' in the algebraic direct sum of the k-fold algebraic tensor products of $\mathcal{H}$; note that the separation-completion of this "algebraic Fock space" is canonically isomorphic to $\mathcal{F}_{\text{free}}$. This implies that $\|\ell_{\text{free}}(\zeta)\eta\| \leq \|\zeta\| \|\eta\|$, so that $\ell_{\text{free}}(\zeta)$ passes to a well-defined bounded operator on $\mathcal{F}_{\text{free}}(\mathcal{H})$. Next, we claim that that $\ell_{\text{free}}(\zeta)$ is adjointable with

$$\ell_{\text{free}}(\zeta)^* b = 0 \text{ for } b \in \mathcal{B}.$$
$$\ell_{\text{free}}(\zeta)^*[\zeta_1 \otimes \cdots \otimes \zeta_k] = \langle \zeta, \zeta_1 \rangle \zeta_2 \otimes \zeta_3 \otimes \cdots \otimes \zeta_k.$$

A direct computation shows that if we define the operator $\ell_{\text{free}}(\zeta)^*$ on the algebraic direct sum of algebraic tensor products by the above formula (with a slight abuse of notation), then

$$\langle \ell_{\text{free}}(\zeta)^* \eta, \eta' \rangle = \langle \eta, \ell_{\text{free}}(\zeta)\eta \rangle.$$

But since $\|\ell_{\text{free}}(\zeta)^*\eta\| = \sup\{\|\langle \ell_{\text{free}}(\zeta)^*\eta, \eta' \rangle\| : \|\eta'\| \leq 1\}$, we see that $\|\ell_{\text{free}}(\zeta)^*\eta\| \leq \|\ell_{\text{free}}(\zeta)\| \|\eta\|$, hence $\ell_{\text{free}}(\zeta)^*$ passes to a well-defined bounded operator on the separation-completion, and it is the adjoint of $\ell_{\text{free}}(\zeta)$.

Now for the other types of independence, using the identification $\mathcal{F}_{\text{ind}}(\mathcal{H}) = P_{\text{ind}}[\mathcal{F}_{\text{free}}(\mathcal{H})]$, we define

$$\ell_{\text{ind}}(\zeta) = P_{\text{ind}} \ell_{\text{free}}(\zeta) P_{\text{ind}},$$

and one can verify that $\ell_{\text{ind}}(\zeta)[\zeta_1 \otimes \cdots \otimes \zeta_k]$ has the formula we asserted at the beginning. (Also, the relation $\ell_{\text{free}}(\zeta) = P_{\text{free}} \ell_{\text{free}}(\zeta) P_{\text{free}}$ holds vacuously in the free case as well.) It follows that $\ell_{\text{ind}}(\zeta)$ is bounded and adjointable with adjoint given by $P_{\text{ind}} \ell_{\text{free}}(\zeta)^* P_{\text{ind}}$.

8.3.2 Multiplication operators of the first kind

Let us denote

$$L^\infty[0,T]' \cap B(\mathcal{H}) = \{z \in B(\mathcal{H}) : zf\zeta = fz\zeta \text{ for } \zeta \in \mathcal{H}, f \in L^\infty[0,T]\},$$

that is, the commutant in $B(\mathcal{H})$ of the $L^\infty[0,T]$-action on $\mathcal{H}$. For $z \in L^\infty[0,T]' \cap B(\mathcal{H})$, we define the *multliplication operator* $\mathfrak{m}_{\text{ind}}(z)$ on $\mathcal{F}_{\text{ind}}(\mathcal{H})$ as follows. Note that $\mathcal{H}$ is a $B(\mathcal{H})$-$\mathcal{B}$-correspondence by definition. Hence, if $k \geq 1$, then $\mathcal{H}^{\otimes_k \mathcal{B}}$ has the structure of a $B(\mathcal{H})$-$\mathcal{B}$-correspondence. We define $\mathfrak{m}_{\text{free}}(z)$ to be the operator on $\mathcal{F}_{\text{free}}(\mathcal{H})$ obtained as the direct sum of the left multiplication by z on $\mathcal{H}^{\otimes_{\mathcal{B}} k}$ for $k \geq 1$ and the zero operator for $k = 0$, that is,

$$\mathfrak{m}_{\text{free}}(T) b = 0 \text{ for } b \in \mathcal{B}$$
$$\mathfrak{m}_{\text{free}}(T)[\zeta_1 \otimes \cdots \otimes \zeta_k] = T\zeta_1 \otimes \zeta_2 \otimes \cdots \otimes \zeta_k.$$

Then we define $\mathfrak{m}_{\text{ind}}(z) = P_{\text{ind}}\mathfrak{m}_{\text{free}}(z)P_{\text{ind}}$ for each of the other types of independence.

We claim that $\mathfrak{m}_{\text{ind}}$ is a $*$-homomorphism $L^\infty[0,T]' \cap B(\mathcal{H}) \to B(\mathcal{F}_{\text{ind}}(\mathcal{H}))$ (and this is the reason that we restricted to operators that commute with the $L^\infty[0,T]$ action). First, note that z commutes with the $L^\infty([0,T]^k)$-action on $\mathcal{H}^{\otimes_{\mathcal{B}} k}$. This is because z commutes with the action of χ_E when $E = E_1 \times \cdots \times E_k$, hence for general E in the algebra of sets generated by rectangles. Then by an approximation argument, it commutes with the action of all $L^\infty([0,T]^k)$ functions. In particular, it commutes with $\chi_{E_{\text{ind}}(k,T)}$, and therefore, $\mathfrak{m}_{\text{free}}(z)$ commutes with P_{ind}. Therefore, for two such operators z_1 and z_2, we have

$$[P_{\text{ind}}\mathfrak{m}_{\text{free}}(z_1)P_{\text{ind}}][P_{\text{ind}}\mathfrak{m}_{\text{free}}(z_2)P_{\text{ind}}] = P_{\text{ind}}\mathfrak{m}_{\text{free}}(z_1)\mathfrak{m}_{\text{free}}(z_2)P_{\text{ind}} = P_{\text{ind}}\mathfrak{m}_{\text{free}}(z_1 z_2)P_{\text{ind}},$$

since $\mathfrak{m}_{\text{free}}$ is clearly a $*$-homomorphism.

However, we remark that the restriction to $L^\infty[0,T]' \cap B(\mathcal{H})$ is only really necessary in the (anti-)monotone case; in the free and boolean cases, $\mathfrak{m}$ is a $*$-homomorphism on all of $B(\mathcal{H})$.

8.3.3 Multiplication operators of the second kind

The last type of operator we will define requires more casework. The purpose of these operators in our main theorem concerning processes with independent increments (Theorem 8.5.1 below) will be to model the "drift" term $b(t)$ in the expression $b(t) + G_{\sigma(\cdot,t)}(z)$.

A bounded linear functional $\beta : L^\infty[0,T] \to \mathcal{B}$ is said to be *normal* if whenever $(f_n)_{n\in\mathbb{N}}$ is a bounded sequence in $L^\infty[0,T]$ and $f_n \to f$ in measure, we have $\beta[f_n] \to \beta[f]$. If β is a normal bounded linear functional, then we define a multiplication operator $\mathfrak{n}_{\text{ind}}(\beta)$ on $\mathcal{F}_{\text{ind}}$ as follows for each of the four types of independence.

Let ξ be the vector 1 in $\mathcal{B} \subseteq \mathcal{F}_{\text{ind}}$, and let P_ξ be the projection onto $\mathcal{B}\xi$.

(1) We define $\mathfrak{n}_{\text{bool}}(\beta) := \beta[1]P_\xi$, that is, $\mathfrak{n}_{\text{bool}}(\beta)$ is left multiplication by $\beta[1]$ on the direct summand $\mathcal{B}$ and zero on the direct summand $\mathcal{H}$ of $\mathcal{F}_{\text{bool}}$.

(2) We define $\mathfrak{n}_{\text{free}}(\beta)$ to be left multiplication by $\beta[1]$, defined using the $\mathcal{B}$-$\mathcal{B}$-correspondence structure.

(3) We define $\mathfrak{n}_{\text{mono}}(\beta)$ as follows. Let $\phi(t) = \beta[\chi_{[t,T]}]$. Because β is normal, the function $\phi : [0,T] \to \mathcal{B}$ is continuous. Then define $z_n \in B(\mathcal{H})$ by

$$z_n\zeta = \sum_{j=0}^{n-1} \phi(jT/n)\chi_{[jT/n,(j+1)T/n)}\zeta.$$

Because of the continuity of ϕ, one can show that z_n converges in operator norm to an operator z. Now z_n and hence z are in $L^\infty[0,T]' \cap B(\mathcal{H})$. Therefore, the multiplication operator $\mathfrak{m}(z)$ is defined as in the previous subsection. Then we set

$$\mathfrak{n}_{\text{mono}}(\beta) := \beta[1]P_\xi + \mathfrak{m}_{\text{mono}}(z).$$

(4) The definition of $\mathfrak{n}_{\mathrm{mono}\,\dagger}$ is the same as the monotone case except that we use $\phi(t) = \beta[\chi_{[0,t]}]$ instead of $\beta[\chi_{[t,T]}]$.

One rationale for the multiplication operators of the second kind is the following relation between these multiplication operators and the creation and annihilation operators.

Lemma 8.3.1. *Let $\mathcal{H}$ be a $\mathcal{B}$-$\mathcal{B}$-correspondence, let* $\mathrm{ind} \in \{\mathrm{bool}, \mathrm{free}, \mathrm{mono}, \mathrm{mono}\,\dagger\}$*, and let $\mathcal{F}_{\mathrm{ind}}(\mathcal{H})$ be the Fock space constructed above. Let $\zeta_1, \zeta_2 \in \mathcal{H}$. The map $L^\infty[0,T] \to \mathcal{B}$ given by $\beta[f] = \langle \zeta_1, f\zeta_2 \rangle$ is normal, and we have*

$$\ell_{\mathrm{ind}}(\zeta_1)^* \ell_{\mathrm{ind}}(\zeta_2) = \mathfrak{n}_{\mathrm{ind}}(\beta).$$

Proof of the boolean case. Note that if $b \in \mathcal{B} \subseteq \mathcal{F}_{\mathrm{bool}}(\mathcal{H})$, we have

$$\ell_{\mathrm{bool}}(\zeta_1)^* \ell_{\mathrm{bool}}(\zeta_2) b = \ell_{\mathrm{bool}}(\zeta_1)^*[\zeta_2 b] = \langle \zeta_1, \zeta_2 \rangle b = \mathfrak{n}_{\mathrm{bool}}(\beta) b.$$

Meanwhile, if $\zeta \in \mathcal{H} \subseteq \mathcal{F}_{\mathrm{bool}}(\mathcal{H})$, we have

$$\ell_{\mathrm{bool}}(\zeta_1)^* \ell_{\mathrm{bool}}(\zeta_2) \zeta = 0 = \mathfrak{n}_{\mathrm{bool}}(\beta).$$ □

Proof of the free case. It is a direct computation that for $\eta \in \mathcal{F}_{\mathrm{free}}(\mathcal{H})$,

$$\ell_{\mathrm{free}}(\zeta_1)^* \ell_{\mathrm{free}}(\zeta_2) \eta = \langle \zeta_1, \zeta_2 \rangle \eta,$$

and indeed the case of this identity where $\zeta_1 = \zeta_2$ can be gleaned from our discussion of the creation and annihilation operators earlier. This proves the claim since $\mathfrak{n}_{\mathrm{free}}(\beta)$ is exactly multiplication by $\langle \zeta_1, \zeta_2 \rangle$. □

Proof of the (anti-)monotone case. First, consider the case of a vector b from the direct summand $\mathcal{B} \subseteq \mathcal{F}_{\mathrm{mono}}(\mathcal{H})$. Since $\chi_{E_{\mathrm{mono}}(1,T)} = 1$, we have in this case that

$$\ell_{\mathrm{mono}}(\zeta_1)^* \ell_{\mathrm{mono}}(\zeta_2) b = \ell_{\mathrm{mono}}(\zeta_1)^*[\zeta_2 b] = \langle \zeta_1, \zeta_2 \rangle b = \mathfrak{n}_{\mathrm{mono}}(\beta) b.$$

Next, consider the action of $\ell_{\mathrm{mono}}(\zeta_1)^* \ell_{\mathrm{mono}}(\zeta_2)$ on vectors in the k direct summand $\chi_{E_{\mathrm{mono}}(k,T)} \mathcal{H}^{\otimes_{\mathcal{B}} k}$ for $k \geq 1$. It is clear that $\ell_{\mathrm{mono}}(\zeta_2)$ maps the kth summand into the $(k+1)$th and $\ell_{\mathrm{mono}}(\zeta_1)^*$ maps the $(k+1)$th into the kth. Therefore, to check that $\ell_{\mathrm{mono}}(\zeta_1)^* \ell_{\mathrm{mono}}(\zeta_2)$ agrees with $\mathfrak{n}_{\mathrm{mono}}(\beta)$ on $\chi_{E_{\mathrm{mono}}(k,T)} \mathcal{H}^{\otimes_{\mathcal{B}} k}$, it suffices to show that for $\eta_1, \eta_2 \in \chi_{E_{\mathrm{mono}}(k,T)} \mathcal{H}^{\otimes_{\mathcal{B}} k}$, we have

$$\langle \ell_{\mathrm{mono}}(\zeta_1) \eta_1, \ell_{\mathrm{mono}}(\zeta_2) \eta_2 \rangle = \langle \eta_1, \mathfrak{n}_{\mathrm{mono}}(\beta) \eta_2 \rangle.$$

Let us continue to view $\mathcal{F}_{\mathrm{mono}}(\mathcal{H})$ as a subspace of $\mathcal{F}_{\mathrm{free}}(\mathcal{H})$. Recalling the definitions of $\ell_{\mathrm{mono}}(\zeta_j)$ and $\mathfrak{n}_{\mathrm{mono}}(\beta)$, we can rewrite the equation we want to prove as

$$\langle \chi_{E_{\mathrm{mono}}(k+1,T)}[\zeta_1 \otimes \eta_1], \chi_{E_{\mathrm{mono}}(k+1,T)}[\zeta_2 \otimes \eta_2] \rangle = \langle \eta_1, \mathfrak{m}_{\mathrm{mono}}(z) \eta_2 \rangle, \tag{8.2}$$

where z is the operator constructed in the definition of $\mathfrak{n}_{\text{mono}}(\beta)$. Now let $F_n \subseteq [0,T]^{k+1}$ be the set

$$F_n = \bigcup_{n-1 \geq j(0) \geq \cdots \geq j(k) \geq 0} [j(0)T/n, (j(0)+1)T/n) \times \cdots \times [j(k)T/n, (j(k)+1)T/n),$$

which is an approximation of $E_{\text{mono}}(k+1,T)$ by a untion of rectangles. Note that $F_n \supseteq E_{\text{mono}}(k+1,T)$ (up to null sets) and $|F_n \setminus E_{\text{mono}}(k+1,T)| \to 0$. Since the action of $L^\infty([0,T]^{k+1})$ on $\mathcal{H}^{\otimes_{\mathcal{B}} k}$ is normal, we have

$$\lim_{n\to\infty} \langle \chi_{F_n}[\zeta_1 \otimes \eta_1], \chi_{F_n}[\zeta_2 \otimes \eta_2] \rangle = \langle \chi_{E_{\text{mono}}(k+1,T)}[\zeta_1 \otimes \eta_1], \chi_{E_{\text{mono}}(k+1,T)}[\zeta_2 \otimes \eta_2] \rangle.$$

Meanwhile, the right hand side of (8.2) can be expressed as

$$\langle \eta_1, \mathfrak{m}_{\text{mono}}(z)\eta_2 \rangle = \lim_{n\to\infty} \langle \eta_1, \mathfrak{m}_{\text{mono}}(z_n)\eta_2 \rangle,$$

where z_n is as in the definition of $\mathfrak{n}_{\text{mono}}(\beta)$, because we already know that $z_n \to z$ in operator norm, hence $\mathfrak{m}_{\text{mono}}(z_n) \to \mathfrak{m}_{\text{mono}}(z)$ in operator norm. Therefore, to prove (8.2), it suffices to prove the "approximate version"

$$\begin{aligned}\langle \chi_{F_n}[\zeta_1 \otimes \eta_1], \chi_{F_n}[\zeta_2 \otimes \eta_2] \rangle &= \langle \eta_1, \mathfrak{m}_{\text{mono}}(z_n)\eta_2 \rangle \\ &= \langle \eta_1, \mathfrak{m}_{\text{free}}(z_n)\chi_{E_{\text{mono}}(k,T)}\eta_2 \rangle \\ &= \langle \eta_1, \mathfrak{m}_{\text{free}}(z_n)\eta_2 \rangle, \end{aligned} \tag{8.3}$$

where the last inequality follows because we assumed that $\eta_2 \in \chi_{E_{\text{mono}}(k,T)}\mathcal{H}^{\otimes_{\mathcal{B}} k}$. Let us start with the right hand side. Letting ϕ be as in the definition of $\mathfrak{n}_{\text{mono}}(\beta)$. Recall that the definition of z_n was that for $\zeta \in \mathcal{H}$, we have

$$\begin{aligned} z_n\zeta &= \sum_{j=0}^{n-1} \phi(jT/n)\chi_{[jT/n,(j+1)T/n)}\zeta \\ &= \sum_{j=0}^{n-1} \beta[\chi_{[jT/n,T]}]\chi_{[jT/n,(j+1)T/n]} \\ &= \sum_{j=0}^{n-1} \langle \zeta_1, \chi_{[jT/n,T]}\zeta_2 \rangle \chi_{[jT/n,(j+1)T/n]}\zeta. \end{aligned}$$

Let $I_j = [jT/n, (j+1)T/n)$ for $j = 0, \dots, n-1$. Therefore, we have

$$\mathfrak{m}_{\text{free}}(z_n)\eta_2 = \sum_{j=0}^{n-1} \langle \zeta_1, \chi_{[jT/n,T]}\zeta_2 \rangle [\chi_{I_j} \otimes 1]\eta_2.$$

Since η_2 is in the image of $\chi_{E_{\text{mono}}(k,T)}$ and $F_n \supseteq E_{\text{mono}}(k,T)$, we have $\chi_{F_n}\eta_2 = \eta_2$. Thus,

$$\begin{aligned}
\mathfrak{m}_{\text{free}}(z_n)\eta_2 &= \sum_{j=0}^{n-1} \langle \zeta_1, \chi_{[jT/n,T]}\zeta_2\rangle [\chi_{I_j} \otimes 1]\chi_{F_n}\eta_2 \\
&= \sum_{j=0}^{n-1} \langle \zeta_1, \chi_{[jT/n,T]}\zeta_2\rangle [\chi_{I_j} \otimes 1] \sum_{n-1\geq j(1)\geq\cdots\geq j(k)\geq 0} [\chi_{I_{j(1)}} \otimes \cdots \otimes \chi_{I_{j(k)}}]\eta_2 \\
&= \sum_{n-1\geq j(1)\geq\cdots\geq j(k)\geq 0} \langle \zeta_1, \chi_{[j(1)T/n,T]}\zeta_2\rangle [\chi_{I_{j(1)}} \otimes \cdots \otimes \chi_{I_{j(k)}}]\eta_2,
\end{aligned}$$

where the last equality follows because only the terms where $j = j(1)$ will survive when we multiply $\chi_{[jT/n,(j+1)T/n]} \otimes 1$ by $\chi_{I_{j(1)}} \otimes \cdots \otimes \chi_{I_{j(k)}}$. Then we express $[j(1)T/n, T]$ as the union of $I_{j(0)}$ over $j(0) = j(1), \ldots, n-1$. Thus, we get

$$\mathfrak{m}_{\text{free}}(z_n)\eta_2 = \sum_{n-1\geq j(0)\geq j(1)\geq\cdots\geq j(k)\geq 0} \langle \zeta_1, \chi_{I_{j(0)}}\zeta_2\rangle [\chi_{I_{j(1)}} \otimes \cdots \otimes \chi_{I_{j(k)}}]\eta_2,$$

so that

$$\begin{aligned}
\langle \eta_1, \mathfrak{m}_{\text{free}}(z_n)\eta_2\rangle &= \sum_{n-1\geq j(0)\geq j(1)\geq\cdots\geq j(k)\geq 0} \left\langle \eta_1, \langle \zeta_1, \chi_{I_{j(0)}}\zeta_2\rangle [\chi_{I_{j(1)}} \otimes \cdots \otimes \chi_{I_{j(k)}}]\eta_2 \right\rangle \\
&= \sum_{n-1\geq j(0)\geq j(1)\geq\cdots\geq j(k)\geq 0} \left\langle \zeta_1 \otimes \eta_1, [\chi_{I_{j(0)}} \otimes \chi_{I_{j(1)}} \otimes \cdots \otimes \chi_{I_{j(k)}}][\zeta_1 \otimes \eta_2] \right\rangle \\
&= \langle \zeta_1 \otimes \eta_1, \chi_{F_n}[\zeta_2 \otimes \eta_2]\rangle \\
&= \langle \chi_{F_n}[\zeta_1 \otimes \eta_1], \chi_{F_n}[\zeta_2 \otimes \eta_2]\rangle,
\end{aligned}$$

which establishes (8.3) and hence completes the proof of the monotone case. The anti-monotone case of course is completely symmetrical. □

8.3.4 General time intervals

Of course, instead of using the time interval $[0, T]$, we could have used an arbitrary interval $[a, b]$. The sets $E_{\text{ind}}(k, T)$ would then be replaced with

$$\begin{aligned}
E_{\text{bool}}(k, a, b) &= \begin{cases} [a, b], & k = 1, \\ \varnothing, & k > 1 \end{cases} \\
E_{\text{free}}(k, a, b) &= [a, b]^k \\
E_{\text{mono}}(k, a, b) &= \{(t_1, \ldots, t_k) \in [a, b]^k : t_1 > t_2 > \cdots > t_k\} \\
E_{\text{mono}\dagger}(k, a, b) &= \{(t_1, \ldots, t_k) \in [a, b]^k : t_1 < t_2 < \cdots < t_k\}.
\end{aligned}$$

All the definitions and results about generalize easily from $[0, T]$ to $[a, b]$ with the appropriate changes in notation.

8.4 Independence on the Fock space

We will next explain how the operators on the Fock space naturally give rise to processes with independent increments. The end goal of the section is the following result.

Proposition 8.4.1. *Let $\mathcal{H}$ be a $\mathcal{B}$-$\mathcal{B}$-correspondence with a normal $L^\infty[0,T]$-action, and let* $\mathrm{ind} \in \{\mathrm{bool}, \mathrm{free}, \mathrm{mono}, \mathrm{mono}\,\dagger\}$*. For $0 \leq s \leq t \leq T$, let $\mathcal{A}_{s,t}$ be the* C**-subalgebra of $B(\mathcal{F}_{\mathrm{ind}}(\mathcal{H}))$ generated by*

$$\begin{aligned}&\{\ell_{\mathrm{ind}}(\zeta) : \zeta \in \mathcal{H}, \chi_{[s,t]}\zeta = \zeta\} \\ \cup&\{\mathfrak{m}_{\mathrm{ind}}(a) : a \in B(\mathcal{H}), a = \chi_{[s,t]} a \chi_{[s,t]}\} \\ \cup&\{\mathfrak{n}_{\mathrm{ind}}(\beta) :\ \beta : L^\infty[0,T] \to \mathcal{B} \text{ normal}, \beta = \chi_{[s,t]}\beta\}.\end{aligned}$$

Then for every $0 = t_0 < t_1 < \cdots < t_N = T$, the subalgebras $\mathcal{A}_{t_0,t_1}$, $\dots$, $\mathcal{A}_{t_{N-t},t_N}$ are ind*-independent in $(\mathcal{F}_{\mathrm{ind}}(\mathcal{H}), E_\xi)$.*

While this proposition can be verified by hand in several ways, we believe the most natural and straightforward method is to show that the Fock space itself is isomorphic to the independent product of the individual $\mathcal{B}$-$\mathcal{B}$-correspondences associated to subintervals of $[0,T]$, and to verify that this isomorphism respects the creation, annihilation, and multiplication operators in a natural sense. That is the content of the theorem below. Here, if $\mathcal{H}$ and $\mathcal{K}$ are correspondences and $T : \mathcal{H} \to \mathcal{K}$ is adjointable, then we denote $\mathrm{Ad}(T) : B(\mathcal{H}) \to B(\mathcal{K})$ by $\mathrm{Ad}(T)(S) = TST^*$.

Theorem 8.4.2. *With the notation of Proposition 8.4.1, let $0 = t_0 < t_1 < \cdots < t_N = T$. Then there is a $\mathcal{B}$-$\mathcal{B}$-correspondence isomorphism*

$$\Phi_{\mathrm{ind},t_0,\dots,t_N} : \bigstar_{\mathrm{ind}}[(\mathcal{F}_{\mathrm{ind}}(\chi_{[t_0,t_1]}\mathcal{H}), \xi_1), \dots (\mathcal{F}_{\mathrm{ind}}(\chi_{[t_{N-1},t_N]}\mathcal{H}), \chi_{[t_0,t_1]}\xi_N)] \to (\mathcal{F}_{\mathrm{ind}}(\mathcal{H}), \xi),$$

where $\chi_{[t_{j-1},t_j]}\mathcal{H}$ is viewed as $\mathcal{B}$-$\mathcal{B}$-correspondence with an $L^\infty[t_{j-1},t_j]$-action, and ξ_j is the vacuum vector in $\mathcal{F}_{\mathrm{ind}}(\chi_{[t_{j-1},t_j]}\mathcal{H})$. Furthermore, let $i_{[s,t]} : \chi_{[s,t]}\mathcal{H} \to \mathcal{H}$ be the canonical inclusion map, and let

$$\rho_{\mathrm{ind},j} : B(\mathcal{F}_{\mathrm{ind}}(\chi_{[t_{j-1},t_j]}\mathcal{H})) \to B\left(\bigstar_{\mathrm{ind}}[(\mathcal{F}_{\mathrm{ind}}(\chi_{[t_0,t_1]}\mathcal{H}), \xi_1), \dots (\mathcal{F}_{\mathrm{ind}}(\chi_{[t_{N-1},t_N]}\mathcal{H}), \chi_{[t_0,t_1]}\xi_N)]\right)$$

be the $*$-homomorphism given by the definition of independence. Then the following diagrams commute:

$$\begin{array}{ccc} \chi_{[t_{j-1},t_j]}\mathcal{H} & \xrightarrow{\quad i_{[t_{j-1},t_j]} \quad} & \mathcal{H} \\ \downarrow{\scriptstyle \ell_{\mathrm{ind}}} & & \downarrow{\scriptstyle \ell_{\mathrm{ind}}} \\ B(\mathcal{F}_{\mathrm{ind}}(\chi_{[t_{j-1},t_j]}\mathcal{H})) & \xrightarrow{\mathrm{Ad}(\Phi_{\mathrm{ind},t_0,\dots,t_N}) \circ \rho_{\mathrm{ind},j}} & B(\mathcal{F}_{\mathrm{ind}}(\mathcal{H})), \end{array}$$

and

$$\begin{array}{ccc} B(\chi_{[t_{j-1},t_j]}\mathcal{H}) & \xrightarrow{\operatorname{Ad}(i_{[t_{j-1},t_j]})} & B(\mathcal{H}) \\ \downarrow \mathfrak{m}_{\mathrm{ind}} & & \downarrow \mathfrak{m}_{\mathrm{ind}} \\ B(\mathcal{F}_{\mathrm{ind}}(\chi_{[t_{j-1},t_j]}\mathcal{H})) & \xrightarrow{\operatorname{Ad}(\Phi_{\mathrm{ind},t_0,\dots,t_N})\circ\rho_{\mathrm{ind},j}} & B(\mathcal{F}_{\mathrm{ind}}(\mathcal{H})), \end{array}$$

and

$$\begin{array}{ccc} \mathcal{L}_{\mathrm{n}}(L^\infty[t_{j-1},t_j],\mathcal{B}) & \xrightarrow{(i_{[t_{j-1},t_j]})_*} & \mathcal{L}_{\mathrm{n}}(L^\infty[0,T],\mathcal{B}) \\ \downarrow \mathfrak{n}_{\mathrm{ind}} & & \downarrow \mathfrak{n}_{\mathrm{ind}} \\ B(\mathcal{F}_{\mathrm{ind}}(\chi_{[t_{j-1},t_j]}\mathcal{H})) & \xrightarrow{\operatorname{Ad}(\Phi_{\mathrm{ind},t_0,\dots,t_N})\circ\rho_{\mathrm{ind},j}} & B(\mathcal{F}_{\mathrm{ind}}(\mathcal{H})), \end{array}$$

where $\mathcal{L}_{\mathrm{n}}(L^\infty[a,b],\mathcal{B})$ denotes the space of bounded normal maps from $L^\infty[a,b]$ to $\mathcal{B}$, and $(i_{[t_{j-1},t_j]})_*\beta$ is the map $f \mapsto \beta[f|_{[t_{j-1},t_j]}]$.

This theorem immediately implies Proposition 8.4.1. Indeed, the theorem shows that if $\zeta \in \mathcal{H}$ with $\chi_{[t_{j-1},t_j]}\zeta = \zeta$, then the creation operator $\ell_{\mathrm{ind}}(\zeta)$ on $\mathcal{F}_{\mathrm{ind}}(\mathcal{H})$ is equivalent via the isomorphism $\Phi_{\mathrm{ind},t_0,\dots,t_N}$ to the image under $\rho_{\mathrm{ind},j}$ of the corresponding creation operator on $\mathcal{F}(\chi_{[t_{j-1},t_j]}\mathcal{H})$. Thus, $\operatorname{Ad}(\Phi_{t_0,\dots,t_N})^{-1}\ell_{\mathrm{ind}}(\zeta)$ is in the image of $\rho_{\mathrm{ind},j}$ on the ind-product $\mathcal{B}$-$\mathcal{B}$-correspondence. By the same token, the multiplication operators of the first and second kind associated to objects "supported in $[t_{j-1},t_j]$" correspond under $\operatorname{Ad}(\Phi_{\mathrm{ind},t_0,\dots,t_N})^{-1}$ to operators in the image of $\rho_{\mathrm{ind},j}$. Thus, we have

$$\operatorname{Ad}(\Phi_{\mathrm{ind},t_0,\dots,t_N})^{-1}[\mathcal{A}_{t_{j-1},t_j}] \subseteq \rho_{\mathrm{ind},j}\Big(B(\mathcal{F}_{\mathrm{ind}}(\chi_{t_{j-1},t_j}\mathcal{H}))\Big).$$

But we know from Theorem 5.3.1 that the images of $\rho_{\mathrm{ind},j}$ are independent, hence also the algebras $\mathcal{A}_{t_{j-1},t_j}$ are also independent. It remains to prove Theorem 8.4.2, which we will do in cases.

8.4.1 The boolean case

In the boolean case, the isomorphism $\Phi_{\mathrm{bool},t_0,\dots,t_N}$ is obtained as follows:

$$\mathcal{F}_{\mathrm{bool}}(\mathcal{H}) = \mathcal{B}\xi \oplus \mathcal{H} \cong \mathcal{B}\xi \oplus \bigoplus_{j=1}^N \chi_{[t_{j-1},t_j]}\mathcal{H} = \mathcal{B}\xi \oplus \bigoplus_{j=1}^N \mathcal{F}_{\mathrm{bool}}(\chi_{[t_{j-1},t_j]}\mathcal{H})^\circ,$$

since $\mathcal{F}_{\mathrm{bool}}(\chi_{[t_{j-1},t_j]}\mathcal{H})^\circ = \chi_{[t_{j-1},t_j]}\mathcal{H}$, and the right-hand side is exactly the boolean product $\mathcal{B}$-$\mathcal{B}$-correspondence.

Suppose $\zeta \in \chi_{[t_{j-1},t_j]}\mathcal{H}$. The creation operator $\ell_{\mathrm{bool}}(\zeta)$ on $\mathcal{F}_{\mathrm{bool}}(\chi_{[t_{j-1},t_j]}\mathcal{H})$ maps $b\xi_j$ to $\zeta b \in \chi_{[t_{j-1},t_j]}\mathcal{H}$ and annihilates $\chi_{[t_{j-1},t_j]}\mathcal{H}$. Now $\rho_{\mathrm{bool},j}(\ell_{\mathrm{bool}}(\zeta))$ annihilates the spaces $\mathcal{F}_{\mathrm{bool}}(\chi_{[t_{k-1},t_k]}\mathcal{H})$ for $k \neq j$, and it applies $\ell_{\mathrm{bool}}(\zeta)$ on $\mathcal{B}\xi \oplus \mathcal{F}_{\mathrm{bool}}(\chi_{[t_{j-1},t_j]}\mathcal{H})^\circ \cong \mathcal{F}_{\mathrm{bool}}(\chi_{[t_{j-1},t_j]}\mathcal{H})$.

Thus, $\rho_{\mathrm{bool},j}(\ell(\zeta))$ maps $b\xi$ to ζb in $\chi_{[t_{j-1},t_j]}\mathcal{H} \subseteq \mathcal{H}$. But under the isomorphim $\Phi_{\mathrm{bool},t_0,\dots,t_N}$, this corresponds exactly to the action of $\ell_{\mathrm{bool}}(i_{[t_{j-1},t_j]}\zeta)$ on $\mathcal{F}_{\mathrm{bool}}(\mathcal{H})$.

Supppose $a \in B(\chi_{[t_{j-1},t_j]}\mathcal{H})$. The multiplication operator $\mathfrak{m}_{\mathrm{bool}}(a)$ on $\mathcal{F}_{\mathrm{bool}}(\chi_{[t_{j-1},t_j]}\mathcal{H})$ applies a on $\chi_{[t_{j-1},t_j]}\mathcal{H}$ and zero on $\mathcal{B}\xi_j$. Thus, $\rho_{[t_{j-1},t_j]}$ applies a on the subspace $\chi_{[t_{j-1},t_j]}\mathcal{H}$ and zero on $\mathcal{B}\xi$ and $\chi_{[t_{k-1},t_k]}\mathcal{H}$ for $k \neq j$. This is the same as the operator $\mathfrak{m}_{\mathrm{bool}}(\mathrm{Ad}(i_{[t_{j-1},t_j]})(a)$.

Suppose $\beta \in \mathcal{L}_{\mathrm{n}}(L^\infty[0,T],\mathcal{B})$ with $\chi_{[t_{j-1},t_j]}\beta = \beta$. Then $\mathfrak{n}_{\mathrm{bool}}(\beta) = \beta[1]P_\xi$ on $\mathcal{F}_{\mathrm{bool}}(\chi_{[t_{j-1},t_j]}\mathcal{H})$. It is straightforward to show that $\rho_{\mathrm{bool},j}(P_{\xi_j}) = P_\xi$ and $\rho_{\mathrm{bool},j}$ is a $\mathcal{B}$-$\mathcal{B}$-bimodule map. Thus, $\rho_{\mathrm{bool},j}(\mathfrak{m}_{\mathrm{bool}}(a))$ is $\beta[1]P_\xi$, which is the equivalent by the isomorphism $\Phi_{\mathrm{bool},t_0,\dots,t_N}$ to $(i_{[t_{j-1},t_j]})_*\beta[1]P_\xi = \mathfrak{n}_{\mathrm{bool}}((i_{[t_{j-1},t_j]})_*\beta)$.

8.4.2 The free case

In the free case, the isomorphism $\Phi_{\mathrm{free},t_0,\dots,t_N}$ is obtained as follows. Note that

$$\begin{aligned}\mathcal{F}_{\mathrm{free}}(\mathcal{H}) &\cong \bigoplus_{k\geq 0}\left(\bigoplus_{j=1}^N \chi_{[t_{j-1},t_j]}\mathcal{H}\right)^{\otimes_{\mathcal{B}} k} \\ &\cong \bigoplus_{k\geq 0}\bigoplus_{j_1,\dots,j_k} (\chi_{[t_{j_1-1},t_{j_1}]}\mathcal{H}) \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} (\chi_{[t_{j_k-1},t_{j_k}]}\mathcal{H}).\end{aligned}$$

The string (or sequence) of indices $j_1 \dots j_k$ may have some consecutive repeated indices. But this string can always be uniquely expressed as k_1 occurrences of an index i_1, then k_2 occurrences of an index i_2, so forth up to k_m occurrences of i_m, where each $k_r \geq 1$ and where $i_1 \neq i_2 \neq \dots \neq i_m$ (or in other words $i_1 \dots i_m$ is alternating). Thus, we have

$$\begin{aligned}\mathcal{F}_{\mathrm{free}}(\mathcal{H}) &\cong \bigoplus_{m\geq 0}\bigoplus_{i_1\neq\dots\neq i_m}\bigoplus_{k_1,\dots,k_m\geq 1} (\chi_{[t_{i_1-1},t_{i_1}]}\mathcal{H})^{\otimes_{\mathcal{B}} k_1} \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} (\chi_{[t_{i_m-1},t_{i_m}]}\mathcal{H})^{\otimes_{\mathcal{B}} k_m} \\ &\cong \bigoplus_{m\geq 0}\bigoplus_{i_1\neq\dots\neq i_m}\left(\bigoplus_{k_1\geq 1}(\chi_{[t_{i_1-1},t_{i_1}]}\mathcal{H})^{\otimes_{\mathcal{B}} k_1}\right) \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \left(\bigoplus_{k_m\geq 1}(\chi_{[t_{i_m-1},t_{i_m}]}\mathcal{H})^{\otimes_{\mathcal{B}} k_m}\right) \\ &\cong \bigoplus_{m\geq 0}\bigoplus_{i_1\neq\dots\neq i_m} \mathcal{F}_{\mathrm{free}}(\chi_{[t_{i_1-1},t_{i_1}]})^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{F}_{\mathrm{free}}(\chi_{[t_{i_N-1},t_{i_N}]})^\circ,\end{aligned}$$

which is exactly the free-product $\mathcal{B}$-$\mathcal{B}$-correspondence obtained from $\mathcal{F}_{\mathrm{free}}(\chi_{[t_0,t_1]}\mathcal{H})$, …, $\mathcal{F}_{\mathrm{free}}(\chi_{[t_{N-1},t_N]}\mathcal{H})$.

Suppose $\zeta \in \chi_{[t_{j-1},t_j]}\mathcal{H}$. The creation operator $\ell_{\mathrm{free}}(\zeta)$ on $\mathcal{F}_{\mathrm{bool}}(\chi_{[t_{j-1},t_j]}\mathcal{H})$ will tensor a given input vector in one of the direct summands $\chi_{[t_{j-1},t_j]}\mathcal{H}^{\otimes_{\mathcal{B}} k}$ with ζ on the left. To determine the action of $\rho_{\mathrm{free},j}(\ell_{\mathrm{free}}(\zeta))$ on the free product $\mathcal{B}$-$\mathcal{B}$-correspondence, consider a vector in one of the direct summands

$$(\chi_{[t_{i_1-1},t_{i_1}]}\mathcal{H})^{\otimes_{\mathcal{B}} k_1} \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} (\chi_{[t_{i_m-1},t_{i_m}]}\mathcal{H})^{\otimes_{\mathcal{B}} k_m}.$$

If $i_1 = j$, then this is contained in the direct summand

$$\mathcal{F}_{\text{free}}(\chi_{[t_{j-1},t_j]}\mathcal{H})^\circ \otimes_{\mathcal{B}} \mathcal{F}_{\text{free}}(\chi_{[t_{i_2-1},t_{i_2}]}\mathcal{H})^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{F}_{\text{free}}(\chi_{[t_{i_m-1},t_{i_m}]}\mathcal{H}) \\ \subseteq \mathcal{F}_{\text{free}}(\chi_{[t_{j-1},t_j]}\mathcal{H}) \otimes_{\mathcal{B}} \mathcal{F}_{\text{free}}(\chi_{[t_{i_2-1},t_{i_2}]}\mathcal{H})^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{F}_{\text{free}}(\chi_{[t_{i_m-1},t_{i_m}]}\mathcal{H}).$$

If $i_1 \neq j$, it is contained in the direct summand

$$\mathcal{F}_{\text{free}}(\chi_{[t_{i_1-1},t_{i_1}]}\mathcal{H})^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{F}_{\text{free}}(\chi_{[t_{i_m-1},t_{i_m}]}\mathcal{H}) \\ \subseteq \mathcal{F}_{\text{free}}(\chi_{[t_{j-1},t_j]}\mathcal{H}) \otimes_{\mathcal{B}} \mathcal{F}_{\text{free}}(\chi_{[t_{i_1-1},t_{i_1}]}\mathcal{H})^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{F}_{\text{free}}(\chi_{[t_{i_m-1},t_{i_m}]}\mathcal{H}).$$

But in either case, $\rho_{[t_{j-1},t_j]}(\ell_{\text{free}}(\zeta))$ will still tensor the input vector with ζ on the left, mapping it into

$$\chi_{[t_{j-1},t_j]}\mathcal{H} \otimes_{\mathcal{B}} (\chi_{[t_{i_1-1},t_{i_1}]}\mathcal{H})^{\otimes_{\mathcal{B}} k_1} \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} (\chi_{[t_{i_m-1},t_{i_m}]}\mathcal{H})^{\otimes_{\mathcal{B}} k_m}.$$

The creation operator $\ell_{\text{free}}((i_{[t_{j-1},t_j]})_*\zeta)$ on $\mathcal{F}_{\text{free}}(\mathcal{H})$ works out to exactly the same thing.

For $a \in B(\chi_{[t_{j-1},t_j]}\mathcal{H})$, the multiplication operator $\rho_{\text{free},j}(\mathfrak{m}_{\text{free}}(a))$ will multiply a vector in

$$(\chi_{[t_{i_1-1},t_{i_1}]}\mathcal{H})^{\otimes_{\mathcal{B}} k_1} \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} (\chi_{[t_{i_m-1},t_{i_m}]}\mathcal{H})^{\otimes_{\mathcal{B}} k_m}$$

by a on the left if $i_1 = j$ and by zero if $i_1 \neq j$, and this is the same thing as $\mathfrak{m}_{\text{free}}(\text{Ad}(i_{[t_{j-1},t_j]})(\zeta))$. Finally, for $\beta \in \mathcal{L}_{\text{n}}(L^\infty[t_{j-1},t_j],\mathcal{B})$, the operators $\rho_{\text{free},j}(\mathfrak{n}_{\text{free}}(\beta))$ and $\mathfrak{n}_{\text{free}}((i_{[t_{j-1},t_j]})_*\beta)$ both work out to multiplication by $\beta[1] \in \mathcal{B}$ on the entire space $\mathcal{F}_{\text{free}}(\mathcal{H})$.

8.4.3 The (anti-)monotone case

In the monotone case, the isomorphism $\Phi_{\text{free},t_0,\dots,t_N}$ is obtained as follows. For the sake of space, let us abbreviate $E_{\text{mono}}(k,T)$ as $E(k,T)$ and $E_{\text{mono}}(k,a,b)$ as $E(k,a,b)$ throughout this subsection. Recall that

$$\mathcal{F}_{\text{free}}(\mathcal{H}) = \bigoplus_{k \geq 0} \chi_{E(k,T)}[\mathcal{H}^{\otimes_{\mathcal{B}} k}].$$

The points $\{t_0,\dots,t_N\}^k$ partition $[0,T]^k$ into the rectangles $[t_{j_1-1},t_{j_1}] \times \cdots \times [t_{j_k-1},t_{j_k}]$ for $j_1, \dots, j_k \in \{1,\dots,N\}$ (we ignore boundary issues since the boundaries of the rectangles have measure zero). Intersecting with $E(k,T)$ yields a partition of $E(k,T)$ into smaller sets. But $[t_{j_1-1},t_{j_1}] \times \cdots \times [t_{j_k-1},t_{j_k}]$ only intersects $E(k,T)$ nontrivially if $j_1 \geq j_2 \geq \cdots \geq j_k$. In this case, we can express $j_1 \dots j_k$ as k_1 occurrences of i_1, k_2 occurrences of i_2, and so forth up to k_m occurrences of i_m, where $i_1 > i_2 > \cdots > i_m$. Then

$$E(k,T) \cap \left([t_{i_1-1},t_{i_1}]^{k_1} \times \cdots \times [t_{i_m-1},t_{i_m}]^{k_m}\right) = E(k_1,t_{i_1-1},t_{i_1}) \times \cdots \times E(k_m,t_{i_m-1},t_{i_m}).$$

Thus, we have

$$\mathcal{F}_{\text{mono}}(\mathcal{H}) \cong \bigoplus_{m\geq 0} \bigoplus_{i_1>\cdots>i_m} \bigoplus_{k_1,\dots,k_m\geq 1} [\chi_{E(k_1,t_{i_1-1},t_{i_1})}\mathcal{H}^{\otimes_{\mathcal{B}} k_1}] \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} [\chi_{E(k_m,t_{i_m-1},t_{i_m})}\mathcal{H}^{\otimes_{\mathcal{B}} k_m}]$$
$$\cong \bigoplus_{m\geq 0} \bigoplus_{i_1\neq\dots\neq i_m} \mathcal{F}_{\text{mono}}(\chi_{[t_{i_1-1},t_{i_1}]})^\circ \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{F}_{\text{mono}}(\chi_{[t_{i_N-1},t_{i_N}]})^\circ,$$

where we have used the identification

$$\chi_{E(k_j,t_{i_j-1},t_{i_j})}\mathcal{H}^{\otimes_{\mathcal{B}} k_j} \cong \chi_{E(k_j,t_{i_j-1},t_{i_j})}(\chi_{[t_{i_j-1},t_{i_j}]}\mathcal{H})^{\otimes_{\mathcal{B}} k_1}.$$

But the right-hand side above is exactly the monotone product $\mathcal{B}$-$\mathcal{B}$-correspondence obtained from $\mathcal{F}_{\text{mono}}(\chi_{[t_0,t_1]}\mathcal{H})$, ..., $\mathcal{F}_{\text{mono}}(\chi_{[t_{N-1},t_N]}\mathcal{H})$.

Suppose $\zeta \in \chi_{[t_{j-1},t_j]}\mathcal{H}$. Consider applying the operator $\rho_{\text{mono},j}(\ell_{\text{mono}}(\zeta))$ to a vector in

$$[\chi_{E(k_1,t_{i_1-1},t_{i_1})}\mathcal{H}^{\otimes_{\mathcal{B}} k_1}] \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} [\chi_{E(k_m,t_{i_m-1},t_{i_m})}\mathcal{H}^{\otimes_{\mathcal{B}} k_m}].$$

If $m = 0$ or if $j > i_1$, then it will tensor ζ on the left. If $j = i_1$, then it will tensor ζ on the left to obtain a vector in $\chi_{[t_{j-1},t_j]}\mathcal{H} \otimes_{\mathcal{B}} [\chi_{E(k_1,t_{j-1},t_j)}\mathcal{H}^{\otimes_{\mathcal{B}} k_1}]$ tensored with the rest of the spaces above, and next it will multiply by $\chi_{E(k_1+1,t_{j-1},t_j)}$ acting on the left-most tensorands $\chi_{[t_{j-1},t_j]}\mathcal{H} \otimes_{\mathcal{B}} [\chi_{E(k_1,t_{j-1},t_j)}\mathcal{H}^{\otimes_{\mathcal{B}} k_1}]$. Finally, if $j < i_1$, then $\rho_{\text{mono},j}(\ell_{\text{mono}}(\zeta))$ will evaluate to zero on the input vector.

Then one can check that $\ell_{\text{mono}}(i_{t_{j-1},t_j}\zeta)$ will do the exact same thing. Indeed, letting $k = k_1 + \cdots + k_m$, this operator will tensor the input vector with ζ and then multiply by $\chi_{E(k+1,T)}$. Now the input vector was "supported in" $E(k_1, t_{i_1-1}, t_{i_1}) \times \cdots \times E(k_m, t_{i_m-1}, t_{i_m})$, and after we tensor with ζ, the resulting vector is "supported in" $[t_{j-1}, t_j] \times E(k_1, t_{i_1-1}, t_{i_1}) \times \cdots \times E(k_m, t_{i_m-1}, t_{i_m})$. Hence, if $j > i_1$, then multiplying by $\chi_{E(k+1,T)}$ does nothing; if $j = i_1$, then it just amounts to multiplying by $\chi_{E(k_1+1,t_{j-1},t_j)}$ in the first $k_1 + 1$ tensorands; and if $j < i_1$, then it will kill the vector.

The argument for the multiplication operators $\mathfrak{m}_{\text{mono}}(a)$ for $a \in B(\chi_{[t_{j-1},t_j]}\mathcal{H})$ is no harder than in the free case, and we leave the details to the reader. On the other hand, the consideration of $\mathfrak{n}_{\text{mono}}(\beta)$ is more involved for the monotone case.

Let $\beta \in \mathcal{L}_{\text{n}}(L^\infty[t_{j-1}, t_j])$, let $\phi(t) = \beta[\chi_{[t,t_j]}]$ for $t \in [t_{j-1}, t_j]$, and let z be as in the definition of $\mathfrak{n}_{\text{mono}}(\beta)$ (intuitively, z is the operator of "left multiplication by ϕ" on $\chi_{[t_{j-1},t_j]}\mathcal{H}$). Recall that $\mathfrak{n}_{\text{mono}}(\beta) = \beta[1]P_\xi + \mathfrak{m}_{\text{mono}}(z)$. Meanwhile, denote

$$\psi(t) = (i_{t_{j-1},t_j})_*\beta[\chi_{[t,T]}] = \begin{cases} \beta[1], & t \in [0, t_{j-1}], \\ \phi(t), & t \in [t_{j-1}, t_j], \\ 0, & t \in [t_j, T], \end{cases}$$

and let w be the operator of "left multiplication by ψ" on $\mathcal{H}$. Consider the action of the two operators $\rho_{\text{mono},j}(\mathfrak{n}_{\text{mono}}(\beta))$ and $\mathfrak{n}_{\text{mono}}((i_{t_{j-1},t_j})_*\beta)$ on an input vector v from

$$[\chi_{E(k_1,t_{i_1-1},t_{i_1})}\mathcal{H}^{\otimes_{\mathcal{B}} k_1}] \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} [\chi_{E(k_m,t_{i_m-1},t_{i_m})}\mathcal{H}^{\otimes_{\mathcal{B}} k_m}],$$

viewed both as a vector in the monotone product space and in $\mathcal{F}_{\text{mono}}(\mathcal{H})$ by a slight abuse of notation.

- Suppose $m = 1$, that is, $v \in \mathcal{B}\xi$. Then $\rho_{\text{mono},j}(\mathfrak{n}_{\text{mono}}(\beta))$ and $\mathfrak{n}_{\text{mono}}((i_{t_{j-1},t_j})_*\beta)$ will both multiply v by $\beta[1]$.

- Suppose $j > i_1$. Then $\rho_{\text{mono},j}(\mathfrak{m}_{\text{mono}}(z))v = 0$ and $\rho_{\text{mono},j}(P_\xi)v = v$. Thus,

$$\rho_{\text{mono},j}(\mathfrak{n}_{\text{mono}}(\beta))v = \beta[1]v.$$

  Meanwhile, $\mathfrak{n}_{\text{mono}}((i_{t_{j-1},t_j})_*\beta)v = \beta[1]v$ because $\psi|_{[t_{i_1-1},t_{i_1}]} = \beta[1]$.

- Suppose $j = i_1$. Then

$$\rho_{\text{mono},j}(\mathfrak{n}_{\text{mono}}(\beta))v = \rho_{\text{mono},j}(\mathfrak{m}_{\text{mono}}(z))v = \mathfrak{m}_{\text{mono}}(w)v = \mathfrak{n}_{\text{mono}}((i_{t_{j-1},t_j})_*\beta)v,$$

  since v represents multiplication by ϕ and w represents multiplication by ψ, which restricts to ϕ on the interval $[t_{j-1}, t_j]$.

- Suppose $j < i_1$. Then $\rho_{\text{mono},j}(\mathfrak{n}_{\text{mono}}(\beta))v = 0$ by the nature of $\rho_{\text{mono},j}$. Meanwhile, note that $\phi|_{[0,t_j]} = 0$ and hence $\mathfrak{n}_{\text{mono}}((i_{t_{j-1},t_j})_*\beta)v = \mathfrak{m}_{\text{mono}}(w)v = 0$.

8.5 Construction of processes

Now we describe the construction of the process with independent increments.

Let ind $\in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\,\dagger\}$. Let $b \in \mathcal{L}(L^1[0,T], \mathcal{B})$ and let $\sigma : L^1[0,T] \times \mathcal{B}\langle Y \rangle \to \mathcal{B}$ be a distributional family of generalized laws. Let $\mathcal{H}$ be a $\mathcal{B}$-$\mathcal{B}$-correspondence with a normal $L^\infty[0,T]$-action, and let $\zeta \in \mathcal{H}$ and $Y \in B(\mathcal{H}) \cap L^\infty[0,T]'$ such that

$$\int_E \sigma(f(Y), t)\, dt = \langle \zeta, f(Y)\chi_E \zeta \rangle;$$

we know that such an $\mathcal{H}$, ζ, and Y exist by Lemma 8.2.6.

Note that $b : L^1[0,T] \to \mathcal{B}$ restricts to a linear map $L^\infty[0,T] \to \mathcal{B}$ which is normal. Moreover, for $E \subseteq [0,T]$ Borel, we may define $\chi_E b : L^1[0,T] \to \mathcal{B}$ by $\int (\chi_B b)(t) f(t)\, dt = \int b(t)[\chi_E f](t)\, dt$. Clearly also $\chi_E b$ restricts to a normal linear map $L^\infty[0,T] \to \mathcal{B}$. Hence, $\mathfrak{n}_{\text{ind}}(\chi_E b)$ is well defined.

For $0 \leq s \leq t \leq T$, we define self-adjoint operators $X_{s,t}$ on $\mathcal{F}_{\text{ind}}(\mathcal{H})$ by

$$X_{s,t} = \mathfrak{n}_{\text{ind}}(\chi_{[s,t]} b) + \ell_{\text{ind}}(\chi_{[s,t]}\zeta) + \ell_{\text{ind}}(\chi_{[s,t]}\zeta)^* + \mathfrak{m}_{\text{ind}}(\chi_{[s,t]} Y). \tag{8.4}$$

Note that for $s \leq t \leq u$, we have $X_{s,t} + X_{t,u} = X_{s,u}$ by linearity of $\mathfrak{n}_{\text{ind}}$, ℓ_{ind}, and $\mathfrak{m}_{\text{ind}}$.

Theorem 8.5.1. *Let* $\text{ind} \in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\,\dagger\}$*, let* $b \in \mathcal{L}(L^1[0,T], \mathcal{B})$*, and let* σ *be a distributional family of generalized laws on* $[0,T]$*. Let* $\mathcal{H}$*,* ζ*, and* Y *be as above, let* $X_{s,t}$ *be given by* (8.4)*, and set* $X_t = X_{0,t}$*. Then* $(X_t)_{t \in [0,T]}$ *is a process with* ind*-independent increments satisfying* $X_0 = 0$ *and*

$$\partial_t F_{X_t}(z) = \begin{cases} -[b(t) + G_{\sigma(\cdot,t)}(z)], & \textit{boolean case,} \\ -DF_{X_t}(z)[b(t) + G_{\sigma(\cdot,t)}(F_{X_t}(z))], & \textit{free case,} \\ -DF_{X_t}(z)[b(t) + G_{\sigma(\cdot,t)}(z)], & \textit{monotone case,} \\ -[b(t) + G_{\sigma(\cdot,t)}(F_{X_t}(z))], & \textit{anti-monotone case.} \end{cases}$$

Proof. Note that $X_t - X_s = X_{s,t}$. It follows from Proposition 8.4.1 that for $0 = t_0 < t_1 < \dots < t_N$, the operators X_{t_0,t_1}, …, X_{t_{N-1},t_N} are ind-independent. Thus, if μ_t is the law of X_t, then $(\mu_t)_{t \in [0,T]}$ is an ind-subordination family. More precisely, $\mu_s \boxplus_{\text{ind}} \mu_{s,t} = \mu_t$, where $\mu_{s,t}$ is the law of $X_{s,t}$.

Furthermore, we claim that it is a Lipschitz subordination family. The mean

$$\mu_t(X) = \langle \xi, X_t \xi \rangle = \langle \xi, \mathfrak{n}_{\text{ind}}(\chi_{[0,t]}\beta)\xi \rangle = \int_0^t b(s)\,ds.$$

Since $b \in \mathcal{L}(L^1[0,T], \mathcal{B})$, we know that $\int_0^t b(s)\,ds$ is Lipschitz in t. Moreover, a direct computation shows that

$$\operatorname{Var}(\mu_t)[1] = \langle \xi, \ell(\chi_{[0,t]}\zeta)^* \ell(\chi_{[0,t]}\zeta)\xi \rangle = \langle \zeta, \chi_{[0,t]}\zeta \rangle = \int_0^t \sigma(1,s)\,ds.$$

Since $\sigma(1,\cdot)$ is in $\mathcal{L}(L^1[0,T], \mathcal{B})$, the right-hand side is Lipschitz in t.

It follows from the results of §7.4 that the F-transforms satisfy the differential equations with respect to some distributional family of generalized laws $\overline{\sigma}$. We must show that $\overline{\sigma} = \sigma$. As in §7.4, let $\tau_{s,t}$ be the generalized law given by

$$F_{\mu_{s,t}}^{(n)}(z) = z - \mu_{s,t}(X)^{(n)} - G_{\tau_{s,t}}^{(n)}(z).$$

Recall that $\overline{\sigma}$ is characterized by

$$G^{(n)}_{\int_s^t \overline{\sigma}(\cdot,t)\,dt}(z) = G^{(n)}_{\tau_{s,t}}(z) + o(|t-s|) \text{ for every } z.$$

Therefore, to check that $\sigma = \overline{\sigma}$, it suffices to show that

$$G^{(n)}_{\tau_{s,t}}(z) = G^{(n)}_{\int_s^t \sigma(\cdot,t)\,dt}(z) + o(|t-s|) \text{ for every } z. \tag{8.5}$$

From the proof of Theorem 4.5.3, recall that if $Q = 1 - P_\xi$, then the generalized law $\tau_{s,t}$ is the distribution of the operator $QX_{s,t}Q$ with respect to the vector $QX_{s,t}\xi$. Now $QX_{s,t}\xi = \ell_{\text{ind}}(\chi_{[s,t]}\zeta)\xi = \zeta$ since the three terms in $X_{s,t}$ other than $\ell_{\text{ind}}(\chi_{[s,t]}\zeta)$ disappear when we apply them to ξ and then project onto $\{\xi\}^\perp$. And of course, $\|\chi_{[s,t]}\zeta\| = \|\int_s^t \sigma(1,t)\,dt\|^{1/2} = O(|t-s|^{1/2})$.

Meanwhile,

$$\begin{aligned} QX_{s,t}Q &= Q\mathfrak{m}_{\mathrm{ind}}(\chi_{[s,t]}Y)Q + Q\mathfrak{n}_{\mathrm{ind}}(\chi_{[s,t]}\beta)Q + Q\ell_{\mathrm{ind}}(\chi_{[s,t]}\zeta)Q + Q\ell_{\mathrm{ind}}(\chi_{[s,t]}\zeta)^*Q \\ &= Q\mathfrak{m}_{\mathrm{ind}}(\chi_{[s,t]}Y)Q + 0 + O(|t-s|^{1/2}), \end{aligned}$$

where the error estimate holds with respect to the operator norm. It follows that if $\operatorname{Im} z \geq \epsilon > 0$, then

$$(z - Q^{(n)}X_{s,t}^{(n)}Q^{(n)})^{-1} = (z - Q^{(n)}\mathfrak{m}_{\mathrm{ind}}(\chi_{[s,t]}Y)^{(n)}Q^{(n)})^{-1} + O(|t-s|^{1/2}),$$

using the standard resolvent-identity trick, where the error estimate depends implicitly upon ϵ. Hence, letting $Q^{(n)}X_{s,t}^{(n)}\xi^{(n)} = \chi_{[s,t]}\zeta^{(n)}$ be the vector in the $n \times n$ matrix amplification of $\mathcal{F}_{\mathrm{ind}}(\mathcal{H})$, then we have

$$\begin{aligned} G_{\tau_{s,t}}^{(n)}(z) &= \langle \chi_{[s,t]}\zeta^{(n)}, (z - Q^{(n)}X_{s,t}^{(n)}Q^{(n)})^{-1}\chi_{[s,t]}\zeta^{(n)}\rangle \\ &= \langle \chi_{[s,t]}\zeta^{(n)}, (z - Q^{(n)}\mathfrak{m}(\chi_{[s,t]}Y)^{(n)}Q^{(n)})^{-1}\chi_{[s,t]}\zeta^{(n)}\rangle + O(|t-s|^{1/2}\|\chi_{[s,t]}\zeta^{(n)}\|^2) \\ &= \langle \chi_{[s,t]}\zeta^{(n)}, (z - Q^{(n)}\mathfrak{m}_{\mathrm{ind}}(\chi_{[s,t]}Y)^{(n)}Q^{(n)})^{-1}\chi_{[s,t]}\zeta^{(n)}\rangle + O(|t-s|^{3/2}). \end{aligned}$$

Now $Q^{(n)}\mathfrak{m}_{\mathrm{ind}}(\chi_{[s,t]}Y)^{(n)}Q^{(n)}$ maps the subspace $\mathcal{H}^{(n)} \subseteq \mathcal{F}_{\mathrm{ind}}(\mathcal{H})^{(n)}$ into itself, and the restriction of $Q^{(n)}\mathfrak{m}_{\mathrm{ind}}(\chi_{[s,t]}Y)^{(n)}Q^{(n)}$ to $\mathcal{H}^{(n)}$ is exactly $\chi_{[s,t]}Y^{(n)}$. And the vector $\chi_{[s,t]}\zeta^{(n)}$ is in the subspace $\mathcal{H}^{(n)} \subseteq \mathcal{F}_{\mathrm{ind}}(\mathcal{H})^{(n)}$. Thus,

$$\begin{aligned} \langle \chi_{[s,t]}\zeta^{(n)}, (z - Q^{(n)}\mathfrak{m}_{\mathrm{ind}}(\chi_{[s,t]}Y)^{(n)}Q^{(n)})^{-1}\zeta^{(n)}\rangle &= \langle \chi_{[s,t]}\zeta^{(n)}, (z - \chi_{[s,t]}Y^{(n)})^{-1}\chi_{[s,t]}\zeta^{(n)}\rangle \\ &= G^{(n)}_{\int_s^t \sigma(\cdot,t)\,dt}(z). \end{aligned}$$

Thus, we have demonstrated (8.5) and finished the proof. □

Remark 8.5.2. In the theorem, we only constructed processes with independent increments where $X_0 = 0$ and hence $\mu_0 = \delta_0$. However, we can arrange an arbitrary initial condition μ_0 as follows. Let X_0 be an operator on a $\mathcal{B}$-$\mathcal{B}$-correspondence $\mathcal{K}$ with $\mathcal{B}$-central unit vector ζ which realizes the law μ_0. Let $X_{s,t}$ be the operator on the Fock space $\mathcal{F}_{\mathrm{ind}}(\mathcal{H})$ as above. Let $(\mathcal{G}, \eta)$ be the independent product of $(\mathcal{K}, \zeta)$ and $(\mathcal{F}_{\mathrm{ind}}(\mathcal{H}), \xi)$. Then the operators $\rho_{\mathrm{ind},1}(X_0) + \rho_{\mathrm{ind},2}(X_{0,t})$ for $t \in [0,T]$ are a process with independent increments which satisfy the evolution equations (7.1), such that the distribution of $\rho_{\mathrm{ind},1}(X_0)$ is μ_0.

CHAPTER 9

Examples and applications I

9.1 Convolution semigroups and Bercovici-Pata bijections

In the previous two chapters, we studied Lipschitz subordination families of $\mathcal{B}$-valued laws with respect to boolean, free, monotone, and anti-monotone independence. An important special case is a convolution semigroup, which has received a lot of attention in the literature (see Tables 7.1 and 8.1).

9.1.1 Differentiation for semigroups

For $\text{ind} \in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\,\dagger\}$, we say that $(\mu_t)_{t \in [0,+\infty)}$ is an ind-*convolution semigroup* if $\mu_s \boxplus_{\text{ind}} \mu_t = \mu_{s+t}$. A few preliminary observations about these semigroups:

- In this situation, $(\mu_t)_{t \in [0,T]}$ is a Lipschitz subordination family for any $T > 0$; indeed, we can take $\mu_{s,t} = \mu_{t-s}$ for $t > s$, and then $\mu_s \boxplus_{\text{ind}} \mu_{s,t} = \mu_t$. Moreover, the mean and variance of μ_t are given by $\mu_t(X) = t\mu_1(X)$ and $\operatorname{Var}(\mu_t) = t \operatorname{Var}(\mu_1)$, which are automatically Lipschitz in t.
- Note that $(\mu_t)_{t \in [0,T]}$ is a monotone convolution semigroup if and only if it is an anti-monotone convolution semigroup.

Since semigroups are Lipschitz subordination families, Theorems 7.4.1 and 8.5.1 apply in the situation of convolution semigroups. In fact, because of the extra symmetry of the semigroup setup, we can conclude that F_{μ_t} is differentiable in t in the pointwise sense (and even smooth). It is instructive to see how these results can be deduced from our more general theory, even though the case of semigroups could be handled alternatively without using distributional differentiation or the complicated Fock spaces coming from normal $L^\infty[0,T]$ actions (see especially [PV13] and [AW16]). The statement and proof given here are essentially from [Jek20, §5.4].

Theorem 9.1.1. *Let* $(\mu_t)_{t \in [0,+\infty)}$ *be an* ind-*convolution semigroup. Then there exists a* *unique self-adjoint* $b_0 \in \mathcal{B}$ *and* $\mathcal{B}$*-valued generalized law* σ_0 *such that*

$$\partial_t F_{\mu_t}(z) = \begin{cases} -[b_0 + G_{\sigma_0}(z)], & \textit{boolean case,} \\ -DF_{\mu_t}(z)[b_0 + G_{\sigma_0}(F_{X_t}(z))], & \textit{free case,} \\ -DF_{\mu_t}(z)[b_0 + G_{\sigma_0}(z)], & \textit{monotone case,} \\ -[b_0 + G_{\sigma_0}(F_{\mu_t}(z))], & \textit{anti-monotone case.} \end{cases} \tag{9.1}$$

The differentiation with respect to t *occurs pointwise with respect to the norm on* $M_n(\mathcal{B})$. *Moreover,* F_{μ_t} *is a* C^∞ *function of* t*, and for all* $k, \ell \in \mathbb{N}_0$,

$$\partial_t^\ell \Delta^k F_{\mu_t}(z_0, \dots, z_k)[w_1, \dots, w_k] = \Delta^k(\partial_t^\ell F_{\mu_t})(z_0, \dots, z_k)[w_1, \dots, w_k]. \tag{9.2}$$

Proof. Fix $T > 0$. By Theorem 7.4.1, there exists $b \in \mathcal{L}(L^1[0,T], \mathcal{B})$ self-adjoint and a distributional family of generalized laws σ satisfying the differential equation. Letting $b_0 = \mu_1(X)$, we have $\int_0^t b(s)\,ds = tb_0$ for all $t \in [0,T]$, and thus the distribution b is given by the constant function b_0 in the sense that $\int_0^T b(s)\phi(s)\,ds = (\int_0^T \phi) b_0$ for all $\phi \in L^1[0,T]$.

Similarly, we claim that σ is given by a generalized law σ_0 which is independent of t. Recall that in §7.4 we constructed σ as the limit over partitions $\mathcal{P}$ of distributional families of generalized laws $\sigma_{\mathcal{P}}$, which were given by

$$\sigma_{\mathcal{P}}(p(Y), t) = \sum_{j=1}^N \frac{1}{t_j - t_{j-1}} \tau_{t_{j-1}, t_j}(p(Y)) \chi_{[t_{j-1}, t_j)}(t),$$

where $\mathcal{P} = \{t_0, t_1, \dots, t_N\}$ and $\tau_{s,t}$ is given by $F_{\mu_{s,t}}(z) = z - \mu_{s,t}(X) - G_{\tau_{s,t}}$. Note that $\tau_{s,t} = \tau_{t-s}$ for $s \leq t$, where τ_t corresponds to μ_t in the same way. In particular, letting $\mathcal{P}_N = \{0, T/N, 2T/N, \dots, T\}$, we have

$$\sigma_{\mathcal{P}_N}(\cdot, t) = \frac{N}{T} \tau_{T/N}.$$

By Lemma 7.4.4, for $f \in \mathcal{B}\langle Y \rangle$ and $\phi \in L^1[0,T]$, we have

$$\int_0^T \sigma(f(Y), t)\phi(t)\,dt = \lim_{N \to \infty} \int_0^T \sigma_{\mathcal{P}_N}(f(Y), t)\,dt = \lim_{N \to \infty} \left(\int_0^T \phi(t)\,dt \right) \frac{N}{T} \tau_{T/N}(f(Y)).$$

This implies that $(N/T)\tau_{T/N}$ converges in moments to some generalized law σ_0 (since $\operatorname{rad}(\tau_{t/N})$ is uniformly bounded) and that $\int_0^T \sigma(\cdot, t)\phi(t)\,dt = (\int_0^T \phi(t)\,dt)\sigma_0$. Therefore, we have found a self-adjoint $b_0 \in \mathcal{B}$ and generalized law σ_0 satisfying the desired differential equation in a distributional sense. It follows from Theorem 7.4.1 that b and σ, and hence b_0 and σ_0, are uniquely determined by $(\mu_t)_{t \in [0,T]}$. In particular, this implies that b_0 and σ_0 are independent of T.

Next, we claim the differential equation holds in a pointwise sense, not merely a distributional sense. To prove this, note that the distribution $\partial_t F_{\mu_t}(z)$ is actually given by a pointwise-defined and continuous function of t, because $b + G_\sigma$ is independent of t and $DF_{\mu_t}(z)$ and $F_{\mu_t}(z)$ are locally Lipschitz in t. Therefore, F_{μ_t} is the antiderivative with respect to t in the distributional sense of a distribution which is a continuous function t, and so by the well-known argument to prove the fundamental theorem of calculus, we see that F_{μ_t} is continuously differentiable with respect to t. Since this holds for $t \in [0, T]$ and T is arbitrary, it holds for $t \in [0, +\infty)$.

Next, one can show by induction on ℓ that F_{μ_t} is a C^ℓ function of t and (9.2) holds for all k (we will not need to induct on k, however). We already showed above that $F_{\mu_t}(z)$ is C^1 in t. The equality of mixed partials (9.2) holds because $\Delta^k F_{\mu_t}(z_0, \dots, z_k)[w_1, \dots, w_k]$ is given as a matrix block of the evaluation of F_{μ_t} on a certain upper triangular matrix. For the induction step from ℓ to $\ell + 1$, one differentiates (9.1) ℓ times with respect to t (which makes sense by the induction hypothesis), and thus expresses $\partial_t^{\ell+1} F_{\mu_t}$ in terms of lower-order t-derivatives of F_{μ_t} and the spacial derivatives $\Delta^k F_{\mu_t}$ to show that F_{μ_t} is $C^{\ell+1}$ in t (we leave the details of this computation to the reader). And it follows that $\Delta^k F_{\mu_t}$ is also $C^{\ell+1}$ in t. □

9.1.2 Fock space construction for semigroups

Conversely, using Theorem 8.5.1, one can show that for every (b_0, σ_0), there exists a corresponding semigroup $(\mu_t)_{t \in [0,+\infty)}$ for each type of independence modeled by operators on the corresponding Fock space. This Fock space turns out to be much simpler to construct in the case of a semigroup; see the references listed in Table 8.1.

In Lemma 8.1.2, for each distributional family of generalized laws σ on $[0, T]$, we constructed a $\mathcal{B}$-$\mathcal{B}$-correspondence $\int_\oplus \mathcal{B}\langle Y \rangle \otimes_{\sigma(\cdot,t)} \mathcal{B}\, dt$ as the separation-completion of $\mathcal{L} \otimes_{\text{alg}} \mathcal{B}\langle Y \rangle \otimes_{\text{alg}} \mathcal{B}$ with respect to a certain inner product, where $\mathcal{L}$ is the algebra of simple functions in $L^\infty[0, T]$. But if $\sigma(\cdot, t) = \sigma_0(\cdot)$, then the inner product reduces to

$$\langle f_1 \otimes p_1(Y) \otimes b_1, f_1 \otimes p_2(Y) \otimes b_2 \rangle = \langle f_1, f_2 \rangle_{L^2[0,T]} \langle p_1(Y) \otimes b_1, p_2(Y) \otimes b_2 \rangle_{\mathcal{B}\langle Y \rangle \otimes_{\sigma_0} \mathcal{B}},$$

and hence

$$\int_\oplus \mathcal{B}\langle Y \rangle \otimes_{\sigma_0(\cdot,t)} \mathcal{B}\, dt \cong L^2[0, T] \otimes_{\mathbb{C}} (\mathcal{B}\langle Y \rangle \otimes_\sigma \mathcal{B}),$$

where the $\mathcal{B}$-$\mathcal{B}$-bimodule structure is the one inherited from $\mathcal{B}\langle Y \rangle \otimes_\sigma \mathcal{B}$. Moreover, the $L^\infty[0, T]$-action on this $\mathcal{B}$-$\mathcal{B}$-correspondence from Lemma 8.2.6 is the standard $L^\infty[0, T]$-action on $L^2[0, T]$ tensored with identity. Letting $\mathcal{H}_0 = \mathcal{B}\langle Y \rangle \otimes_{\sigma_0} \mathcal{B}$ and $\mathcal{H} = L^2[0, T] \otimes_{\mathbb{C}} \mathcal{H}_0$, it is straightforward to check that

$$\mathcal{H}^{\otimes_{\mathcal{B}} k} \cong (L^2[0, T])^{\otimes_{\mathbb{C}} k} \otimes_{\mathbb{C}} \mathcal{H}_0^{\otimes_{\mathcal{B}} k} \cong L^2([0, T]^k) \otimes_{\mathbb{C}} \mathcal{H}_0^{\otimes_{\mathcal{B}} k}.$$

Moreover, the $L^\infty([0, T]^k)$-action on $\mathcal{H}^{\otimes_{\mathcal{B}} k}$ from Lemma 8.2.7 is exactly the canonical $L^\infty([0, T]^k)$-action on $L^2([0, T]^k)$-action tensored with the identity on $\mathcal{H}_0^{\otimes_{\mathcal{B}} k}$.

Therefore, recalling the notation $E_{\text{ind}}(k,T)$ defined in §8.3, we have that

$$\mathcal{F}_{\text{ind}}(\mathcal{H}) \cong \mathcal{B} \oplus \bigoplus_{k\geq 1} L^2(E_{\text{ind}}(k,T)) \otimes_{\mathbb{C}} \mathcal{H}_0^{\otimes_{\mathcal{B}} k}.$$

Let ζ be the vector $1 \otimes 1 \otimes 1$ in $\mathcal{K}$, which we can express as $1 \otimes \zeta_0$, where $\zeta_0 = 1 \otimes 1$ in $\mathcal{K}_0$. Let Y be the operator on $\mathcal{H}$ given by the GNS construction, which we can express as the identity on $L^2[0,T]$ tensor the multiplication operator Y_0 on $\mathcal{H}_0$. Recall that the process with independent increments was given by

$$X_{s,t} = \mathfrak{n}_{\text{ind}}(\chi_{[s,t]} b_0) + \ell_{\text{ind}}(\chi_{[s,t]}\zeta) + \ell_{\text{ind}}(\chi_{[s,t]}\zeta)^* + \mathfrak{m}_{\text{ind}}(\chi_{[s,t]} Y). \tag{9.3}$$

Here each of the operators has a simpler form in the setting where b_0 and σ_0 are independent of t:

- The operator $\ell_{\text{ind}}(\chi_{[s,t]}\zeta)$ maps $L^2(E_{\text{ind}}(k,T)) \otimes_{\mathbb{C}} \mathcal{H}_0^{\otimes_{\mathcal{B}} k}$ into $L^2(E_{\text{ind}}(k+1,T)) \otimes_{\mathbb{C}} \mathcal{H}_0^{\otimes_{\mathcal{B}} k+1}$ by sending $f \otimes \eta$ (where $f \in L^2(E_{\text{ind}}(k,T))$ and $\eta \in \mathcal{H}_0^{\otimes_{\mathcal{B}} k}$) to $(\chi_{[s,t]} \otimes f)|_{E_{\text{ind}}(k+1,T)} \otimes (\zeta_0 \otimes \eta)$.
- Its adjoint can be described in a similar way.
- The multiplication $\mathfrak{m}_{\text{ind}}(\chi_{[s,t]} Y)$ acts on $L^2(E_{\text{ind}}(k,T)) \otimes_{\mathbb{C}} \mathcal{H}_0^{\otimes_{\mathcal{B}} k}$ by multiplying by $\chi_{[s,t]}$ in the first coordinate on $L^2(E_{\text{ind}}(k,T))$ and multiplying by Y_0 in the first coordinate on $\mathcal{H}_0^{\otimes_{\mathcal{B}} k}$.
- The operator $\mathfrak{n}(\chi_{[s,t]} b_0)$ is given by left multiplication by $(t-s)b_0 P_\xi$ in the boolean case and $(t-s)b_0$ in the free case. For the monotone case, it is $(t-s)b_0 P_\xi$ plus multiplication by the function $\phi(u) b_0$ on the direct summands $k \geq 1$ in the Fock space, where $\phi(u) = \int_u^T \chi_{[s,t]}(v)\,dv$. The anti-monotone case is the similar with $\phi(u) = \int_0^u \chi_{[s,t]}(v)\,dv$.

We claim that the law $\mu_{s,t}$ of the operator $X_{s,t}$ only depends on $t-s$. In light of Theorem 8.4.2, the law of $X_{s,t}$ is the same as the law of a creation-plus-annihilation-plus-multiplication operator on the Fock space over the subinterval $[s,t]$ rather than the entire interval $[0,T]$. But the Fock space over the subinterval $[s,t]$ is clearly isomorphic to the Fock space over the subinterval $[0,t-s]$ by a time translation, which also respects the creation, annihilation, and multiplication operators in a canonical way.

In fact, the Fock space can easily be extended to the time interval $[0,+\infty)$ rather than $[0,T]$. We did not address this in the general case in the last chapter simply to minimize the amount of technical explanation needed. But for the setting with b_0 and σ_0 constant, it is clear that if $E_{\text{ind}}(k,\infty) = \bigcup_{T>0} E_{\text{ind}}(k,T)$, then the Fock space on $[0,\infty)$ should be

$$\mathcal{B} \oplus \bigoplus_{k\geq 1} L^2(E_{\text{ind}}(k,\infty)) \otimes_{\mathbb{C}} \mathcal{H}_0^{\otimes_{\mathcal{B}} k}.$$

The creation, annihilation, and multiplication operators in (9.3) make perfect sense on this larger space. Letting μ_t be the law of $X_{0,t}$, we have that $(\mu_t)_{t\geq 0}$ is a convolution semigroup for the appropriate type of independence.

9.1.3 Bercovici-Pata bijections

The classification of semigroups leads to the following result. This is a summary of the outcome of a lot of prior work (see the references in 7.1).

Theorem 9.1.2. *For each* ind $\in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\,\dagger\}$*, there is a bijective correspondence between convolution semigroups* $(\mu_t)_{t\in[0,T]}$ *and pairs* (b_0, σ_0)*, where* $b_0 \in \mathcal{B}$ *is self-adjoint and* σ_0 *is a* $\mathcal{B}$*-valued generalized law. In particular, the convolution semigroups for free, boolean, and monotone independence exist in bijection with each other.*

Proof. Fix a type of independence. By the previous theorem, we have an injective map that sends a semigroup $(\mu_t)_{t\in[0,\infty)}$ to the corresponding pair (b_0, σ_0) satisfying the differential equation. This map is also surjective because the Fock space construction produces a semigroup $(\mu_t)_{t\in[0,\infty)}$ for each pair (b_0, σ_0). □

The bijection between semigroups and pairs (b_0, σ_0) is a version of the *Lévy-Hinčin formula*, while the bijections between the semigroups for different types of independence are called *Bercovici-Pata bijections* because of their relationship with the work of Bercovici and Pata [BP99] (see §1.3). However, let us reiterate that [BP99] not only proved bijections between semigroups (which could easily be deduced from the Lévy-Hinčin formulas), but also showed much deeper bijections between limit theorems for classical, free, and boolean independence, which we will not handle here.

In the scalar-valued setting, the Bercovici-Pata bijections work without any assumptions of bounded support or finite moments (although in general, we cannot express things in terms of a finite measure σ as we have done here). For the $\mathcal{B}$-valued setting, hardly anything is known about the case of measures with "unbounded support." However, under the assumption of bounded support, the Bercovici-Pata bijections between free, boolean, and monotone convolution semigroups were studied both for multivariable distributions and $\mathcal{B}$-valued distributions by the papers listed in Table 7.1. Theorem 9.1.2 thus summarizes the state of the art for $\mathcal{B}$-valued semigroups indexed by $t \in [0,\infty)$. However, we should also mention that in the free and boolean cases, we can take convolution powers of a distribution indexed by completely positive maps rather than real numbers; see [ABF13].

The results of the previous two chapters amount to a generalization of Theorem 9.1.2 to Lipschitz subordination families rather than semigroups. For instance, for the time interval $[0,T]$, we can state the result as follows. This follows from Theorems 7.4.1 and 8.5.1 by the same argument as above.

Theorem 9.1.3. *For each of the four types of independence studied here, the differential equations* (7.23) *define a bijection between Lipschitz subordination families* $(\mu_t)_{t\in[0,T]}$ *with* $\mu_t = \delta_0$ *and pairs* (b, σ) *where* $b \in \mathcal{L}(L^1[0,T], \mathcal{B})$ *is self-adjoint and* σ *is a distributional family of* $\mathcal{B}$*-valued generalized laws on* $[0,T]$*. In particular, the Lipschitz subordination families for the four types of independence are in bijection with each other.*

	boolean	free	(anti-)monotone
scalar-valued	[SW97, Thm. 3.4].	[VDN92, §3.5]	[Mur01]
best scalar-valued	[AS18]	[CG08]	[ASW19]
operator-valued	[BPV13, §2.1]	[Voi95, Thm. 8.4]	[BPV13, §2.3]
		[Spe98, §4.2]	[HS14, Thm. 3.6]

Table 9.1: References on non-commutative central theorems.

9.2 The central limit theorem

The central limit theorem of classical probability states that $X_1, \ldots, X_N$ are independent and identically distributed with mean zero and variance 1, and if $S_N = (X_1 + \cdots + X_N)/\sqrt{N}$, then the law of S_N approaches the standard normal distribution as $N \to \infty$. There is also a central limit theorem for each of the types of independence studied here. We list the references in Table 9.1 for the reader's convenience; the second row lists the sharpest known estimates for the scalar-valued setting.

9.2.1 The Bernoulli, semicircle, and arcsine laws

Which laws play the role of limiting distribution for the boolean, free, and monotone central limit theorems? In hindsight, the Bercovici-Pata bijection provides a clear heuristic. The normal distributions of mean zero and variance t form a classical convolution semigroup which corresponds in the classical version of the Lévy-Hinčin formula to the pair $(b_0, \sigma_0) = (0, \delta_0)$, and it is modeled on a symmetric Fock space.

Thus, our candidates for the central limit laws for boolean, free, and (anti-)monotone independence are as follows:

- In the boolean case, we take $(b_0, \sigma_0) = (0, \delta_0)$, and then by Proposition 7.4.5, this produces the boolean convolution semigroup of laws $\nu_{\text{bool},t}$ satisfying $K_{\nu_{\text{bool},t}}(z) = t/z$, which amounts to $\nu_{\text{bool},t} = \frac{1}{2}(\delta_{-t^{1/2}} + \delta_{t^{1/2}})$. This is the *Bernoulli law of mean zero and variance* t.

- In the free case, by Proposition 7.4.6, we get $\Phi_{\nu_{\text{free},t}}(z) = t/z$, which amounts to $F_{\mu_t}^{-1}(z) = z + t/z$. Inverting the function and then applying the Stieltjes inversion formula, we get
$$d\nu_{\text{free},t}(x) = \frac{1}{2\pi t}\sqrt{4t - x^2}\, 1_{|x|<2t^{1/2}}\, dx.$$
This is the *semicircle law of mean zero and variance* t.

- In the (anti-)monotone case, by Proposition 7.4.8, the F-transforms should satisfy the differential equation
$$\partial_t F_{\nu_{\text{mono},t}}(z) = -F_{\nu_{\text{mono},t}}(z)^{-1}.$$

This is an ODE so it clearly has a unique solution for z in the upper half-plane, and one can directly check that $F_{\nu_{\text{mono},t}}(z) = \sqrt{z^2 - 2t}$ is the solution. By Stieltjes inversion we get

$$d\nu_{\text{mono},t}(x) = \frac{1}{\pi\sqrt{2 - x^2}} 1_{|x|<\sqrt{2}}\, dx.$$

This is the *arcsine law of mean zero and variance* t, so named because the cumulative distribution function is $1/2 + (1/\pi)\arcsin(x/\sqrt{2})$ on $[-\sqrt{2}, \sqrt{2}]$.

We will not dwell on the details of these computations since they have been explained long ago in the references in Table 9.1, but move on immediately to the most general setting. In the $\mathcal{B}$-valued setting, the variance of a law μ is not a scalar, but rather a completely positive map $\operatorname{Var}(\mu) : \mathcal{B} \to \mathcal{B}$ given by $\operatorname{Var}(\mu)[b] = \mu[(X - \mu(X))b(X - \mu(X))]$.

Definition 9.2.1. For each $\text{ind} \in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\dagger\}$, each $b \in \mathcal{B}$, and each $\eta : \mathcal{B} \to \mathcal{B}$, we define $\nu_{\text{ind},b,\eta}$ as follows. Let $\widehat{\eta} : \mathcal{B}\langle Y\rangle \to \mathcal{B}$ be given by $\widehat{\eta}(p(Y)) = \eta(p(0))$, which is a $\mathcal{B}$-valued generalized law. Consider the ind-convolution semigroup corresponding to $(b, \widehat{\eta})$ in the Lévy-Hinčin formula, and then let $\nu_{\text{ind},b,\eta}$ be the law at $t = 1$ in this semigroup. We call $\nu_{\text{ind},b,\eta}$ the *$\mathcal{B}$-valued Bernoulli (resp. semicircle, arcsine) law of mean b and variance η* in the boolean (resp. free, monotone) case.

Proposition 9.2.2.

(1) For each $\text{ind} \in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\dagger\}$*, the laws* $(\nu_{\text{ind},tb,t\eta})_{t\in[0,\infty)}$ *form a convolution semigroup.*

(2) If $X \sim \nu_{\text{ind},b,\eta}$ *and* $c \in \mathbb{R}$*, then* $cX \sim \nu_{\text{ind},cb,c^2\eta}$.

(3) In the free and boolean cases, we have $\nu_{\text{ind},b_1,\eta_1} \boxplus_{\text{ind}} \nu_{\text{ind},b_2,\eta_2} = \nu_{\text{ind},b_1+b_2,\eta_2}$.

(4) $\operatorname{rad}(\nu_{\text{ind},b,\eta}) \leq 2\|\eta(1)\|^{1/2} + \|b\|$.

Proof. (1) Fix b and η and let $(\mu_t)_{t\in[0,\infty)}$ be the semigroup given by $(b, \widehat{\eta})$. We claim that $\mu_t = \nu_{\text{ind},tb,t\eta}$. This is because we can reparametrize time to map the interval $[0, t]$ to the interval $[0, 1]$. This results in rescaling the time derivative in (9.1) by a factor of t, which is equivalent to multiplying b and $\widehat{\eta}$ by t. Therefore, $(\nu_{\text{ind},tb,t\eta})_{t\in[0,\infty)} = (\mu_t)_{t\in[0,\infty)}$ is a convolution semigroup.

(2) Let $(X_t)_{t\in[0,\infty)}$ be a process with independent increments satisfying $X_t \sim \nu_{\text{ind},tb,t\eta}$. Then we have $F_{cX_t}(z) = cF_{X_t}(cz)$, and hence if F_{X_t} satisfies the differential equation (9.1) with $b + \widehat{\eta}(z^{-1})$, then F_{cX_t} satisfies it with $cb + c\widehat{\eta}((cz)^{-1}) = cb + c^2\widehat{\eta}(z^{-1})$.

(3) In the boolean case the K-transform of $\nu_{\text{bool},b,\eta}$ is $b + \eta(z^{-1})$ and in the free case the Φ-transform of $\nu_{\text{free},b,\eta}$ is $b + \eta(z^{-1})$, and the K-transform or Φ-transform respectively is additive under convolution.

(4) The central limit law is realized on the Fock space by $\mathfrak{n}_{\text{ind}}(\chi_{[0,1]}b) + \ell_{\text{ind}}(\chi_{[0,1]}\zeta) + \ell_{\text{ind}}(\chi_{[0,1]}\zeta)^*$, where $\langle\zeta, \zeta\rangle = \eta(1)$. Then observe that $\|\mathfrak{n}_{\text{ind}}(\chi_{[0,1]}b)\| \leq \|b\|$ and $\|\ell_{\text{ind}}(\chi_{[0,1]}\zeta)\| \leq$

$\|\zeta\|$ because $\ell_{\mathrm{ind}}(\chi_{[0,1]}\zeta)$ is a compression of $\ell_{\mathrm{free}}(\chi_{[0,1]}\zeta)$ and $\ell_{\mathrm{free}}(\chi_{[0,1]}\zeta)^*\ell_{\mathrm{free}}(\chi_{[0,1]}\zeta) = \langle \zeta, \zeta \rangle$. □

Remark 9.2.3. However, in the monotone case, we do not have in general that $\mu \triangleright \nu = \nu \triangleright \mu$, so there is no reason to expect the relation (3) to hold for general b_1, η_1, b_2, and η_2.

9.2.2 Central limit theorem via spatial coupling

The proof of the central limit theorem that we present here is based on "coupling" two different non-commutative laws together on the same Hilbert space rather than by computing moments or studying the analytic transforms. The results of this section are based on [JL19, §8.2] and [Jek20, §7], which was the first time such a coupling proof was presented as far as we know.

We will prove the following version of the central limit theorem. We provide this simple statement immediately as a goal of the section for the sake of exposition, although as we will explain later, the method of proof also yields more refined versions of the theorem. For a scalar c, we denote $\mathrm{dil}_c(\mu)$ the law given by $\mathrm{dil}_c(\mu)(f(X)) = \mu(f(cX))$. Thus, for instance, if $X_1, \dots, X_N$ are i.i.d., then the law of $(X_1 + \cdots + X_N)/N^{1/2}$ is the $N^{1/2}$ dilation of the N-fold convolution power of the law of X_1.

Theorem 9.2.4. *Let* $\mathrm{ind} \in \{\mathrm{bool}, \mathrm{free}, \mathrm{mono}, \mathrm{mono}\dagger\}$. *Let μ be a $\mathcal{B}$-valued non-commutative law with mean 0 and variance η. Let $\nu_{\mathrm{ind},0,\eta}$ be the law defined in the previous subsection. Then*

$$\|\mathrm{Mom}_k(\mathrm{dil}_{N^{-1/2}}(\mu^{\boxplus_{\mathrm{ind}} N})) - \mathrm{Mom}_k(\nu_{\mathrm{ind},0,\eta})\|_{\#} \leq 2N^{-1/2}(\mathrm{rad}(\mu)^k + \mathrm{rad}(\nu_{\mathrm{ind},0,\eta})^k).$$

Definition 9.2.5. Let σ_1 and σ_2 be $\mathcal{B}$-valued generalized laws with $\sigma_1|_{\mathcal{B}} = \sigma_2|_{\mathcal{B}}$. A *spatial coupling* of σ_1 and σ_2 is a tuple $(\mathcal{H}, \xi, X_1, X_2)$, where $\mathcal{H}$ is a $\mathcal{B}$-$\mathcal{B}$ correspondence, ξ is a vector in $\mathcal{H}$ and X_1 and X_2 are self-adjoint operators with distributions σ_1 and σ_2 with respect to the vector ξ. We define

$$d_{\mathrm{coup}}(\sigma_1, \sigma_2) := \inf\big\{\|X_1 - X_2\| : (\mathcal{H}, \xi, X_1, X_2) \text{ a spatial coupling of } \sigma_1, \sigma_2\big\}.$$

Note that $\sigma_1(1) = \sigma_2(1)$ is a necessary condition because $\sigma_j(b) = \langle \xi, b\xi \rangle$.

We do not claim that d_{coup} is a metric. To check the triangle inequality one would have to be able to "glue together" a spatial coupling of σ_1 and σ_2 and a spatial coupling of σ_2 and σ_3. (One could try to take an independent product amalgamated over the subalgebra generated by $\mathcal{B}$ and the operator X_2 for σ_2, but this runs into problems because the expectation might not be faithful as a $\mathcal{B}$-valued state, and because of the lack of orthogonal complements in C*-correspondences.) Despite possibly not being a metric, d_{coup} as we will see relates nicely to independent products and allows for an easy and direct proof of the central limit theorem (for measures with "bounded support") with explicit estimates on moments.

Lemma 9.2.6. *Let σ_1 and σ_2 be $\mathcal{B}$-valued generalized laws with $\sigma_1|_{\mathcal{B}} = \sigma_2|_{\mathcal{B}}$. Then*

$$\|\operatorname{Mom}_k(\operatorname{dil}_{N^{-1/2}}(\mu_N)) - \operatorname{Mom}_k(\nu_{\mathrm{ind},0,\eta})\| \leq 3k\|\eta(1)\|\big(2\|\eta(1)\|^{1/2} + N^{-1/2}\operatorname{rad}(\mu)\big)^{k-1}\operatorname{rad}(\mu).$$

Proof. Let $(\mathcal{H}, \xi, X_1, X_2)$ be a spatial coupling of σ_1 and σ_2. Then

$$\operatorname{Mom}_k(\sigma_j)[w_0, \dots, w_k] = \sigma_j(w_0 X_j w_1 \dots X_j w_k) = \langle \xi, w_0 X_j w_1 \dots X_j w_k \xi \rangle.$$

Now

$$\begin{aligned}
&\langle \xi, w_0 X_1 w_1 \dots X_1 w_k \xi \rangle - \langle \xi, w_0 X_2 w_1 \dots X_2 w_k \xi \rangle \\
\leq &\sum_{i=1}^k \|\langle \xi, w_0 X_1 \dots w_{i-1}(X_1 - X_2) w_i X_2 \dots w_k \rangle\| \\
\leq &\sum_{i=1}^k \|w_{i-1} X_1 \dots w_0 \xi\| \|X_1 - X_2\| \|w_i X_2 \dots w_k \xi\| \\
\leq &\|X_1 - X_2\| \sum_{i=1}^k \|\sigma(w_0^* X_1 \dots w_{i-1}^* w_{i-1} \dots X_1 w_0)\|^{1/2} \|\sigma(w_k^* X_1 \dots w_i^* w_i \dots X_1 w_k)\|^{1/2} \\
\leq &\|X_1 - X_2\| \|w_0\| \dots \|w_k\| \sum_{i=1}^k \|\sigma_1(1)\|^{1/2} \operatorname{rad}(\sigma_1)^{i-1} \|\sigma_2(1)\|^{1/2} \operatorname{rad}(\sigma_2)^{k-i} \\
\leq &\|X_1 - X_2\| \|w_0\| \dots \|w_k\| \|\sigma_1(1)\| k \max(\operatorname{rad}(\sigma_1), \operatorname{rad}(\sigma_2))^{k-1}.
\end{aligned}$$

This proves the desired bound on $\|\operatorname{Mom}_k(\sigma_1) - \operatorname{Mom}_k(\sigma_2)\|$. But since any coupling on $\mathcal{H}$ automatically produces a coupling on the matrix amplification $\mathcal{H}^{(n)}$, the same estimate holds with σ_j replaced by $\sigma_j^{(n)}$. Thus, our moment estimate holds for the completely bounded norm of $\operatorname{Mom}_k(\sigma_1) - \operatorname{Mom}_k(\sigma_2)$. ☐

Observation 9.2.7. *We have $d_{\mathrm{coup}}(\operatorname{dil}_c(\sigma_1), \operatorname{dil}_c(\sigma_2)) = c d_{\mathrm{coup}}(\sigma_1, \sigma_2)$.*

Lemma 9.2.8. *Let μ_1 and μ_2 be two non-commutative laws with the same mean and variance, and let (b_j, σ_j) correspond to μ_j as in Theorem 4.5.3 (so $b_1 = b_2$ and $\sigma_1|_{\mathcal{B}} = \sigma_2|_{\mathcal{B}}$). Then*

$$d_{\mathrm{coup}}(\mu_1, \mu_2) \leq d_{\mathrm{coup}}(\sigma_1, \sigma_2).$$

Proof. Let $(\mathcal{K}, \zeta, Y_1, Y_2)$ be a coupling of σ_1 and σ_2. Then define $\mathcal{H} = \mathcal{B}\xi \oplus \mathcal{K}$ and define X_j by

$$X_j(c\xi \oplus \kappa) = (b_j c + \langle \zeta, \kappa \rangle)\xi \oplus (\zeta c + Y_j \kappa).$$

for $c \in \mathcal{B}$ and $\kappa \in \mathcal{K}$ as in the proof of Theorem 4.5.3. Then we have $X_j \sim \mu_j$ and $\|X_1 - X_2\| = \|Y_1 - Y_2\|$. ☐

Lemma 9.2.9. *Let μ be a $\mathcal{B}$-valued law with mean b and variance η, and let (b,σ) correspond to μ as in Theorem 4.5.3. Then*

$$d_{\text{coup}}(\mu, \nu_{\text{bool},b,\eta}) \leq d_{\text{coup}}(\sigma, \widehat{\eta}) \leq \operatorname{rad}(\sigma),$$

where $\widehat{\eta}(p(Y)) = \eta(p(0))$.

Proof. Let $\mathcal{K} = \mathcal{B}\langle Y \rangle \otimes_\sigma \mathcal{B}$ and $\zeta = 1 \otimes 1 \in \mathcal{K}$. Then $(\mathcal{K}, \zeta, Y, 0)$ is a spatial coupling of σ and $\widehat{\eta}$. So $d_{\text{coup}}(\sigma, \widehat{\eta}) \leq \|Y\| = \operatorname{rad}(\sigma)$. The inequality $d_{\text{coup}}(\mu, \nu_{\text{bool},b,\eta}) \leq d_{\text{coup}}(\sigma, \widehat{\eta})$ follows from Lemma 9.2.8. □

Lemma 9.2.10. *Fix* ind $\in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\dagger\}$. *Let $\mu_1, \dots, \mu_N$ and $\nu_1, \dots, \nu_N$ be $\mathcal{B}$-valued laws such that for each j, the laws μ_j and ν_j have the same mean and variance, and let μ and ν be the* ind*-convolutions of $\mu_1, \dots, \mu_N$ and $\nu_1, \dots, \nu_N$ respectively.*

Let (b_j, σ_j) and (b_j, τ_j) correspond to μ_j and ν_j respectively as in Theorem 4.5.3, and let (b, σ) and (b, τ) correspond to μ and ν. Then we have

$$d_{\text{coup}}(\mu, \nu) \leq d_{\text{coup}}(\sigma, \tau) \leq \max_{j=1,\dots,N} d_{\text{coup}}(\sigma_j, \tau_j).$$

Proof. The inequality $d_{\text{coup}}(\mu, \nu) \leq d_{\text{coup}}(\sigma, \tau)$ follows from the previous lemma.

To prove the second inequality, let $(\mathcal{K}_j, \zeta_j, Y_j, Y_j')$ be a spatial coupling of (σ_j, τ_j). Let $\mathcal{H}_j = \mathcal{B}\xi_j \oplus \mathcal{K}_j$, and let $(\mathcal{H}_j, \xi_j, X_j, X_j')$ be the coupling of μ_j and ν_j as in the proof of the previous lemma. Then let $(\mathcal{H}, \xi)$ be the independent product of $(\mathcal{H}_1, \xi_1)$, ..., $(\mathcal{H}_N, \xi_N)$. Note that

$$X = \sum_{j=1}^N \rho_{\text{ind},j}(X_j), \qquad X' = \sum_{j=1}^N \rho_{\text{ind},j}(X_j')$$

are a coupling of μ and ν. Let $Q = 1 - P_\xi \in B(\mathcal{H})$, and let $\mathcal{K} = Q\mathcal{H} = \mathcal{H} \ominus \mathcal{B}\xi$. Let $Y = QXQ$ and $Y' = QX'Q$, and let ζ be the sum from $j = 1$ to $N$of the vectors $\zeta_j \in \mathcal{K}_j = \mathcal{H}_j^\circ \subseteq \mathcal{H}$. Then $(\mathcal{K}, \zeta, Y|_{\mathcal{K}}, Y'|_{\mathcal{K}})$ form a coupling of σ and τ as is clear from the proof of Theorem 4.5.3. Also, we have

$$Y = \sum_{j=1}^N \rho_{\text{ind},j}(Y_j), \qquad Y' = \sum_{j=1}^n \rho_{\text{ind},j}(Y_j').$$

Since $(Y_j - Y_j')\xi_j = 0$ for each j, we have by Lemma 6.1.1 that

$$\|Y - Y'\| \leq \max_{j=1,\dots,N} \|Y_j - Y_j'\|.$$

Since Y_j and Y_j' represented an arbitrary spatial coupling of σ_j and τ_j, we have

$$d_{\text{coup}}(\sigma, \tau) \leq \max_{j=1,\dots,N} d_{\text{coup}}(\sigma_j, \tau_j).$$ □

Proof of Theorem 9.2.4. Let μ have mean zero and variance η and let σ correspond to μ as in Theorem 4.5.3. By Lemma 9.2.9, $d_{\text{coup}}(\sigma, \widehat{\eta}) \leq \operatorname{rad}(\sigma)$.

Let $\mu_N = \mu^{\boxplus_{\text{ind}} N}$ and $\nu_N = \nu_{\text{bool},0,\eta}^{\boxplus_{\text{ind}} N}$. By Lemma 9.2.10,

$$d_{\text{coup}}(\mu_N, \nu_N) \leq \max_{j=1,\dots,N} d_{\text{coup}}(\sigma, \widehat{\eta}) \leq \operatorname{rad}(\sigma) \leq \operatorname{rad}(\mu).$$

Hence, by Observation 9.2.7,

$$d_{\text{coup}}(\operatorname{dil}_{N^{-1/2}}(\mu_N), \operatorname{dil}_{N^{-1/2}}(\nu_N)) \leq N^{-1/2} \operatorname{rad}(\mu).$$

So by Lemma 9.2.6,

$$\begin{aligned} &\|\operatorname{Mom}_k(\operatorname{dil}_{N^{-1/2}}(\mu_N)) - \operatorname{Mom}_k(\operatorname{dil}_{N^{-1/2}}(\nu_N))\| \\ &\qquad \leq k\|\eta(1)\| \max(\operatorname{rad}(\operatorname{dil}_{N^{-1/2}}(\mu_N)), \operatorname{rad}(\operatorname{dil}_{N^{-1/2}}(\nu_N)))^{k-1} \operatorname{rad}(\mu). \end{aligned}$$

However, by (6.1),

$$\operatorname{rad}(\mu_N) \leq 2\left\|\sum_{j=1}^N \eta(1)\right\|^{1/2} + \max_{j=1,\dots,N} \operatorname{rad}(\sigma) = 2N^{1/2}\|\eta(1)\|^{1/2} + \operatorname{rad}(\sigma),$$

and so

$$\operatorname{rad}(\operatorname{dil}_{N^{-1/2}}(\mu_N)) \leq 2\|\eta(1)\|^{1/2} + N^{-1/2} \operatorname{rad}(\sigma) \leq 2\|\eta(1)\|^{1/2} + N^{1/2} \operatorname{rad}(\mu).$$

By the same reasoning, since $\operatorname{rad}(\widehat{\eta}) = 0$, we have

$$\operatorname{rad}(\operatorname{dil}_{N^{-1/2}}(\nu_N)) \leq 2\|\eta(1)\|^{1/2}.$$

Therefore,

$$\begin{aligned} &\|\operatorname{Mom}_k(\operatorname{dil}_{N^{-1/2}}(\mu_N)) - \operatorname{Mom}_k(\operatorname{dil}_{N^{-1/2}}(\nu_N))\| \\ &\qquad \leq k\|\eta(1)\|\big(2\|\eta(1)\|^{1/2} + N^{-1/2} \operatorname{rad}(\mu)\big)^{k-1} \operatorname{rad}(\mu). \end{aligned}$$

We can apply the same reasoning with μ replaced by $\mu' = \nu_{\text{ind},0,\eta}$, and $\mu'_N = (\mu')^{\boxplus_{\text{ind}} N}$. But note that $\operatorname{dil}_{N^{-1/2}}(\mu'_N) = \mu = \nu_{\text{ind},0,\eta}$ and $\operatorname{rad}(\nu_{\text{ind},0,\eta}) \leq 2\|\eta(1)\|^{1/2}$ by Lemma 9.2.2. Therefore, we get

$$\begin{aligned} &\|\operatorname{Mom}_k(\nu_{\text{ind},0,\eta}) - \operatorname{Mom}_k(\operatorname{dil}_{N^{-1/2}}(\nu_N))\| \\ &\quad \leq k\|\eta(1)\|\big(2\|\eta(1)\|^{1/2}\big)^{k-1}(2\|\eta(1)\|^{1/2}) \\ &\quad \leq 2k\|\eta(1)\|\|\eta(1)\|\big(2\|\eta(1)\|^{1/2} + N^{-1/2} \operatorname{rad}(\mu)\big)^{k-1} \operatorname{rad}(\mu). \end{aligned}$$

Therefore, by the triangle inequality,

$$\|\operatorname{Mom}_k(\operatorname{dil}_{N^{-1/2}}(\mu_N)) - \operatorname{Mom}_k(\nu_{\text{ind},0,\eta})\| \leq 3k\|\eta(1)\|\big(2\|\eta(1)\|^{1/2} + N^{-1/2} \operatorname{rad}(\mu)\big)^{k-1} \operatorname{rad}(\mu).$$

□

9.2.3 Central limit theorem for general mean and variance

There are several ways in which we can generalize Theorem 9.2.4 using the same proof. First, we do not have to consider only convolutions of the same law. Rather, since Lemma 9.2.10 does not require the μ_j's to be equal, we may compare $\operatorname{dil}_{N^{-1/2}}(\mu_1 \boxplus_{\text{ind}} \cdots \boxplus_{\text{ind}} \mu_N)$ with $\nu_{\text{ind},0,\eta}$ whenever the μ_j's all have mean zero and variance η. In this situation, the term $\operatorname{rad}(\mu)$ in Theorem 9.2.4 is simply replaced by $\max_j \operatorname{rad}(\mu_j)$.

In the free and boolean cases, we do not have to assume that the variances are the same either. If μ_j has variance η_j, then we can compare $\operatorname{dil}_{N^{-1/2}}(\mu_1 \boxplus_{\text{ind}} \cdots \boxplus_{\text{ind}} \mu_N)$ with

$$\operatorname{dil}_{N^{-1/2}}(\nu_{\text{ind},0,\eta_1} \boxplus_{\text{ind}} \cdots \boxplus_{\text{ind}} \nu_{\text{ind},0,\eta_N}) = \nu_{\text{ind},0,(1/N)\sum_j \eta_j}.$$

Here the term η in Theorem 9.2.4 is indeed replaced by $(1/N)\sum_j \eta_j$, as one can check by examining the application of (6.1) in the proof.

In the (anti-)monotone case, this breaks down because $\operatorname{dil}_{N^{-1/2}}(\nu_{\text{ind},0,\eta_1} \rhd \cdots \rhd \nu_{\text{ind},0,\eta_N})$ may not be the operator-valued arcsine law itself. However, we could view this law as a sort of "generalized $\mathcal{B}$-valued arcsine law." It can be modeled on a Fock space on the time interval $[0,1]$ corresponding to the distributional family of generalized laws $\sigma(\cdot,t) = \sum_{j=1}^N \chi_{[(j-1)/N,j/N)}(t)\widehat{\eta}_j$.

Moreover, one can handle the case of nonzero mean as well. In the boolean and free cases, we can subtract off the mean from any operator and thus express the CLT in terms of the mean zero case. More precisely, in the boolean case, one should subtract off P_ξ times the mean, but in the free case one should subtract off the mean itself (as a left multiplication operator with respect to the left $\mathcal{B}$-module structure of the underlying correspondence). However, in the monotone case, one runs into the same issue that $\nu_{\text{mono},b_1,\eta_1} \rhd \nu_{\text{mono},b_2,\eta_2}$ is not necessarily $\nu_{\text{mono},b_1+b_2,\eta_1+\eta_2}$. Actually, this fails even in the scalar-valued setting with the same variance and different means, as we encourage the reader to verify on their own. The best solution seems to be expand our notion of what is considered a valid central limit law to include "generalized arcsine laws" of the form

$$\nu_{\text{mono},b_1,\eta_1} \rhd \cdots \rhd \nu_{\text{mono},b_N,\eta_N}.$$

Then one can apply the same proof strategy as Theorem 9.2.4, except that the estimates on the radius of the laws will be different and not as sharp. We leave the details as an exercise.

9.2.4 Central limit theorem for general test functions

We can also generalize the central limit theorem by considering other types of test functions rather than more general collections of laws. Theorem 9.2.4 merely states a bound on the difference in kth moments for $\mu_N := \operatorname{dil}_{N^{-1/2}}(\mu^{\boxplus_{\text{ind}} N})$ and $\nu_{\text{ind},0,\eta}$. This implies a bound for $\mu_N(f) - \nu_{\text{ind},0,\eta}(f)$ for a non-commutative polynomial f by writing it as a sum of monomials, but this bound may not be sharp for general f. Furthermore, we may want to consider f in some completion of polynomials, such as fully matricial functions in some ball around the origin, or even a non-commutative C^k function.

In §3.6, we saw that $f(X) \in \mathcal{B}\langle X \rangle$ can be interpreted as a fully matricial function of the variable X as X ranges over any given C*-algebra $\mathcal{A} \subseteq \mathcal{B}$. Moreover, given X_1 and $X_2 \in \mathcal{A}$, we have

$$f(X_1) - f(X_2) = \Delta f(X_1, X_2)[X_1 - X_2].$$

A version of non-commutative C^1 functions, and in fact more generally non-commutative C^k functions, on $[-R, R]$ can be defined by taking the supremum of $\Delta^k f(X_1, X_2)$ over all self-adjoint X_1 and X_2 bounded by R. Specifically, define

$$\begin{aligned} &\|\Delta^k f\|_R^* := \\ &\sup\{\|\Delta^k f(X_0, \dots, X_k)\|_\# : \mathcal{A} \subseteq \mathcal{B} \text{ a C}^*\text{-algebra}, X_1, \dots X_k \in \mathcal{A} \text{ self-adjoint}, \|X_j\| \leq R\}. \end{aligned}$$

(The collection of probability spaces is not a set. However, we can rephrase the definition by taking the supremum over all possible joint laws of X_0, …, X_k. The space of joint laws is a set because it consists of functions from a formal polynomial algebra into $\mathcal{B}$.) We can then define $C^k_{\text{nc}}(\mathcal{B}, R)$ to be the completion of the non-commutative polynomials with respect to the norm

$$\|f\|_{C^k_{\text{nc}}(\mathcal{B},R)} = \sum_{j=0}^k \|\Delta^j f\|_R^*$$

It is clear that for $\mathcal{B}$-valued laws μ and ν with $\operatorname{rad}(\mu) \leq R$ and $\operatorname{rad}(\nu) \leq R$ and for a non-commutative polynomial f, we have

$$\|\mu_1(f(X)) - \mu_2(f(X))\| \leq \|\Delta f\|_R^* d_{\text{coup}}(\sigma_1, \sigma_2),$$

since if $(\mathcal{H}, \xi, X_1, X_2)$ are a spatial coupling of μ_1 and μ_2, then we can use $f(X_1) - f(X_2) = \Delta f(X_1, X_2)[X_1 - X_2]$ and applying the definition of $\|\Delta f\|_R^*$ with $\mathcal{A} = B(\mathcal{H})$. This inequality will automatically extend to f in the completion $C^k_{\text{nc}}(\mathcal{B}, R)$.

Hence, for instance, suppose μ_1, …, μ_N have mean zero and variance η and $\mu = \operatorname{dil}_{N^{-1/2}}(\mu_1 \boxplus_{\text{ind}} \cdots \boxplus_{\text{ind}} \mu_N)$. Suppose $R > 2\|\eta(1)\| + N^{-1/2} \max_{j=1,\dots,N} \operatorname{rad}(\mu_j)$. Then

$$\|\mu(f(X)) - \nu_{\text{ind},0,\eta}(f(X))\| \leq 2\|\Delta f\|_R^* \operatorname{rad}(\mu),$$

and this holds for all $f \in C^1_{\text{nc}}(\mathcal{B}, R)$. The same method can be applied for different variances and different means with all the considerations we discussed before.

But the question remains of how to compute or estimate $\|f\|_{C^k_{\text{nc}}(\mathcal{B},R)}$ when f is a non-commutative polynomial or more generally some power series. Of course, it is not hard to find upper bounds when f is a monomial. Moreover, in the case $\mathcal{B} = \mathbb{C}$, we can consider applying a smooth function $\phi : \mathbb{R} \to \mathbb{R}$ to self-adjoint operators. Methods for estimating the non-commutative derivatives for ϕ will be explained in §18.1.

At present, we will conclude with central limit estimates for the Cauchy-Stieltjes transform and (in the scalar-valued case) the Fourier transform of laws obtained by iterated convolution.

Proposition 9.2.11. *Let $\mu_1, \dots, \mu_N$ and $\nu_1, \dots, \nu_N$ be $\mathcal{B}$-valued non-commutative laws such that $\mu_j(X) = \nu_j(X)$ and $\operatorname{Var}(\mu_j) = \operatorname{Var}(\nu_j)$ for each j. Let*

$$\mu = \operatorname{dil}_{N^{-1/2}}(\mu_1 \boxplus_{\text{ind}} \cdots \boxplus_{\text{ind}} \mu_N)$$
$$\nu = \operatorname{dil}_{N^{-1/2}}(\nu_1 \boxplus_{\text{ind}} \cdots \boxplus_{\text{ind}} \nu_N)$$

Let R be any common upper bound for $\operatorname{rad}(\mu_j)$ and $\operatorname{rad}(\nu_j)$ and $\operatorname{rad}(\mu)$ and $\operatorname{rad}(\nu)$. Suppose that $z \in M_n(\mathcal{B})$ with either $\operatorname{Im} z \geq \epsilon$ or $\|z^{-1}\| \leq 1/(R+\epsilon)$. Then we have

$$\|G_\mu(z) - G_\nu(z)\| \leq \frac{2R}{N^{1/2}\epsilon^2}$$
$$\|F_\mu(z) - F_\nu(z)\| \leq \frac{2R\|\operatorname{Var}(\mu)[1]\|}{N^{1/2}\epsilon^2}$$
$$\|G_\mu(z) - G_\nu(z)\| \leq \frac{2R\|\operatorname{Var}(\mu)[1]\|}{N^{1/2}\epsilon^4}$$

Proof. If σ_1 and σ_2 are $\mathcal{B}$-valued generalized laws with $\sigma_1|_{\mathcal{B}} = \sigma_2|_{\mathcal{B}}$ and $\operatorname{rad}(\sigma) \leq R$ and $\operatorname{rad}(\sigma') \leq R$, then we claim that

$$\|G_{\sigma_1}(z) - G_{\sigma_2}(z)\| \leq \frac{\|\sigma_1(1)\|}{\epsilon^2} d_{\text{coup}}(\sigma, \sigma') \tag{9.4}$$

whenever $\operatorname{Im} z \geq \epsilon$ or more generally when $\|z^{-1}\| \leq 1/(R+\epsilon)$. To prove this, let $(\mathcal{H}, \xi, Y_1, Y_2)$ be a spatial coupling of σ_1 and σ_2. Note that for $\operatorname{Im} z$ bounded below by some constant, we have

$$\begin{aligned} G_{\sigma_1}(z) - G_{\sigma_2}(z) &= \langle \xi, [(z - Y_1)^{-1} - (z - Y_2)^{-1}]\xi \rangle \\ &= \langle (z^* - Y_1)^{-1}\xi, (Y_1 - Y_2)(z - Y_2)^{-1}\xi \rangle. \end{aligned}$$

Now there is an isometric $\mathcal{B}$-$\mathcal{B}$-bimodule map from $\mathcal{B}\langle Y \rangle \otimes_{\sigma_j} \mathcal{B}$ to $\mathcal{H}$ sending $1 \otimes 1$ to ξ, and therefore, the cyclic subspace of $\mathcal{H}$ generated by $\mathcal{B}\langle Y_j \rangle$ acting on ξ on the left and $\mathcal{B}$ acting on the right is isomorphic to $\mathcal{B}\langle Y \rangle \otimes_{\sigma_j} \mathcal{B}$. Hence, $z \mapsto (z - Y_j)^{-1}\xi$ extends to be analytic whenever $z - Y$ is invertible on $\mathcal{B}\langle Y \rangle \otimes_{\sigma_j} \mathcal{B}$. In particular, when $\|z^{-1}\| \leq 1/(R+\epsilon)$ or $\operatorname{Im} z \geq \epsilon$, then we have by analytic continuation that

$$\begin{aligned} \|G_{\sigma_1}(z) - G_{\sigma_2}(z)\| &\leq \|(z^* - Y_1)^{-1}\xi\| \|Y_1 - Y_2\| \|(z - Y_2)^{-1}\xi\| \\ &\leq \frac{\|\xi\|^2}{\epsilon^2} \|Y_1 - Y_2\|. \end{aligned}$$

Since the coupling was arbitrary and since $\langle \xi, \xi \rangle = \sigma_1(1)$, we obtain (9.4).

Now to prove the first estimate on $G_\mu - G_\nu$, let λ be the $N^{-1/2}$ dilation of the convolution of the Bernoulli distributions $\nu_{\text{bool},b_j,\eta_j}$ with the same mean and variance as μ_j. Then by Lemmas 9.2.9 and 9.2.10, we have $d_{\text{coup}}(\mu, \lambda) \leq N^{-1/2}R$. Hence, $\|G_\mu(z) - G_\lambda(z)\| \leq R/N^{1/2}\epsilon^2$ whenever $\operatorname{Im} z \geq \epsilon$ or $\|z^{-1}\| \leq 1/(R + \epsilon$ by (9.4). The same estimate applies to $\|G_\nu(z) - G_\lambda(z)\|$, and so we conclude the proof by the triangle inequality.

For the second estimate, let ρ, σ, and τ be the generalized laws that correspond to λ, μ, and ν under Theorem 4.5.3. Since μ and λ have the same mean, we have

$$F_\mu(z) - F_\lambda(z) = -G_\sigma(z) + G_\rho(z).$$

We also have $d_{\text{coup}}(\sigma, \rho) \leq N^{-1/2}R$ by Lemma 9.2.10. Thus, using (9.4), we get $\|G_\sigma(z) - G_\rho(z)\| \leq R\|\operatorname{Var}(\mu)[1]\|/N^{1/2}\epsilon^2$ for $\operatorname{Im} z \geq \epsilon$ or $\|z^{-1}\| \leq 1/(R + \epsilon)$.

For the third estimate, observe that

$$G_\mu(z) - G_\nu(z) = G_\mu(z)(F_\nu(z) - F_\mu(z))G_\nu(z).$$

Then apply the previous estimate with the fact that $\|G_\mu(z)\| \leq 1/\epsilon$ and $\|G_\nu(z)\| \leq 1/\epsilon$ for $\operatorname{Im} z \geq \epsilon$ or $\|z^{-1}\| \leq 1/(R + \epsilon)$. □

Proposition 9.2.12. *Let μ and ν be compactly supported probability measures on the real line. Let $\mathcal{F}\mu$ and $\mathcal{F}\nu$ denote their Fourier transforms, so that for instance $\mathcal{F}\mu(t) = \int_{\mathbb{R}} e^{2\pi itx}\, d\mu(x)$. Then*

$$|\mathcal{F}\mu(t) - \mathcal{F}\nu(t)| \leq 2\pi|t| d_{\text{coup}}(\mu, \nu).$$

In particular, if μ and ν are the $N^{-1/2}$ of the convolution of μ_1, ..., μ_N and ν_1, ..., ν_N respectively, and if R is an upper bound for the radii of μ_j and ν_j, then

$$|\mathcal{F}\mu(t) - \mathcal{F}\nu(t)| \leq \frac{4\pi R|t|}{N^{1/2}}.$$

Proof. The second claim follows from the first by the same reasoning as we have used before throughout the section. To prove the first claim, note that if $(\mathcal{H}, \xi, X, Y)$ is a coupling of μ and ν, then

$$\mathcal{F}\mu(t) - \mathcal{F}\nu(t) = \langle \xi, (e^{2\pi itX} - e^{2\pi itY})\xi \rangle.$$

Thus, it suffices to show that $\|e^{2\pi itX} - e^{2\pi itY}\| \leq 2\pi|t|\|X - Y\|$.

Let us give a short argument here for this fact, although we will revisit the argument in greater generality in §18.1. Since X and Y are arbitrary bounded self-adjoint operators, we can reduce by rescaling to the claim that $\|e^{iX} - e^{iY}\| \leq \|X - Y\|$. For $n \in \mathbb{N}$, observe that

$$e^{iX} - e^{iY} = \sum_{j=1}^n [e^{ijX/n} e^{i(n-j)Y/n} - e^{i(j-1)X/n} e^{i(n-j+1)Y/n}] = \sum_{j=1}^n e^{i(j-1)X/n} [e^{iX/n} - e^{iY/n}] e^{i(n-j)Y/n}.$$

Hence, since $e^{i(j-1)X/n}$ and $e^{i(n-j)Y/n}$ are unitary, we have

$$\|e^{iX} - e^{iY}\| \leq n\|e^{iX/n} - e^{iY/n}\|.$$

But looking at the power series expansions of $e^{iX/n}$ and $e^{iY/n}$, we see that $n(e^{iX/n} - e^{iY/n}) \to i(X - Y)$ in operator norm as $n \to \infty$, and hence $\|e^{iX} - e^{iY}\| \leq \|X - Y\|$. □

9.3 Spatial coupling for subordination families

9.3.1 General estimates

The same coupling techniques that we used for central limit sums can also be used for processes with independent increments. The intuition is that X_t represents the "sum" of the family of random variables $(X_s)_{0 \leq s \leq t}$, and the Fock space can also be thought of the "independent product of a family of correspondences indexed by $t \in [0,T]$." We will not attempt to make this intuition precise here, but rather proceed using the formalism of normal $L^\infty[0,T]$ actions from §8.2. The analogue of Definition 9.2.5 for distributional families of generalized laws is as follows.

Definition 9.3.1. Let $\sigma_1, \sigma_2 : \mathcal{B}\langle Y \rangle \otimes_{\text{alg}} L^1[0,T] \to \mathcal{B}$ be distributional families of generalized laws with $\sigma_1|_{\mathcal{B} \otimes_{\text{alg}} L^1[0,T]} = \sigma_2|_{\mathcal{B} \otimes_{\text{alg}} L^1[0,T]}$. A *spatial coupling* of σ_1 and σ_2 is a tuple $(\mathcal{H}, Y_1, Y_2, \zeta)$, where $\mathcal{H}$ is a $\mathcal{B}$-$\mathcal{B}$ correspondence with a normal $L^\infty[0,T]$ action, Y_1 and Y_2 are operators in $B(\mathcal{H})$ that commute with the $L^\infty[0,T]$ action, and ζ is a vector in $\mathcal{H}$ such that such that for $j = 1, 2$, for every Borel set E and $p \in \mathcal{B}\langle Y \rangle$,

$$\langle \xi, p(Y_j) \chi_E \xi \rangle = \int_E \sigma_j(p(Y), t)\, dt.$$

We define

$$d_{\text{coup}}(\sigma_1, \sigma_2) := \inf \left\{ \|Y_1 - Y_2\| : (\mathcal{H}, Y_1, Y_2, \zeta) \text{ a spatial coupling of } \sigma_1, \sigma_2 \right\}.$$

Suppose we have a spatial coupling $(\mathcal{H}, Y_1, Y_2, \zeta)$ of distributional families σ_1 and σ_2 as above. Let $b_1, b_2 \in \mathcal{L}(L^1[0,T], \mathcal{B})$. Then for $0 \leq s \leq t \leq T$, define operators $X_{s,t}^{(1)}$ and $X_{s,t}^{(2)}$ on $\mathcal{F}_{\text{ind}}(\mathcal{H})$ by

$$X_{s,t}^{(j)} = \mathfrak{n}_{\text{ind}}(\chi_{[s,t]} b_j) + \ell_{\text{ind}}(\chi_{[s,t]}\zeta) + \ell_{\text{ind}}(\chi_{[s,t]}\zeta)^* + \mathfrak{m}_{\text{ind}}(\chi_{[s,t]} Y_j),$$

as in Theorem 8.5.1. Then it is clear that

$$\begin{aligned} \left\| X_{s,t}^{(1)} - X_{s,t}^{(2)} \right\| &\leq \left\| \mathfrak{n}_{\text{ind}}(\chi_{[s,t]}(b_1 - b_2)) \right\| + \left\| \mathfrak{m}_{\text{ind}}(\chi_{[s,t]}(Y_1 - Y_2)) \right\| \\ &\leq (t-s) \|b_1 - b_2\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + \|Y_1 - Y_2\|. \end{aligned}$$

Given our earlier results Theorems 7.4.1 and 8.5.1, this implies the following estimate.

Proposition 9.3.2. *Let* $\text{ind} \in \{\text{bool}, \text{free}, \text{mono}, \text{mono}\dagger\}$. *Let* σ_1 *and* σ_2 *be families of distributional generalized laws on* $[0,T]$ *with* $\sigma_1|_{\mathcal{B} \otimes_{\text{alg}} L^1[0,T]} = \sigma_2|_{\mathcal{B} \otimes_{\text{alg}} L^1[0,T]}$, *and let* $b_1, b_2 \in \mathcal{L}(L^1[0,T], \mathcal{B})$. *Let* $\mu_{s,t}^{(j)}$ *be the* $\mathcal{B}$*-valued law of the operator* $X_{s,t}^{(j)}$ *defined above, so that, in particular,* $(\mu_{0,t}^{(j)})_{t \in [0,T]}$ *is the Lipschitz* ind*-subordination family corresponding to* σ_j *under Theorem 8.5.1. Then we have*

$$d_{\text{coup}}\left(\mu_{s,t}^{(1)}, \mu_{s,t}^{(2)}\right) \leq (t-s)\|b_1 - b_2\|_{\mathcal{L}(L^1[0,T],\mathcal{B})} + d_{\text{coup}}(\sigma_1, \sigma_2).$$

As a consequence, we can obtain estimates for $E[f(X_{s,t}^{(1)})] - E[f(X_{s,t}^{(1)})]$ for various functions f using the same reasoning as in §9.2. On the other hand, $E[f(X_{s,t}^{(1)})] - E[f(X_{s,t}^{(1)})]$ could have been estimated analytically by studying the differential equations for the F-transforms of $\mu_{s,t}^{(j)}$; for instance, see [Jek20, §7] for a comparison of the two techniques for estimating the difference between $F_{\mu_{s,t}^{(1)}}$ and $F_{\mu_{s,t}^{(2)}}$ in the monotone case. The advantage of the coupling technique are that it gives us fairly sharp estimates, and we do not have to do any more work to prove them at this point. We are merely collecting on our investment in Fock space models in §8.

Remark 9.3.3. Definition 9.3.1 assumes that $\int_E \sigma_1(\cdot, t)\, dt$ and $\int_E \sigma_2(\cdot, t)$ are the same for every Borel set E, which is equivalent to $\mu_{s,t}^{(1)}$ and $\mu_{s,t}^{(2)}$ in Proposition 9.3.2 having the same variance for every s, t. In order to study the case where the variances do not agree, we could use spatial couplings of σ_1 and σ_2 which use two different vectors ζ_1 and ζ_2 for the two families of generalized laws.

Remark 9.3.4. Due to the general $\mathcal{B}$-valued setting, we have defined $d_{\mathrm{coup}}(\sigma_1, \sigma_2)$ abstractly using normal $L^\infty[0,T]$ actions on $\mathcal{B}$-$\mathcal{B}$-correspondences. However, under certain assumptions, the distributional families of generalized laws and the coupling distance can be understood using more standard measure-theoretic techniques.

For example, suppose that $\mathcal{B}$ is finite-dimensional. Then using Lebesgue differentiation theory, every distributional family of laws σ_j will be given by a family of generalized laws $(\sigma_j(\cdot, t))_{t \in [0,T]}$ that is defined pointwise for almost every t, such that $t \mapsto \sigma_j(p(Y), t)$ is measurable for every p. Thus, $\int_0^T \mathcal{B}\langle Y \rangle \otimes_{\sigma_j(\cdot,t)} \mathcal{B}\, dt$ will be an honest direct integral of $\mathcal{B}$-$\mathcal{B}$-correspondences.

Moreover, suppose that $\mathcal{H}$ is a direct integral of $\mathcal{B}$-$\mathcal{B}$-correspondences $(\mathcal{H}_t)_{t \in [0,T]}$. If $(Y_t^{(j)})_{t \in [0,T]}$ is a measurable family of operators on $\mathcal{H}_t$ for each j and $(\zeta_t)_{t \in [0,T]}$ is a measurable family of vectors such that $(\mathcal{H}_t, Y_t^{(1)}, Y_t^{(2)}, \zeta_t)$ is a coupling of $\sigma_1(\cdot, t)$ and $\sigma_2(\cdot, t)$ for almost every $t \in [0,T]$, then the direct integrals $Y_j = \int_0^T Y_t^{(j)}\, dt$ and $\zeta = \int_0^T \zeta_t\, dt$ furnish a spatial coupling of the distributional families σ_1 and σ_2 in the sense of Definition 9.3.1.

9.3.2 The example of Loewner chains driven by a function

To bring Theorem 8.5.1 and Proposition 9.3.2 down to earth, let us consider an explicit complex-analytic application to Loewner chains in the upper half-plane driven by a function. Consider $\mathcal{B} = \mathbb{C}$, so that in particular a $\mathcal{B}$-$\mathcal{B}$-correspondence reduces to a Hilbert space, and a $\mathcal{B}$-valued law reduces to a compactly supported measure on $\mathbb{R}$.

Let $\psi : [0,T] \to \mathbb{R}$ be a bounded measurable function. Then the *chordal Loewner equation with driving function* ψ is the equation

$$\partial_t F_t(z) = -\frac{F_t'(z)}{z - \psi(t)}$$

for a family of functions $F_t : \mathbb{H}_+ \to \mathbb{H}_+$. This is a special case of the monotone version of (7.1) where $b(t) = 0$, and where the distributional family σ is given by

$$\sigma(\cdot, t) = \delta_{\psi(t)},$$

or more explicitly

$$\int_0^T \phi(t)\sigma(p(Y), t)\, dt = \int_0^T \phi(t) p(\psi(t))\, dt \text{ for } \psi \in L^1[0, T],$$

and hence $G_{\sigma(\cdot,t)}(z) = 1/(z - \psi(t))$.

The Fock space construction used in Theorem 8.5.1 can be evaluated in the following way. Note that $\mathcal{H} = L^2[0, T]$ has a normal action of $L^\infty[0, T]$ given by multiplication. If ζ is the vector 1 and Y is the operator of multiplication by ψ, then

$$\langle \zeta, p(Y)\chi_E \zeta \rangle = \int_E p(\psi(t))\, dt = \int_E \sigma(p(Y), t)\, dt$$

for every polynomial p and Borel set E. Thus, Y realizes the distributional family σ.

The k-fold tensor product of $L^2[0, T]$ over $\mathbb{C}$ is $L^2([0, T]^k)$. When we multiply by the indicator function $\chi_{E_{\text{mono}}(k,T)}$, we obtain simply $L^2(E_{\text{mono}}(k, T))$, and thus

$$\mathcal{F}_{\text{mono}}(\mathcal{H}) = \mathbb{C}\xi \oplus \bigoplus_{k \geq 1} L^2(E_{\text{mono}}(k, T)).$$

The process with monotone independent increments given by σ is then

$$X_{s,t} = \ell_{\text{mono}}(\chi_{[s,t]}) + \ell_{\text{mono}}(\chi_{[s,t]})^* + \mathfrak{m}(Y\chi_{[s,t]}).$$

Here $\ell(\chi_{[s,t]})$ maps $L^2(E_{\text{mono}}(k, T))$ into $L^2(E_{\text{mono}}(k + 1, T))$ by $f \mapsto (\chi_{[s,t]} \otimes f)|_{E_{\text{mono}}(k+1,T)}$, and $\mathfrak{m}(Y)$ acts on $L^2(E_{\text{mono}}(k, T))$ by multiplying by $\psi\chi_{[s,t]}$ of the first coordinate. Theorem 8.5.1 says that the solution F_t of the Loewner equation is given by

$$\frac{1}{F_t(z)} = \left\langle \xi, (z - X_{0,t})^{-1}\xi \right\rangle,$$

or $F_t = F_{\mu_{0,t}}$, where $\mu_{s,t}$ is the spectral measure associated to $X_{s,t}$ and ξ.

The coupling estimate Proposition 9.3.2 leads to the following result in this setting.

Proposition 9.3.5. *Let ψ_1 and $\psi_2 \in L^\infty[0, T]$, and let $\mu_t^{(1)}$ and $\mu_t^{(2)}$ be the probability measures such that $F_{\mu_t^{(j)}}$ solves the chordal Loewner equation with driving function ψ_j. Then*

$$d_{\text{coup}}\left(\mu_t^{(1)}, \mu_t^{(2)}\right) \leq \|\psi_1 - \psi_2\|_{L^\infty[0,T]}$$

and hence the Fourier transforms satisfy

$$|\mathcal{F}\mu_t^{(1)}(t) - \mathcal{F}\mu_t^{(2)}(t)| \leq 2\pi|t|\, \|\psi_1 - \psi_2\|_{L^\infty[0,T]}.$$

Proof. Let $\sigma_j(\cdot, t) = \delta_{\psi_j(t)}$. Note that the multiplication operators by ψ_1 and ψ_2 on $L^2[0,T]$ are a coupling of σ_1 and σ_2. As explained in Proposition 9.3.2, this leads to a pair of operators

$$X_{s,t}^{(j)} = \ell_{\text{mono}}(\chi_{[s,t]}) + \ell_{\text{mono}}(\chi_{[s,t]})^* + \mathfrak{m}(\psi_j),$$

on the Hilbert space

$$\mathcal{F} = \mathbb{C}\xi \oplus \bigoplus_{k \geq 1} L^2(E_{\text{mono}}(k,T)),$$

where ψ_j is shorthand for the operator of multiplication by ψ_j on $L^2[0,T]$. Now $\mu_t^{(j)}$ is the spectral distribution of $X_{0,t}^{(j)}$ with respect to ξ, and we have $\|X_{s,t}^{(1)} - X_{s,t}^{(2)}\| \leq \|\psi_1 - \psi_2\|_{L^\infty[0,T]}$. The proves the desired estimate on $d_{\text{coup}}(\mu_t^{(1)}, \mu_t^{(2)})$, and the estimate on the Fourier transforms follows from Proposition 9.2.12. $\square$

Part II

Real evolution equations

CHAPTER 10

Introduction II

Now we begin the second part of the thesis, which deals with random matrices and the tracial W*-algebras which describe their large-n limit. It is based on [Jek19], with some of the supporting results from the the earlier paper [Jek18]. We caution that the second part of the thesis uses largely different notation from the first part (see §10.2).

10.1 Motivation

Free probability initiated a fruitful exchange between random matrix theory and operator algebras. In many situations, the large-n behavior of tuples of $n \times n$ random matrices $(X_1^{(n)}, \dots, X_d^{(n)})$ can be described by non-commutative random variables X_1, ..., X_d which are operators in a tracial W*-algebra. Conversely, many properties of non-commutative random variables (and the W*-algebras that they generate) are easier to understand when they can be simulated by finite-dimensional random matrix models. For instance, Voiculescu used free entropy, defined in terms of matricial microstates, to prove the absence of Cartan subalgebras in free group W*-algebras $L(\mathbb{F}_d)$ [Voi96]; similar techniques were used to give sufficient conditions for a von Neumann algebra to be non-prime and non-Gamma (a convenient list of results and references can be found in [CN19]). Further applications of random matrices to the properties of C*- and W*-algebras can be found for instance in [HT05] and [GS09, §4].

Free Gibbs laws are a prototypical example of the connection between random matrices and W*-algebras. Free Gibbs laws describe the large-n behavior of self-adjoint tuples of random matrices $X^{(n)} = (X_1^{(n)}, \dots, X_d^{(n)})$ given by a probability measure $\mu^{(n)}$ of the form

$$d\mu^{(n)}(x) = \frac{1}{\int e^{-n^2 V^{(n)}}} e^{-n^2 V^{(n)}(x)}\, dx,$$

where $x \in M_n(\mathbb{C})_{\text{sa}}^d$ is a self-adjoint tuple, dx denotes Lebesgue measure, $V^{(n)} : M_n(\mathbb{C})_{\text{sa}}^d \to \mathbb{R}$ is a function (known as a potential) chosen so that $e^{-n^2 V^{(n)}(x)}$ is integrable. Here $V^{(n)}(x)$ could be given by $V^{(n)}(x) = \tau_n(p(x_1, \dots, x_d))$, where $\tau_n = (1/n)\operatorname{Tr}$ and p is a non-commutative polynomial; for instance, taking

$$V^{(n)}(x) = \frac{1}{2} \sum_{j=1}^m \tau_n(x_j^2)$$

produces the Gaussian unitary ensemble (GUE). Under certain assumptions on $V^{(n)}$ (e.g. convexity and good asymptotic behavior as $n \to \infty$), there will be non-commutative random variables $X_1, \dots, X_d$ in a tracial W*-algebra $(\mathcal{M}, \tau)$ such that

$$\tau_N(p(X_1^{(n)}, \dots, X_d^{(n)})) \to \tau(p(X_1, \dots, X_d)) \text{ in probability}$$

for every non-commutative polynomial p; see [GM06, Theorems 3.3 and 3.4], [DGS16, Proposition 50 and Theorem 51], [Jek18, Theorem 4.3], and §15.1. The random matrix models satisfy the relation, derived from integration by parts, that

$$E[\tau_n(\nabla_{x_j} V^{(n)}(X^{(n)}) p(X^{(n)}))] = E[\tau_n \otimes \tau_n(\mathcal{D}_{X_j} p(X^{(n)}))],$$

where $\nabla_{x_j} V^{(n)}$ is a normalized gradient with respect to the coordinates of x_j and $\mathcal{D}_{X_j}$ denotes the free difference quotient of Voiculescu. Hence, if the $V^{(n)}$'s have a large-n limit described by a suitable "function" V (see §13 below), then the non-commutative tuple $X = (X_1, \dots, X_d)$ satisfies

$$\tau(\nabla_{x_j} V(X) p(X)) = \tau \otimes \tau(\mathcal{D}_{X_j} p(X));$$

see [GM06, §2.2 - 2.3] and §16.2. The non-commutative law of a tuple X satisfying such an equation is known as a *free Gibbs law* for V.

Given sufficient assumptions on $V^{(n)}$ (for instance, Assumption 15.1.1), many of the classical quantities associated to $X^{(n)}$ will converge in the large-n limit to their free counterparts, besides obviously the convergence of the non-commutative moments $\tau_n(p(X^{(n)}))$. For instance, the normalized classical entropy will converge to the microstates free entropy (see [Voi93, §2], [GS09, Theorem 5.1], [Jek18, §5.2], and §16.1), and the normalized classical Fisher information will converge to the free Fisher information (see §16.2). The monotone transport maps of Guionnet and Shlyakhtenko are well-approximated by classical transport maps for the random matrix models [GS14, Theorem 4.7]. The solutions of classical SDE associated to the random matrix models approximate the solutions of free SDE; see for instance [BCG03], [GS09, §2], [Dab17, §4]. For further examples of free probability describing the large-n limit of random matrix theory, see [Bia97, DHK13, Kem16, Kem17].

10.2 Notation and background on tracial W*-algebras

10.2.1 Tracial W*-algebras

Since we do not need or want to get entangled in the definitions of weak and strong operator topology, the theory of general von Neumann algebras, and so forth, we will take as our definition the following characterization of tracial W*-algebras; see [AP17b, §7].

A *tracial* W*-*algebra* is pair $(\mathcal{M}, \tau)$, where $\mathcal{M}$ is a unital C*-algebra and $\tau : \mathcal{M} \to \mathbb{C}$ is a state such that:

(1) τ is *tracial*, that is, $\tau(xy) = \tau(yx)$ for all $x, y \in \mathcal{M}$.

(2) τ is *faithful*, that is, $\tau(x^*x) = 0$ if and only if $x = 0$.

(3) The closed unit ball of $\mathcal{M}$ is complete with respect to the norm $\|x\|_2 = \tau(x^*x)^{1/2}$.

Given a tracial W*-algebra $(\mathcal{M}, \tau)$, we denote by $L^2(\mathcal{M}, \tau)$ (or $L^2(\mathcal{M})$ when τ is clear from context) the completion of $\mathcal{M}$ with respect to $\|\cdot\|_2$. This is a Hilbert space with respect to the inner product given by $\langle x, y\rangle_2 = \tau(x^*y)$ when x, $y \in \mathcal{M}$. Then $\mathcal{M}$ acts by left multiplication on $L^2(\mathcal{M}, \tau)$ because this is an example of the GNS construction (see §2.1). But using the fact that τ is tracial, we can also see that $\|xy\|_2 \leq \|x\|_2 \|y\|_\infty$, and hence there is also a right multiplication action of $\mathcal{M}$ on $L^2(\mathcal{M})$.

Because of the faithfulness of τ, $\mathcal{M}$ may be regarded as a subset of $L^2(\mathcal{M}, \tau)$. Although it is helpful when developing the general theory to use different notation for elements of $\mathcal{M}$ and elements of $L^2(\mathcal{M})$, it would be unnecessary and distracting to do so in our situation. Thus, we will be content to write for instance $x \in \mathcal{M}$ or $x \in L^2(\mathcal{M})$, or to write xy or yx for the left and right multiplication actions when $x \in \mathcal{M}$ and $y \in L^2(\mathcal{M})$. Similarly, we will not hesitate to write $\tau(x) = \langle x, 1\rangle_2$ for $x \in L^2(\mathcal{M})$. The following fact is well known (see [AP17b, §7]).

Proposition 10.2.1. *Let $(\mathcal{M}, \tau)$ be a tracial* W**-algebra. For $x \in L^2(\mathcal{M})$, let us write*

$$\|x\|_\infty = \sup\{\tau(y^*xz) : y, z \in \mathcal{M}, \|y\|_2, \|z\|_2 \leq 1\}.$$

Then $x \in \mathcal{M}$ if and only if $\|x\|_\infty < \infty$, and in that case, $\|x\|_\infty$ equals the operator norm of x.

More generally, for $z = (z_1, \dots, z_d) \in L^2(\mathcal{M})^d$, we will write

$$\|z\|_2 = \left(\sum_{j=1}^d \|z_j\|_2^2 \right)^{1/2}$$

and

$$\|z\|_\infty = \max_{j=1,\dots,d} \|z_j\|_\infty.$$

It is well known that the map $\mathcal{M} \to \mathcal{M}$ given by $x \mapsto x^*$ extends to a conjugate-linear isometry of $L^2(\mathcal{M})$. Thus, we will denote by $L^2(\mathcal{M})_{\mathrm{sa}}$ elements of $L^2(\mathcal{M})$ with $x^* = x$, and we denote by $\mathcal{M}_{\mathrm{sa}}$ the self-adjoint elements of $\mathcal{M}$. It follows from Proposition 10.2.1 that $L^2(\mathcal{M})_{\mathrm{sa}} \cap \mathcal{M} = \mathcal{M}_{\mathrm{sa}}$. Since τ is a trace, we also have $\tau(xy) \in \mathbb{R}$ whenever $x, y \in L^2(\mathcal{M})_{\mathrm{sa}}$, and hence $L^2(\mathcal{M})_{\mathrm{sa}}$ is a real Hilbert space with respect to $\langle \cdot, \cdot \rangle_2$.

By an *isomorphism of tracial* W**-algebras*, we mean a trace-preserving $*$-isomorphism. Similarly, an *inclusion or embedding of tracial* W**-algebras* is a trace-preserving $*$-homomorphism $(\mathcal{N}, \tau_{\mathcal{N}}) \to (\mathcal{M}, \tau)$ (which is necessarily injective due to the faithfulness of τ). In that case, we say that $\mathcal{N}$ is a W*-subalgebra of $(\mathcal{M}, \tau)$.

If $x = (x_1, \dots, x_d)$ is a d-tuple from $(\mathcal{M}, \tau)$, we denote by $\mathrm{W}^*(x)$ the W^*-subalgebra generated by x. We can evaluate $\mathrm{W}^*(x)$ by taking the $\|\cdot\|_2$-closure of the $*$-algebra generated by x, and then restricting to the elements that are bounded in $\|\cdot\|_\infty$.

If $\mathcal{N}$ is a W^*-subalgebra of $(\mathcal{M}, \tau)$, then we denote by $E_{\mathcal{N}}$ the orthogonal projection $L^2(\mathcal{M}) \to L^2(\mathcal{N})$. Using Proposition 10.2.1, we see that $E_{\mathcal{N}}$ maps $\mathcal{M}$ into $\mathcal{N}$ with $\|E_{\mathcal{N}}(x)\|_\infty \leq \|x\|_\infty$. Moreover, $E_{\mathcal{N}}|_{\mathcal{N}} = \mathrm{id}$ and $E_{\mathcal{N}}$ is an $\mathcal{N}$-$\mathcal{N}$-bimodule map. In other words, $E_{\mathcal{N}}$ is an $\mathcal{N}$-valued expectation in the sense of §2.5.

10.2.2 Non-commutative laws

We denote by $\mathbb{C}\langle X_1, \dots, X_d\rangle$ the $*$-algebra of non-commutative polynomials in d self-adjoint indeterminates. A *tracial non-commutative law* is a linear map $\lambda : \mathbb{C}\langle X_1, \dots, X_m\rangle \to \mathbb{C}$ satisfying

(A) *Unitality:* $\lambda(1) = 1$.

(B) *Positivity:* $\lambda(p^*p) \geq 0$ for all $p \in \mathbb{C}\langle X_1, \dots, X_d\rangle$.

(C) *Traciality:* $\lambda(pq) = \lambda(qp)$ for all $p, q \in \mathbb{C}\langle X_1, \dots, X_d\rangle$.

(D) *Exponential boundedness:* $|\lambda(x_{i(1)} \dots x_{i(k)})| \leq R^k$ for some constant R.

The set of non-commutative laws that satisfy (D) for a fixed value of R is denoted $\Sigma_{d,R}$, and it is equipped with the topology of pointwise convergence on $\mathbb{C}\langle X_1, \dots, X_d\rangle$. Likewise, the space of all laws, equipped with the topology of pointwise convergence, will be denoted by Σ_d.

Proposition 10.2.2. *If $x = (x_1, \dots, x_d)$ is a tuple of self-adjoint elements of $(\mathcal{M}, \tau)$, then the linear functional λ_X by $\lambda_x(p) = \tau(p(x))$. Conversely, every non-commutative law can be realized in this way through the GNS construction.*

Proof. First, we proceed as in Theorem 2.6.6, except that we define the power series ring in several variables instead of a single variable. We thus obtain a realization of λ by a self-adjoint tuple $(x_1, \dots, x_d)$ in a C^*-algebra $\mathcal{A}$ on a Hilbert space $L^2(\lambda)$ with a state τ given by the vector $\xi = 1$. Because λ is tracial, it follows that τ is tracial on $\mathcal{A}$. Using traciality, we check that $\mathcal{A}$ acts on $L^2(\lambda)$ by right multiplication and from this, we check that the vector ξ is separating for $\mathcal{A}$, so that τ is faithful. Then let $\mathcal{M}$ be defined as the set of $x \in L^2(\lambda)$ such that $\|x\|_\infty < \infty$. The fact that $(\mathcal{M}, \tau)$ is a tracial W^*-algebra is proved similarly to Proposition 10.2.1. □

Definition 10.2.3. Given self-adjoint d-tuples x and y, we say that $x \sim y$ *in non-commutative law* if $\lambda_x = \lambda_y$. Also, we say that $x^{(n)}$ *converges to x in non-commutative law* if $\lambda_{x^{(n)}} \to \lambda_x$. (Here x, y, and $x^{(n)}$ each come from some tracial W^*-algebra, but not necessarily the same one.)

Definition 10.2.4. Given a tracial W*-algebra $(\mathcal{M}, \tau)$ and a set $S \subseteq \mathcal{M}$, we denote by W$^*(S)$ the smallest W*-algebra of $\mathcal{M}$ containing S. This is equivalently the set of elements of $\mathcal{M}$ that are in the $\|\cdot\|_2$-closure of non-commutative polynomials in the elements of S and their adjoints.

Lemma 10.2.5. *We have $x \sim y$ if and only if there exists a tracial* W**-isomorphism α :* W$^*(x) \to$ W$^*(y)$ *with $\phi(x_j) = y_j$ for each j.*

Proof. Suppose that $\lambda_x = \lambda_y$. Then for any non-commutative polynomials p and q, we have $\tau(p(x)^* q(x)) = \tau(p(y)^* q(y))$. Therefore, there is a unitary transformation $\alpha : L^2(\mathrm{W}^*(x)) \to L^2(\mathrm{W}^*(y))$ sending $p(x)$ to $p(y)$ for every non-commutative polynomial p. Using Proposition 10.2.1, we conclude that α restricts to a $\|\cdot\|_\infty$-isometry W$^*(x) \to$ W$^*(y)$. Since α respects adjoints and multiplication of non-commutative polynomials, it follows from standard approximation arguments that α is a $*$-homomorphism everywhere.

Conversely, if α : W$^*(x) \to$ W$^*(y)$ is an isomorphism with $\alpha(x) = y$, then we have $\alpha(p(x)) = p(y)$ for any non-commutative polynomial p, and hence $\tau(p(x)) = \tau(p(y))$. ☐

10.2.3 Free independence and free products

Free independence of W*-subalgebras is defined as in §5.2.

Definition 10.2.6. If $(\mathcal{M}, \tau)$ is a tracial W*-algebra, then we say that W*-subalgebras $\mathcal{M}_1$, ..., $\mathcal{M}_N$ are *freely independent* if whenever a_1, ..., $a_k \in \mathcal{M}$ with $a_j \in \mathcal{M}_{i(j)}$ and $\tau(a_j) = 0$, we have $\tau(a_1 \dots a_k) = 0$. We say that two tuples x and y are *freely independent* if W$^*(x)$ and W$^*(y)$ are freely independent.

Given tracial W*-algebras $(\mathcal{M}_1, \tau_1)$, ..., $(\mathcal{M}_N, \tau_N)$, we construct the free product as follows. Let $(\mathcal{H}_j, \xi_j) = (L^2(\mathcal{M}_j), 1)$, and let $(\mathcal{H}, \xi)$ be the free product Hilbert space constructed in §5.3 (with $\mathcal{B} = \mathbb{C}$). Let $\rho_j : B(\mathcal{H}_j) \to B(\mathcal{H})$ be the corresponding inclusion.

Lemma 10.2.7. *The vector state τ given by ξ is tracial on the algebra $\mathcal{A}$ generated by $\rho_1(\mathcal{M}_1)$, ..., $\rho_N(\mathcal{M}_N)$.*

Proof. We must show that $\tau(xy) = \tau(yx)$ for all $x, y \in \mathcal{A}$. By linearity, it suffices to consider the case where x and y have the form a_1, ..., a_k with $a_j \in \rho_{i(j)}(\mathcal{M}_{i(j)})$. In fact, by transitivity, it suffices to show that

$$\tau(a_1 \dots a_k) = \tau(a_k a_1 \dots a_{k-1})$$

whenever $a_j \in \rho_{i(j)}(\mathcal{A}_{i(j)})$. We proceed by induction on k. We can write each a_j as a scalar multiple of the identity plus something of trace zero, so by linearity, it suffices to consider when each a_j is one of those two cases.

If some a_j is a scalar multiple of the identity, we can remove it from the string and apply the induction hypothesis. So assume that each a_j has trace zero. Furthermore, if

$i(j) = i(j+1)$ for some j with $1 \leq j \leq k-2$, then we can combine a_j and a_{j+1} into a single term and use the induction hypothesis, so assume that the consecutive indices are distinct except possibly for the last two.

If $k = 2$ and $i(1) = i(2)$, then the claim follows from traciality of $\tau_{i(1)}$. In any other case, we claim that both sides are zero. If $i(k-1) \neq i(k)$, then the left-hand side is zero by free independence. On the other hand, if $k > 2$ and $i(k-1) = i(k)$, then by free independence,

$$\tau(a_1 \dots a_k) = \tau(a_1 \dots a_{k-2}(a_{k-1}a_k - \tau(a_{k-1}a_k)) + \tau(a_1 \dots a_{k-2})\tau(a_{k-1}a_k) = 0.$$

The argument for the right-hand side is similar except that the terms $a_k a_1$ play the role of $a_{k-1}a_k$. □

Since τ is tracial on $\mathcal{A}$, $(\mathcal{A}, \tau)$ generates a tracial W*-algebra $(\mathcal{M}, \tau)$. We define the *free product* of $(\mathcal{M}_1, \tau_1)$, ..., $(\mathcal{M}_N, \tau_N)$ to be $(\mathcal{M}, \tau)$. Note that ρ_j defines a tracial W*-inclusion $\mathcal{M}_j \to \mathcal{M}$.

10.2.4 Notation for matrix algebras

We denote by τ_n the normalized trace $(1/n)\operatorname{Tr}$ on $M_n(\mathbb{C})$. Then $(M_n(\mathbb{C}), \tau_n)$ is a tracial W*-algebra. In particular, we use the notation $\|x\|_2$, $\|x\|_\infty$, and λ_x as defined above when x is an d-tuple of matrices. The notation $\|\cdot\|_2$ and $\|\cdot\|_\infty$ will never be used for the L^2 or L^∞ norms of *functions* on matrices, but if we write an L^p norm it will be denoted by $\|\cdot\|_{L^p}$.

Now $M_n(\mathbb{C})_{\text{sa}}^d$ is a real inner-product space with respect to $\langle \cdot, \cdot \rangle_2$. It has (real) dimension dn^2, and hence can be identified isometrically with $\mathbb{R}^{dn^2}$. For a smooth function $u : M_n(\mathbb{C})_{\text{sa}}^d \to \mathbb{R}$, we denote by ∇u and Hu the gradient and Hessian with respect to such an isometric identification. (See §14.1 for further explanation.) Note that this inner product $\langle \cdot, \cdot \rangle_2$ differs by a factor of $1/n$ from the inner product $\langle x, y \rangle_{\operatorname{Tr}} = \operatorname{Tr}(x^* y)$ which many authors use in random matrix theory. Moreover, since diagonal entries of a self-adjoint matrix are real, and the off-diagonal entries are in complex-conjugate pairs, we are *not* using entrywise coordinates to identify $M_n(\mathbb{C})_{\text{sa}}^d$ with $\mathbb{R}^{dn^2}$ or compute derivatives (even after taking account of the renormalization from Tr to τ_n).

10.3 Summary of main results

The second part of the thesis will further develop the connection between classical and free probability for convex free Gibbs laws by studying conditional expectation (§15), conditional entropy and Fisher information (§16), and conditional transport (§17).

We consider a sequence of random matrix tuples $(X^{(n)}, Y^{(n)}) = (X_1^{(n)}, \dots, X_{d_1}^{(n)}, Y_1^{(n)}, \dots, Y_{d_2}^{(n)})$ given by a uniformly convex and semi-concave sequence of potentials $V^{(n)}$ such that the normalized gradient $\nabla V^{(n)}$ is asymptotically approximable by trace polynomials (a notion of good asymptotic behavior as $n \to \infty$ defined in §13.5). Then the following results hold:

(1) The non-commutative moments $\tau_n(p(X^{(n)}, Y^{(n)}))$ converge almost surely to $\tau(p(X,Y))$ for some tuple (X,Y) of non-commutative random variables in a tracial W*-algebra. See Theorem 15.1.5.

(2) The classical conditional expectation $E[f^{(n)}(X^{(n)}, Y^{(n)})|Y^{(n)}]$ behaves asymptotically like the non-commutative conditional expectation $E_{\mathrm{W}^*(Y)}[f(X,Y)]$ where f comes from an appropriate non-commutative function space and $f^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to M_n(\mathbb{C})$ is a sequence of functions that satisfy some mild growth bounds at ∞ and "behave like f in the large-n limit" in the sense of §13.5. See Theorem 15.1.7.

(3) The classical conditional entropy $n^{-2}h(X^{(n)}|Y^{(n)}) + d_1 \log n$ converges to the conditional free entropy $\chi^*(X : \mathrm{W}^*(Y))$. This is a similar to a conditional version of $\chi = \chi^*$. See Theorem 16.4.1.

(4) There exists a function $f(X,Y)$ such that $(f(X,Y),Y) \sim (Z,Y)$ in non-commutative law, where Z is a free semicircular m-tuple freely independent of Y, and this function also arises from functions $f^{(n)}$ such that $(f^{(n)}(X^{(n)}, Y^{(n)}), Y^{(n)}) \sim (Z^{(n)}, Y^{(n)})$, where $Z^{(n)}$ is an independent Gaussian d-tuple. This is the conditional version of transport to the Gaussian/semicircular law. See Theorems 17.1.1 and 17.1.4.

(5) This transport map also witnesses the conditional entropy-cost inequality for the law of X relative to semicircular conditioned on Y.

(6) This transport map furnishes an isomorphism $\mathrm{W}^*(X,Y) \cong \mathrm{W}^*(S,Y) \cong \mathrm{W}^*(S) * \mathrm{W}^*(Y)$, which shows that $\mathrm{W}^*(Y)$ is freely complemented in $\mathrm{W}^*(X,Y)$.

(7) Actually, a second application of transport shows that $\mathrm{W}^*(Y)$ is isomorphic to the W*-algebra generated by a semicircular n-tuple, or in other words $L(\mathbb{F}_{d_1})$. So altogether there is an isomorphism $\mathrm{W}^*(X,Y) \to L(\mathbb{F}_{d_1+d_2})$ that maps $\mathrm{W}^*(Y)$ to the canonical copy of $L(\mathbb{F}_{d_2})$ inside $L(\mathbb{F}_{d_1+d_2})$.

Furthermore, the results about transport can be iterated to produce a "lower-triangular transport" as shown in Theorems 17.1.8 and 17.1.9 and discussed further in §10.6. This is analogous to the classical results on triangular transport of measure such as [BKM05].

In the rest of the introduction, we will motivate and explain the main results in more detail. It will become clear in the course of the argument that our main results and their proofs are tightly interrelated.

10.4 Main results on conditional expectation

Consider a tuple

$$(X^{(n)}, Y^{(n)}) = (X_1^{(n)}, \dots, X_{d_1}^{(n)}, Y_1^{(n)}, \dots, Y_{d_2}^{(n)})$$

of random self-adjoint matrices given by a probability density $(1/\int e^{-n^2 V^{(n)}}) e^{-n^2 V^{(n)}(x,y)}\,dx\,dy$. We assume that $V^{(n)}$ is uniformly convex and semi-concave and that the normalized gradient $\nabla V^{(n)}$ is asymptotically approximable by trace polynomials (a certain notion of good asymptotic behavior as $n \to \infty$, explained below). The precise hypotheses are listed in Assumption 15.1.1. We show in Theorem 15.1.5 that in this case, there exists an $(d_1 + d_2)$-tuple (X, Y) of non-commutative random variables such that $\tau_n(p(X, Y)) \to \tau(p(X, Y))$ almost surely.

Our first main result (Theorem 15.1.2) says roughly that the classical conditional expectation given $Y^{(n)}$ well approximates the W*-algebraic conditional expectation $E_{\mathrm{W}^*(Y)} : \mathrm{W}^*(X, Y) \to \mathrm{W}^*(Y)$. This is motivated in general by the importance of conditional expectation in free probability, e.g. its relationship to free independence with amalgamation and to free score functions. See [BCG03, §4] for a study of the large n limits of conditional expectations related to matrix SDE. The relationship between classical and free conditional expectation also has implications for the study of relative matricial microstate spaces, such as the "external averaging property" introduced in the upcoming joint work with Hayes, Nelson, and Sinclair [HJN19].

Applications of conditional expectation within this paper include our results on free Fisher information and entropy (see Theorem 16.4.1), as well as our proof that Assumption 15.1.1 is preserved under marginals (see Proposition 15.1.8).

The statement and proof of Theorem 15.1.2 rely on a notion of asymptotic approximation for functions on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ explained in §13.5. We define a class of non-commutative functions $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ as a certain Fréchet space completion of trace polynomials, such that if $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ and $X_1, \dots, X_d$ are self-adjoint elements in a Connes-embeddable tracial W*-algebra $(\mathcal{M}, \tau)$, then $f(X_1, \dots, X_d)$ is a well-defined element of $L^2(\mathcal{M})$. In particular, every $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ can be evaluated on a tuple of self-adjoint matrices. Now if $f^{(n)} : M_N(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})$, we say that $f^{(n)} \rightsquigarrow f$ if for every $R > 0$,

$$\lim_{n\to\infty} \sup_{\substack{x \in M_n(\mathbb{C})_{\mathrm{sa}}^m \\ \|x\|_\infty \le R}} \left\| f^{(n)}(x) - f(x) \right\|_2 = 0,$$

Moreover, if such an f exists, then we say that $f^{(n)}$ is *asymptotically approximable by trace polynomials.*

Consider the random matrices $(X^{(n)}, Y^{(n)})$ and non-commutative random variables (X, Y) as above, and suppose that $f^{(n)} : M_N(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to M_n(\mathbb{C})$ satisfies $f^{(n)} \rightsquigarrow f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*(d_1+d_2)}, \|\cdot\|_2)$ and satisfies some reasonable growth bounds at ∞. Then we show that $E[f^{(n)}(X^{(n)}, Y^{(n)}) | Y^{(n)}]$ is given by a function $g^{(n)}(Y^{(n)})$ such that $g^{(n)} \rightsquigarrow g \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d_2}, \|\cdot\|_2)$, and moreover $E_{\mathrm{W}^*(Y)}[f(X, Y)] = g(Y)$.

A curious feature of this result is that the function g is defined for all self-adjoint d_2-tuples of non-commutative random variables, not only for the specific d_2-tuple Y that we are concerned with. Similarly, the claim that $g^{(n)} \rightsquigarrow g$ describes the asymptotic behavior of $g^{(n)}(y)$ for all $y \in M_n(\mathbb{C})_{\mathrm{sa}}^{d_2}$, even though the distribution of the random matrix $Y^{(N)}$ is highly concentrated as $N \to \infty$ on much smaller sets, namely the "matricial microstate spaces" consisting of tuples $y \in M_n(\mathbb{C})_{\mathrm{sa}}^{d_2}$ with non-commutative moments close to those of Y (see §16.1). Thus, the statement we prove about the functions $g^{(n)}$ is stronger than an asymptotic result about L^2 approximation such as [GS14, Theorem 4.7].

10.5 Main results on entropy

Voiculescu defined two types of free entropy (see [Voi94], [Voi98a], [Voi02b]). The first, called $\chi(X)$, measures the asymptotic volume of matricial microstate spaces, which is closely related to the classical entropy of the random matrix models (see §16.1). The second, called $\chi^*(X)$, is defined in terms of free Fisher information, which has to do with how the distribution of X interacts with differentiation and is analogous to the classical Fisher information (see §16.2). Either one should heuristically be the large-n limit of the classical entropy of random matrix models, but there were many technical obstacles to proving this (see [Voi02b]). The inequality $\chi \leq \chi^*$ is known in general thanks to [BCG03]. However, even for non-commutative laws as well-behaved and explicit as free Gibbs laws given by uniformly convex and semi-concave potentials, the equality of χ and χ^* when $d > 1$ was not proved until Dabrowski's paper [Dab17], and the problem is still open for non-convex Gibbs laws.

In [Jek18], we gave a proof of this equality in the uniformly convex and semi-concave case based on the asymptotic analysis of functions and PDE related to the random matrix models which had some parallels to the SDE techniques of [Dab17]. Here, as in [Jek19], we will use similar techniques for the conditional setting. We will show (Theorem 16.4.1) that for a random tuple of matrices $(X^{(n)}, Y^{(n)})$ given by a sequence of convex potentials as above, the classical conditional entropy $n^{-2} h(X^{(n)} | Y^{(n)}) + d_1 \log n$ converges to the conditional free entropy $\chi^*(X : \mathrm{W}^*(Y))$.

In particular, in the non-conditional setting when $d_2 = 0$ (so there is no Y), this argument implies that $\chi(X) = \chi^*(X)$ (see Corollary 16.4.2), because $\chi(X)$ is evaluated as the $\limsup$ of normalized classical entropies (Proposition 16.1.4). In fact, the proof here is shorter than those of [Dab17] and [Jek18].

In the conditional setting, we focus only on the non-microstates entropy (defined using Fisher information). It is not yet resolved in the literature what the correct definition of conditional microstates free entropy should be. In light of §16.1, the conditional classical entropy for the random matrix models seems to be a reasonable substitute for microstates entropy. In the setting where $V^{(n)}$ is uniformly convex, both the overall distribution and the conditional distribution of $X^{(n)}$ given $Y^{(n)}$ exhibit concentration of measure, and hence we expect the large-n limit of classical conditional entropy to agree with any plausible definition of conditional microstates entropy.

10.6 Main results on transport

A *transport map* from a probability measure μ and to another probability measure ν is a function f such that $f_*\mu = \nu$. In probabilistic language, if $X \sim \mu$ and $Y \sim \nu$ are random variables, then $f_*\mu = \nu$ means that $f(X) \sim Y$ in distribution. The theory of transport (and in particular optimal transport) has numerous and significant applications in the classical setting. For instance, if we have a function f such that $f(X) \sim Y$ and we can numerically simulate the random variable X, then we can also simulate Y.

In the non-commutative world, transport is even more significant. As remarked in [GS14, §1.1], there is no known analogue of a probability density in free probability. However, the existence of transport maps that would express our given random variables as functions of a free semicircular family (for instance) would serve a similar purpose to a density, namely to provide a fairly explicit and analytically tractable model for a large class of non-commutative laws.

Moreover, in contrast to the classical setting, the very existence of transport maps is a nontrivial condition. Being able to express a non-commutative tuple Y as a function of another non-commutative tuple X implies that $\mathrm{W}^*(Y)$ embeds into $\mathrm{W}^*(X)$, and if in addition this transport map is invertible, then $\mathrm{W}^*(Y) \cong \mathrm{W}^*(X)$. In the classical setting, any two diffuse (non-atomic) standard Borel probability spaces are isomorphic. On the other hand, there are many non-isomorphic diffuse tracial W^*-algebras, even after restricting our attention to factors (those which cannot be decomposed as direct sums, or equivalently have trivial center); see [McD69]. Moreover, Ozawa [Oza04] showed that there is no separable tracial factor that contains an isomorphic copy of each of the others. Thus, there are many instances where it is not even possible to transport one given non-commutative law to another.

The papers [GS09] and [DGS16] showed the existence of monotone transport maps between certain free Gibbs laws given by convex potentials and the law of a free semicircular family, and thus concluded that each of the corresponding W^*-algebras was isomorphic to a free group factor $L(\mathbb{F}_d)$. In particular, this result applies to the q-Gaussian variables for sufficiently small q. These transport techniques have been extended to type III von Neumann algebras [Nel15a], to planar algebras [Nel15b], and to interpolated free group factors [HN18].

We will focus on "conditional transport" in the tracial setting. Our first main result about transport is contained in Theorems 17.1.1 and 17.1.4. Let $(X^{(n)}, Y^{(n)})$ be an (d_1+d_2)-tuple of random matrices arising from a sequence of convex potentials satisfying Assumption 15.1.1. Let (X, Y) be an $(d_1 + d_2)$-tuple of non-commutative self-adjoint variables realizing the limiting free Gibbs law. Then we construct functions $F^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to M_N(\mathbb{C})_{\mathrm{sa}}^{d_1}$ such that $(F^{(n)}(X^{(n)}, Y^{(n)}), Y^{(n)}) \sim (Z^{(n)}, Y^{(n)})$ in distribution, where $Z^{(n)}$ is a GUE d_1-tuple independent of $Y^{(n)}$. We think of this as a conditional transport, which transports the law of $X^{(n)}$ to the law of $Z^{(n)}$ *conditioned on* $Y^{(n)}$.

Moreover, we show that the transport maps satisfy $F^{(n)} \rightsquigarrow F \in C_{\mathrm{tr,app}}(\mathbb{R}^{*(d_1+d_2)}, \|\cdot\|_2)_{\mathrm{sa}}^{d_1}$. In the large-$n$ limit, we obtain $(F(X,Y), Y) \sim (Z, Y)$ in non-commutative law, where Z is a free semicircular d_1-tuple freely independent of Y. In particular, this means that $\mathrm{W}^*(X,Y) \cong \mathrm{W}^*(Z,Y) = \mathrm{W}^*(Z) * \mathrm{W}^*(Y)$ (where $*$ denotes free product). In other words, $\mathrm{W}^*(Y)$ is freely complemented in $\mathrm{W}^*(X,Y)$.

By iterating this result, we can show that if $X = (X_1, \dots, X_d)$ is a tuple of non-commutative random variables given by a convex free Gibbs law as above, then there is an isomorphism $\mathrm{W}^*(X) \to \mathrm{W}^*(Z)$ such that $\mathrm{W}^*(X_1, \dots, X_k)$ is mapped onto $\mathrm{W}^*(Z_1, \dots, Z_k)$ for each $k = 1, \dots, d$. In other words, there is a "lower-triangular transport." See Theorems 17.1.8 and 17.1.9. This is a (partial) free analogue of the classical result [BKM05, Corollary 3.10].

This result implies in particular that $\mathrm{W}^*(X_1)$ is a maximal abelian subalgebra and in fact maximal amenable (since the subalgebra $\mathrm{W}^*(S_1)$ is known to be maximal amenable thanks to Popa [Pop83]), and the same holds for each $\mathrm{W}^*(X_j)$ by symmetry. For context on maximal amenable subalgebras, see for instance [Pop83, BC15, BH18]. More generally, *any von Neumann algebraic property* of the sequence of inclusions $\mathrm{W}^*(X_1) \subseteq \mathrm{W}^*(X_1, X_2) \subseteq \dots \subseteq \mathrm{W}^*(X_1, \dots, X_d)$ will behave the same way as in the case of free semicirculars, that is, for the standard inclusions $L(\mathbb{Z}) \subseteq L(\mathbb{F}_2) \subseteq \dots \subseteq L(\mathbb{F}_d)$.

We will show in Theorem 18.3.1 that if Z is a free semicircular family and p is a polynomial, then for δ sufficiently small, the non-commutative d-tuple $X = (Z_1 + \delta p_1(Z), \dots, Z_d + \delta p_d(Z))$ is given by a free Gibbs law coming from random matrix models satisfying our assumptions, and hence X can be triangularly transported back to Z, and hence the sequence of tracial W^* inclusions coming from X is isomorphic to that coming from Z. This produces an abundance of new examples of non-commutative random variables X which behave the same as a free semicircular family from the von Neumann algebraic standpoint.

The triangular transport maps also relate naturally to the free Talagrand inequality. Let X be a non-commutative d-tuple realizing the limiting free Gibbs law for random matrix models as above, and denote by F the triangular transport map from the law of X to the law of a free semicircular family Z in our construction, so that $F(X) \sim Z$. Then F witnesses the Talagrand entropy-cost inequality relative to the semicircular law, that is,

$$\|F(X) - X\|_2^2 \leq \|X\|_2^2 + d \log 2\pi - 2\chi^*(X),$$

where the left-hand side is twice the entropy relative to semicircular (see Theorem 17.1.9).

This is not surprising because it was already known in the classical case that the Talagrand inequality can be witnessed by some triangular transport [BKM05, Corollary 3.10]. Moreover, our construction of the transport maps is a direct application of the same method that Otto and Villani used to prove the Talagrand entropy-cost inequality under the assumption of the log-Sobolev inequality [OV00, §4]; see also §12.3. Thus, our main new contribution is to study the large-n limit of the transport maps using asymptotic approximation by trace polynomials and some explicit dimension-independent Lipschitz estimates. Specifically, we show that F is $\|\cdot\|_2$-Lipschitz, and we estimate $\|F(X) - X\|_\infty$ in terms of the constants c and C specifying the uniform convexity and semi-concavity of $V^{(n)}$. These estimates will go to zero as $c, C \to 1$.

Unfortunately, the maps constructed here are not optimal triangular transport maps with respect to the L^2-Wasserstein distance, since Otto and Villani's proof of [OV00, Theorem 1] uses a diffusion-semigroup interpolation between the two measures, not the displacement interpolation from optimal transport theory. In that sense, the results of this paper do not fully prove an analogue of [BKM05, Corollary 3.10]. Even in the work of Guionnet and Shlyakhtenko [GS09], which constructed monotone transport maps in the free setting, the question of whether these maps furnish an optimal coupling between X and S inside a tracial von Neumann algebra was left unresolved. Future research should study optimal transport in the free setting, and determine whether the classical optimal transport (or more generally optimal triangular transport) maps for the random matrix models converge in the large-n limit in the sense of this paper (or perhaps only in some weaker sense).

10.7 Overview of the second part

The second part of the thesis is organized as follows.

In §11, we review some relatively standard background concerning log-concave measures on $\mathbb{R}^d$, particularly those given as e^{-V} for some uniformly convex function V. After some elementary estimates for semi-convex and semi-concave functions V and the associated measures, we review the log-Sobolev inequality, Herbst's concentration inequality, and their consequences for random matrix models.

In §12, we review background on classical entropy and Fisher information. We include an exposition of Otto and Villani's proof that the log-Sobolev inequality implies the Talagrand inequality [OV00]. This chapter serves as motivation and technical support for our results about non-commutative entropy and transport in §16 and §17 respectively.

In §13, we define a space $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ of "functions of d non-commuting real variables from a tracial W*-algebra." Functions in $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ are suitable for describing the large-n limit of certain sequences of functions $f^{(n)}$ on $M_n(\mathbb{C})_{\text{sa}}^d$. This is the technical framework that will be used to state and prove most of our main results. The role of §13 in the second part of the thesis is loosely analogous to the role of §3 in the first part; §3 explores "non-commutative complex analysis" while §13 explores "non-commutative real analysis."

Because the space $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ is not yet standard in random matrix theory, we describe two other characterizations of this space to build motivation and intuition — first, we show that it consists of continuous sections of a certain vector bundle over the space of non-commutative laws (§13.6), and second, we show that it consists of quantifier-free definable functions in d variables in the sense of model theory (§13.7). These characterizations are not used in the proof of the main results, but are included because of their inherent interest.

In §14, we develop further technical tools to prove our main results. First, we explain how to compute the gradient and Laplacian of trace polynomials. This leads into an analysis of the standard heat semigroup (almost known as convolution with the Gaussian measure) for functions on $M_n(\mathbb{C})_{\text{sa}}^d$ in the large-n limit. Finally, we study ODE for functions $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$

and for sequences of functions on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ are asymptotic to some function $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ in the large-n limit.

In §15, we prove our first main result (Theorem 15.1.2). As consequences, we show that the random matrix models associated to a sequence of potentials $V^{(n)}$ satisfying Assumption 15.1.1 converge in the large-n limit to some non-commutative law (Theorem 15.1.5). Furthermore, if $(X^{(n)}, Y^{(n)})$ is a random matrix tuple given by such a potential, we show in Theorem 15.1.7 that (roughly speaking) the large-n limit of classical conditional expectations given $Y^{(n)}$ behaves in the large-n limit like the W*-algebraic conditional expectation.

In §16, we study the microstates free entropy, free Fisher information, and non-microstates free entropy. We state sufficient conditions for the microstates free entropy $\chi(X)$ to be given as the large-n limit of classical entropies of $n \times n$ random matrix models (Proposition 16.1.4), and for the free Fisher information to given as the large-n limit of classical Fisher information (Proposition 16.2.4). We conclude with the main result that $\chi(X) = \chi^*(X)$ whenever X arises as the large-n limit of matrix models as in Assumption 15.1.1. We also show that the large-n limit of classical conditional entropy for such matrix models is described by the conditional version of χ^*.

In §17, we apply Otto and Villani's transport construction to the random matrix models satisfying Assumption 15.1.1. The result is collection of maps $F^{(n)}$ transporting $\mu^{(n)}$ to the Gaussian measure, which behave well in the large-n limit (Theorem 17.1.1) and thus produce an isomorphism of tracial W*-algebras (Theorem 17.1.4). By iterating our conditional transport results, we obtain a lower-triangular transport map (Theorems 17.1.8 and 17.1.9).

In §18, we discuss two types of examples of potentials $V^{(n)}$ to which our results in §15 - §17 apply, both of a perturbative nature. First, we show that a potential $V^{(n)}$ which is a small perturbation of the potential $(1/2)\|x\|_2^2$ on an operator norm ball can be extended to potential satisfying Assumption 15.1.1, allowing us to apply our previous results. Next, we show that if Z is a free semicircular family, and f is a tuple of self-adjoint non-commutative polynomials, then $Z + \delta f(Z)$ has random matrix models satisfying Assumption 15.1.1 for sufficiently small δ; hence in particular, it can be triangularly transported to another free semicircular family.

CHAPTER 11

Background: Log-concave measures

The random matrix distributions studied in this work are log-concave measures on the finite-dimensional inner product space $M_n(\mathbb{C})_{\text{sa}}^d$, which is isometrically isomorphic to $\mathbb{R}^{dn^2}$. This chapter will review some background on convex functions and log-concave measures on $\mathbb{R}^d$, including the log-Sobolev and Braskamp-Lieb inequalities. The goal is to aid readers who are interested in the von Neumann algebraic results of this paper, but less familiar with the tools from classical analysis. The results that we state for $\mathbb{R}^d$ will clearly extend to any finite-dimensional inner product space, and we will explicitly state their application to $M_n(\mathbb{C})_{\text{sa}}^d$ with free probabilistic normalizations in §11.5.

11.1 Semi-convex and semi-concave Functions

Although we restrict our attention to $\mathbb{R}^d$ for the moment, we use the notation $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ for the dot product and norm because we are thinking about the application of these results to arbitrary finite-dimensional real inner product spaces.

Definition 11.1.1. Let $A : \mathbb{R}^d \to \mathbb{R}^d$ be a self-adjoint linear transformation, and let $u : \mathbb{R}^d \to \mathbb{R}$ be a function. We say that $Hu \geq A$ if $u(x) - \frac{1}{2}\langle Ax, x \rangle$ is a convex function, and we say that $Hu \leq A$ if $u(x) - \frac{1}{2}\langle Ax, x \rangle$ is a concave function.

Definition 11.1.2. We say that u is *semi-convex* if $Hu \geq cI$ for some $c \in \mathbb{R}$ and *semi-concave* if $Hu \leq CI$ for some $C \in \mathbb{R}$.

Lemma 11.1.3. *Let $\mathcal{E}(A, B) = \{u : A \leq Hu \leq B\}$. Then u is closed under translation, averaging with respect to probability measures, and pointwise limits. Hence, in particular, if μ is a probability measure, $u \in \mathcal{E}(A, B)$ and $\mu * u(x) := \int u(x + y)\, d\mu(y)$ is finite for all x, then $\mu * u \in \mathcal{E}(A, B)$.*

Proof. First, one can show that convex functions are closed under all these operations. Indeed, convexity is given by the conditions

$$u\left(\sum_{j=1}^{n} t_j x_j\right) \leq \sum_{j=1}^{n} t_j u(x_j),$$

whenever $t_j \geq 0$ and $\sum_{j=1}^n t_j = 1$. This is a family of linear inequalities on the values of u pointwise, hence is preserved by pointwise limits and averaging. The family of inequalities is also translation invariant. The convolution property follows from translation and averaging.

Next, to extend this characterization to $\mathcal{E}(A, B)$, we first observe that

$$\langle A(x - x_0), x - x_0 \rangle = \langle Ax, x \rangle - 2\langle Ax, x_0 \rangle + \langle Ax_0, x_0 \rangle.$$

Since $x \mapsto \langle A(x - x_0), x - x_0 \rangle$ and $x \mapsto \langle Ax, x \rangle$ differ by an affine function, convexity of $u(x) - (1/2)\langle Ax, x \rangle$ is equivalent to convexity of $u(x) - (1/2)\langle A(x - x_0), x - x_0 \rangle$. With this computation in mind, it is clear that functions with $Hu \geq A$ are closed under all the asserted operations, and the holds for $Hu \leq B$ by symmetrical reasoning. □

Lemma 11.1.4. *Let $u : \mathbb{R}^d \to \mathbb{R}$. The following are equivalent:*

(1) $A \leq Hu \leq B$.

(2) For each $x_0 \in \mathbb{R}^d$, there exists $y_0 \in \mathbb{R}^d$ such that

$$\frac{1}{2}\langle A(x - x_0), x - x_0 \rangle \leq u(x) - u(x_0) - \langle y_0, x - x_0 \rangle \leq \frac{1}{2}\langle B(x - x_0), x - x_0 \rangle$$

for all $x \in \mathbb{R}^d$.

(3) u is continuously differentiable and

$$\langle A(x - x_0), x - x_0 \rangle_2 \leq \langle \nabla u(x) - \nabla u(x_0), x - x_0 \rangle_2 \leq \langle B(x - x_0), x - x_0 \rangle_2$$

for all $x, x_0 \in \mathbb{R}^d$, where $\nabla u(x)$ is the gradient of u.

Moreover, in this case, we have $\|\nabla u(x) - \nabla u(x_0)\| \leq \max(\|A\|, \|B\|)\|x - x_0\|$.

Proof. It is well known that u is convex if and only if for each x_0, there exists y_0 such that

$$u(x) - u(x_0) \geq \langle y_0, x - x_0 \rangle \text{ for all } x.$$

To show (1) $\implies$ (2), we apply this characterization to $u(x) - (1/2)\langle A(x - x_0), x - x_0 \rangle$ at the point x_0 to find a vector y_0 with

$$\frac{1}{2}\langle A(x - x_0), x - x_0 \rangle \leq u(x) - u(x_0) - \langle y_0, x - x_0 \rangle.$$

Similarly, there is a vector y_1 with

$$u(x) - u(x_0) - \langle y_1, x - x_0 \rangle \leq \frac{1}{2}\langle A(x - x_0), x - x_0 \rangle.$$

Subtracting these inequalities shows that

$$\langle y_1 - y_0, x - x_0 \rangle \leq \frac{1}{2}\langle x - x_0, (A - B)(x - x_0) \rangle \leq 0.$$

Substituting $x = x_0 + (y_1 - y_0)$ shows that $y_1 = y_0$, and hence (2) holds.

Conversely, if (2) holds, then the characterization of convvex function by supporting hyperplanes, in the reverse direction, implies that $A \le Hu \le B$.

(3) $\implies$ (2). Since u is continuously differentiable,

$$u(x) - u(x_0) = \int_0^1 \langle \nabla u(tx + (1-t)x_0), x - x_0 \rangle \, dt,$$

hence

$$\begin{aligned} u(x) - u(x_0) - \langle \nabla u(x_0), x - x_0 \rangle &= \int_0^1 \langle \nabla u(tx + (1-t)x_0) - \nabla u(x_0), x - x_0 \rangle \, dt \\ &\le \int_0^1 \langle B[(tx + (1-t)x_0) - x_0], x - x_0 \rangle \, dt \\ &= \int_0^1 t \langle B(x - x_0), x - x_0 \rangle \, dt \\ &= \frac{1}{2} \langle B(x - x_0), x - x_0 \rangle, \end{aligned}$$

which proves the upper bound of (2), and the lower bound is symmetrical.

(2) $\implies$ (3). Let ρ_k be a smooth probability density supported in $B(0, 1/k)$. Now (2) clearly implies continuity of u. Hence, have $u_k \to u$ pointwise as $k \to \infty$, and u_k is smooth by basic facts about convolution. Because of the previous lemma, u_k satisfies (1) and hence (2). Being smooth, u_k has the Taylor expansion

$$u_k(x) = u_k(x_0) + \langle \nabla u_k(x_0), x - x_0 \rangle + \frac{1}{2} \langle x - x_0, Hu(x_0)(x - x_0) \rangle + o(\|x - x_0\|^2),$$

where ∇u_k is the gradient and Hu_k is the Hessian of u_k (which we view as a linear transformation). Comparing this with the expansion (2) at the point x_0, we can deduce that $y_0 = \nabla u_k(x_0)$ and $A \le Hu_k(x_0) \le B$. This holds for every x_0, and hence the inequality $A \le Hu_k \le B$ is true in the pointwise sense (not merely in the sense of Definition 11.1.1).

Recall that Hu_k is the differential of ∇u_k, and so

$$\nabla u_k(x) - \nabla u_k(x_0) = \int_0^1 Hu_k(tx + (1-t)x_0)(x - x_0) \, dt.$$

Now $A \le Hu_k(tx + (1-t)x_0) \le B$ implies that $\|Hu_k(tx + (1-t)x_0)\| \le \max(\|A\|, \|B\|)$, and hence $\|\nabla u_k(x) - \nabla u_k(x_0)\| \le \max(\|A\|, \|B\|)\|x - x_0\|$. Moreover, since $A \le Hu_k \le B$ pointwise,

$$\begin{aligned} \langle \nabla u_k(x) - \nabla u_k(x_0), x - x_0 \rangle &\le \int_0^1 \langle Hu_k(tx + (1-t)x_0)(x - x_0), x - x_0 \rangle \, dt \\ &\le \int_0^1 \langle B(x - x_0), x - x_0 \rangle \\ &= \langle B(x - x_0), x - x_0 \rangle, \end{aligned}$$

and the analogous lower bound holds with A instead of B.

So (3) holds for u_k; to show that it holds for u, we want to apply the Arzelà-Ascoli theorem to $(\nabla u_k)_{k\in\mathbb{N}}$. Clearly, $(\nabla u_k)_{k\in\mathbb{N}}$ is equicontinuous because ∇u_k is $\max(\|A\|, \|B\|)$-Lipschitz. To get pointwise boundedness, we substitute $x = x_0 + y$ in (2) for some vector y and obtain

$$|\langle \nabla u_k(x_0), y\rangle| \leq |u_k(x_0 + y) - u_k(x_0)| + \max(\|A\|, \|B\|)\|y\|^2,$$

and the right hand side is clearly bounded as $k \to \infty$. So by Arzelà-Ascoli, we may assume without loss of generality that ∇u_k converges as $k \to \infty$ locally uniformly to some continuous function f. It is easy to check that f must be the gradient of u and hence u is continuously differentiable and $\nabla u_k \to \nabla u$. Since (3) holds for u_k, it must also hold for u. □

Lemma 11.1.5. *Suppose that $0 \leq Hu \leq A$ for some self-adjoint linear transformation A. Then u is differentiable and we have*

$$|\langle \nabla u(x) - \nabla u(x_0), y\rangle| \leq \langle A(x - x_0), x - x_0\rangle^{1/2}\langle Ay, Ay\rangle^{1/2},$$

so that in particular, $\|\nabla u(x) - \nabla u(x_0)\| \leq \|A\|\|x - x_0\|$.

Proof. Differentiability follows from the previous lemma. Also, as in the proof of (3) $\implies$ (2) of the previous lemma, it suffices to prove the claim in the case where u is smooth. But in this case

$$\begin{aligned}
&\langle \nabla u(x) - \nabla u(x_0), y\rangle \\
&\quad = \int_0^1 \langle Hu(tx + (1-t)x_0)(x - x_0), y\rangle \, dt \\
&\quad \leq \int_0^1 \langle Hu(tx + (1-t)x_0)(x - x_0), x - x_0\rangle^{1/2}\langle Hu(tx + (1-t)x_0)y, y\rangle^{1/2} \, dt \\
&\quad \leq \int_0^1 \langle A(x - x_0), x - x_0\rangle^{1/2}\langle Ay, y\rangle^{1/2} \, dt \\
&\quad = \langle A(x - x_0), x - x_0\rangle^{1/2}\langle Ay, y\rangle^{1/2}.
\end{aligned}$$

□

Lemma 11.1.6. *Suppose that $u : \mathbb{R}^d \to \mathbb{R}$ with $Hu \geq cI$ for some $c > 0$. Then u achieves a global minimum at a unique point x_0, and moreover,*

$$u(x) - u(x_0) \geq \frac{c}{2}\|x - x_0\|^2.$$

Proof. Since $Hu \geq cI$, there exists a vector y_0 such that

$$u(x) - u(0) \geq \langle y_0, x\rangle + \frac{c}{2}\|x\|^2.$$

The function on the right-hand side is bounded below, hence so is u. Let $\delta = \inf u(x)$, and choose a sequence x_k such that $u(x_k) \to \delta$ as $k \to \infty$. By convexity of $u(x) - (c/2)\|x\|^2$, we have

$$u\left(\frac{x_j + x_k}{2}\right) - \frac{c}{2}\left\|\frac{x_j + x_k}{2}\right\|^2 \leq \frac{1}{2}\left(u(x_j) - \frac{c}{2}\|x_j\|^2 + u(x_k) - \frac{c}{2}\|x_k\|^2\right).$$

We substitute $u((x_j + x_k)/2) \geq \delta$ and rearrange using the parallelogram identity to obtain

$$\frac{c}{4}\|x_j - x_k\|^2 \leq (u(x_j) - \delta) + (u(x_k) - \delta),$$

which implies that $(x_k)_{k\in\mathbb{N}}$ is Cauchy and hence converges to some x_0. Then $u(x_0) = \delta$, so x_0 is a global minimizer.

Next, we apply convexity of the function $u(x) - (c/2)\|x - x_0\|^2$ for the convex combination $tx + (1-t)x_0$ to conclude that for $t \in (0,1)$,

$$\begin{aligned} u(x_0) - \frac{c}{2}\|(tx + (1-t)x_0) - x_0\|^2 &\leq u(x + (1-t)x_0) - \frac{c}{2}\|(tx + (1-t)x_0) - x_0\|^2 \\ &\leq t\left(u(x) - \frac{c}{2}\|x - x_0\|^2\right) + (1-t)u(x_0), \end{aligned}$$

which implies

$$tu(x_0) \leq t\left(u(x) - \frac{c}{2}\|x - x_0\|^2\right) + t^2\|x - x_0\|^2.$$

Dividing by t and sending $t \to 0^+$, we obtain the desired estimate $u(x) - u(x_0) \geq (c/2)\|x - x_0\|^2$, which also implies uniqueness of the minimizer. □

11.2 Basic lemmas on log-concave measures

Definition 11.2.1. A probability measure μ on $\mathbb{R}^d$ is said to be *log-concave* if it has a density ρ with respect to Lebesgue measure such that $\log \rho$ is a concave function $\mathbb{R}^d \to [-\infty, +\infty)$.

Definition 11.2.2. If $V : \mathbb{R}^d \to \mathbb{R}$ and e^{-V} is integrable, then we can define a probability measure μ by

$$d\mu(x) = \frac{1}{\int e^{-V}} e^{-V(x)}\, dx.$$

We say that μ is the *measure associated to the potential* V.

Conversely, if μ has density $\rho > 0$, then μ is the measure associated to $V := -\log \rho$. Since we included the normalizing factor $1/\int e^{-V}$, the measure, μ is unchanged if we add a constant to the function V, and μ only determines V up to an additive constant. Nonetheless, by a slight abuse of terminology, we will often call V *the potential associated to* μ.

Observation 11.2.3. *If $\int e^{-V}$ is finite and μ is the measure associated to V, then μ is log-concave if and only if V is convex.*

Observation 11.2.4. *If $HV \geq cI$ for some $c > 0$, then $\int e^{-V} < \infty$ and hence there is a log-concave probability measure μ associated to V.*

Proof. By Lemma 11.1.6, V achieves a global minimum at some x_0, and $V(x) \geq V(x_0) + (c/2)\|x - x_0\|^2$. It is well known that $e^{-(c/2)\|x-x_0\|^2}$ is integrable, and hence e^{-V} is integrable. $\square$

We will often use the following integration-by-parts formula.

Lemma 11.2.5. *Let $V : \mathbb{R}^d \to \mathbb{R}$ with $cI \leq HV \leq CI$ for some $0 < c \leq C$, and let μ be the associated measure. Let $f : \mathbb{R}^d \to \mathbb{R}^d$ be a C^1 function satisfying*

$$\|f(x)\| \leq K_1 e^{K_2\|x\|}$$
$$|\nabla^\dagger f| \leq K_1' e^{K_2'\|x\|}$$

for some constants K_1, K_2, K_1', K_2'. Then

$$\int \langle \nabla V(x), f(x) \rangle \, d\mu(x) = \int \nabla^\dagger f(x) \, d\mu(x),$$

where both the integrals are well-defined (i.e. the functions are integrable).

Proof. By Lemma 11.1.4, V is continuously differentiable, and ∇V is C-Lipschitz. As in the previous lemma, note that $d\mu(x) = (1/\int e^{-V})e^{-V(x)}\,dx$ and $e^{-V(x)} \leq e^{-V(x_0)}e^{-(c/2)\|x-x_0\|^2}$, where x_0 is the minimizer of V. Now $\langle \nabla V, f \rangle$ and $\nabla^\dagger f$ both grow polynomially, so it is easy to check these functions times e^{-V} are integrable.

To check the formula, observe that

$$\nabla^\dagger[f e^{-V}] = (\nabla^\dagger f - \langle \nabla V, \nabla f \rangle)e^{-V}.$$

Thus, by the divergence theorem, for a ball R, we have

$$\int_{B(0,R)} (\nabla^\dagger f - \langle \nabla V, f \rangle)e^{-V} = \int_{\partial B(0,R)} \langle x, f(x) \rangle e^{-V(x)} \, d\sigma(x),$$

where σ is the $(d-1)$-dimensional surface measure. As in the previous observation, $e^{-V(x)} \leq e^{-V(x_0)}e^{-c\|x-x_0\|^2/2}$, where x_0 is the minimizer of V. Thus, our bounds on f are sufficient to guarantee that

$$\lim_{R\to\infty} \int_{B(0,R)} (\nabla^\dagger f - \langle \nabla V, f \rangle)e^{-V} = \int (\nabla^\dagger f - \langle \nabla V, f \rangle)e^{-V},$$

and

$$\lim_{R\to\infty} \int_{\partial B(0,R)} \langle x, f(x) \rangle e^{-V(x)} \, d\sigma(x) = 0.$$

Therefore,

$$\int (\nabla^\dagger f - \langle \nabla V, \nabla f \rangle) e^{-V} = 0.$$

This is what we wanted to prove since $d\mu(x) = (1/\int e^{-V}) e^{-V(x)}\, dx$. □

Corollary 11.2.6. *Let $V : \mathbb{R}^d \to \mathbb{R}$ satisfy $cI \le HV \le CI$ for some scalars $0 < c \le C$. Let μ be the probability measure associated to the potential V, and let X be a random variable whose distribution is μ. Then*

(1) $E[\nabla V(X)] = 0$,

(2) $E\langle \nabla V(X), X \rangle = d$,

(3) $dc \le E\|DV(X)\|^2 \le dC$,

(4) $d/C \le E\|X - E(X)\|_2^2 \le d/c$,

(5) $\|E(X)\|^2 \le E\|X\|^2 \le 2d/c + (1/c^2)\|\nabla V(0)\|^2$.

Proof. (1) Apply the previous lemma with $f(x) = y$ for some vector $y \in \mathbb{R}^d$ to obtain $E\langle \nabla V(X), y \rangle = \nabla^\dagger y = 0$. This holds for all y, hence $E[\nabla V(X)] = 0$.

(2) Use the previous lemma with $f(x) = x$ and $\nabla^\dagger f = d$.

(3) Let ρ_k be a smooth probability measure supported in $B(0, 1/k)$ and let $V_k = V * \rho_k$. Since V_k is smooth satisfies the assumptions of the previous lemma,

$$E\langle \nabla V(X), \nabla V_k(X) \rangle = E[\Delta V_k(X)].$$

But $c \le HV_k \le C$, and hence $dc \le E[\Delta V_k(X)] \le dC$. It is straightforward to check that $E\langle \nabla V(X), \nabla V_k(X) \rangle \to E\|\nabla V(X)\|^2$ as $k \to \infty$.

(4) Observe that by Lemma 11.1.4,

$$cE\|X - E(X)\|^2 \le E\langle \nabla V(X) - \nabla V(E(X)), X - E(X) \rangle \le CE\|X - E(X)\|^2.$$

We can evaluate the middle term above as

$$\begin{aligned} &E\langle \nabla V(X) - \nabla V(E(X)), X - E(X) \rangle \\ &\qquad = \langle \nabla V(X), X \rangle - \langle E[\nabla V(X)], X \rangle + E\langle \nabla V(E(X)), X - E(X) \rangle = d + 0 + 0, \end{aligned}$$

and thus rearranging the inequality finishes the proof.

(5) It follows from Hölder's inequality that $\|E(X)\|^2 \le E\|X\|^2$. Now observe that

$$\begin{aligned} E\|X\|^2 &\le \frac{1}{c} E\langle \nabla V(X) - \nabla V(0), X \rangle \\ &= \frac{d}{c} + \frac{1}{c} E\langle \nabla V(0), X \rangle. \end{aligned}$$

Then we use the arithmetic-geometric mean inequality to conclude that

$$E\langle \nabla V(0), X\rangle = E\langle c^{-1/2}\nabla V(0), c^{1/2}X\rangle \le \frac{1}{2c}\|\nabla V(0)\|^2 + \frac{c}{2}E\|X\|^2.$$

Hence,

$$E\|X\|^2 \le \frac{d}{c} + \frac{1}{2c^2}\|\nabla V(0)\|^2 + \frac{1}{2}E\|X\|^2,$$

which rearranges to the desired inequality. □

Lemma 11.2.7. *Let X be a random variable in $\mathbb{R}^d$, let $G : \mathbb{R}^d \to \mathbb{R}^k$ be Lipschitz, and let $\|G\|_{\mathrm{Lip}}$ denote its Lipschitz semi-norm. Then*

$$\|G(x) - E(G(X))\| \le \|G\|_{\mathrm{Lip}} \left(\|x - E(X)\|_2 + (E\|X - E(X)\|_2^2)^{1/2}\right).$$

Proof. Note that

$$\begin{aligned}\|G(x) - E(G(X))\| &\le \|G\|_{\mathrm{Lip}} E\|x - X\|_2 \\ &\le \|G\|_{\mathrm{Lip}} (\|x - E(X)\| + E\|X - E(X)\|_2) \\ &\le \|G\|_{\mathrm{Lip}} \left(\|x - E(X)\| + (E\|X - E(X)\|_2^2)^{1/2}\right).\end{aligned}$$ □

Corollary 11.2.8. *Let V satisfy $cI \le HV \le CI$ for some $0 < c \le C$, and let μ be the associated measure. Let $G : \mathbb{R}^d \to \mathbb{R}^k$ be Lipschitz. Then*

$$\left\|G(x) - \int G\,d\mu\right\| \le \|G\|_{\mathrm{Lip}} \left(\|x\| + \frac{1}{c}\|\nabla V(0)\| + \frac{3d^{1/2}}{c^{1/2}}\right)$$

Proof. Using the previous lemma,

$$\left\|G(x) - \int G\,d\mu\right\| \le \|G\|_{\mathrm{Lip}} \left(\|x - E(X)\| + (E\|X - E(X)\|^2)^{1/2}\right).$$

But by Corollary 11.2.6 (5) and the triangle inequality,

$$\|E(X)\| \le \left(\frac{2d}{c} + \frac{1}{c^2}\|\nabla V(0)\|^2\right)^{1/2} \le \frac{2d^{1/2}}{c^{1/2}} + \frac{1}{c}\|\nabla V(0)\|,$$

and by Corollary 11.2.6 (4),

$$(E\|X - E(X)\|^2)^{1/2} \le \frac{d^{1/2}}{c^{1/2}}.$$ □

Corollary 11.2.9. *Let $V : \mathbb{R}^d \to \mathbb{R}$ satisfy $cI \le HV \le CI$ for some $0 < c \le C$, let μ be the corresponding measure, and let X be a random variable with distribution μ. Then*

$$\|\nabla V(x)\| \le C\left(\|x - E(X)\|_2 + \frac{d^{1/2}}{c^{1/2}}\right).$$

Proof. We apply Lemma 11.2.7 to $DV(X)$. Also, DV is C-Lipschitz by Lemma 11.1.4. By Lemma 11.2.6 $E(DV(X)) = 0$ and $E\|X - E(X)\|_2^2 \le d/c$. □

11.3 Transformations of log-concave measures

Log-concave measures are closed under affine changes of variables, independent joins, and marginals. As we will see, the first two claims are straightforward, but the third relies on substantial work of Prékopa and Leindler and Brascamp and Lieb.

Lemma 11.3.1. *Let μ be a probability measure associated to a potential $V : \mathbb{R}^d \to \mathbb{R}$ with $\int e^{-V} < \infty$. Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be an invertible linear transformation. Then $T_*\mu$ is given by the potential $\tilde{V} = V \circ T^{-1}$, and we have $\nabla \tilde{V} = (T^{-1})^* \nabla V \circ T^{-1}$.*

(1) If $HV \geq A$ for some self-adjoint linear transformation A, then $H\tilde{V} \geq (T^{-1})^* A T^{-1}$.

(2) If $HV \leq B$ for some self-adjoint linear transformation B, then $H\tilde{V} \leq (T^{-1})^* B T^{-1}$.

Proof. For any nonnegative function f, we have

$$\int f(y) d(T_*\mu)(y) = \int f(Tx)\, d\mu(x) = \frac{1}{\int e^{-V}} \int f(Tx) e^{-V(x)}\, dx.$$

Using the change of variables formula for integration,

$$\int f(Tx) e^{-V(x)}\, dx = \int f(y) e^{-V(T^{-1}y)} |\det T^{-1}|\, dy$$
$$\int e^{-V(x)}\, dx = \int e^{-V(T^{-1}y)} |\det T^{-1}|\, dy.$$

Thus,

$$\int f(y)\, d(T_*\mu)(y) = \frac{1}{\int e^{-\tilde{V}}} \int f(y) e^{-\tilde{V}(y)}\, dy$$

since the two terms $|\det T^{-1}|$ cancel. This is true for all f, hence $T_*\mu$ is given by $\tilde{V}$.

For (1), the composition of a convex function with a linear function is convex, so if $V(x) - (1/2)\langle Ax, x\rangle$ is convex, then so is $\tilde{V}(y) - (1/2)\langle AT^{-1}y, T^{-1}y\rangle = \tilde{V}(y) - (1/2)\langle (T^{-1})^* AT^{-1}y, y\rangle$. Claim (2) is symmetrical. □

Observation 11.3.2. *Let $V_1 : \mathbb{R}^{d_1} \to \mathbb{R}$ and $V_2 : \mathbb{R}^{d_2} \to \mathbb{R}$ such that $\int e^{-V_j} < +\infty$ and let μ_j be the probability measure given by the potential V_j for $j = 1, 2$. Then $\mu_1 \otimes \mu_2$ on $\mathbb{R}^{d_1+d_2}$ is the measure given by the potential $V(x_1, x_2) = V(x_1) + V(x_2)$. Hence, $\nabla V(x_1, x_2) = (\nabla V_1(x_1), \nabla V_2(x_2))$ and $HV(x_1, x_2) = HV_1(x_1) \oplus HV_2(x_2)$. Moreover,*

(1) If $HV_j \geq A_j$ for each j, then $HV \geq A_1 \oplus A_2$.

(2) If $HV_j \leq B_j$ for each j, then $HV \leq B_1 \oplus B_2$.

Prékopa and Leindler [Pre71, Lei72, Pre73] showed that if μ is a log-concave measure on $\mathbb{R}^{d_1+d_2}$, then so is the marginal ν given by $(\pi_1)_*\mu$, where $\pi_1 : \mathbb{R}^{d_1+d_2}$ is the projection onto the first d coordinates. This result was sharpened by Brascamp and Lieb in [BL76]. We summarize the setup and result of [BL76, Thm. 4.3] in our notation as follows.

Theorem 11.3.3. *Let μ be the probability measure on $\mathbb{R}^{d_1+d_2}$ given by a potential $V : \mathbb{R}^d \to \mathbb{R}$. Then the marginal μ_1 of μ on the first d_1 coordinates is the measure given by the potential*

$$V_1(x_1) = -\log \int_{\mathbb{R}^{d_2}} e^{-V(x_1,x_2)}\, dx_2,$$

which is a map $\mathbb{R}^{d_1} \to [-\infty, \infty)$. Let A and B be positive definite linear transformations of $\mathbb{R}^{d_1+d_2}$, written in block form as

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} \qquad B = \begin{pmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{pmatrix}$$

where $A_{i,j}$ and $B_{i,j}$ have size $d_i \times d_j$.

(1) If $HV \geq A$, then V_1 is finite everywhere and $HV_1 \geq A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1}$

(2) If $HV \leq B$, then $HV_1 \leq B_{1,1} - B_{1,2}B_{2,2}^{-1}B_{2,1}$.

The formula for V_1 holds because the marginal density should be given by

$$\frac{e^{-V_1(x_1)}}{\int e^{-V_1}} = \frac{1}{\int e^{-V}} \int e^{-V(x_1,x_2)}\, dx_2.$$

Since V_1 is only determined up to an additive constant, we can normalize it so that $\int e^{-V_1} = \int e^{-V}$.

The differences in notation between our statement and [BL76, Thm. 4.3] are as follows. The variables (x_1, x_2) in our statement are denoted (x, y) in [BL76]. The matrices $A_{1,1}$, $A_{1,2}$, $A_{2,1}$, and $A_{2,2}$ in our claim (1) are denoted as A, B, B^*, C in [BL76]. The matrix $A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1}$ is also known as the *Schur complement* $A/A_{2,2}$, and this matrix is denoted by D in [BL76]. The function $e^{-V(x_1,x_2)}$ in our statement is $F(x,y)\Phi(x,y)$ and $e^{-V_1(x_1)}$ is $G(x)e^{-\langle x,Dx\rangle}$.

The hard part of the theorem is claim (1), which [BL76] prove from the Brunn-Minkowski inequality. Claim (2) is easier because it can be proved by applying Hölder's inequality to appropriate functions.

As in the case of the other two types of transformations, let's record what happens to ∇V when we take a marginal.

Lemma 11.3.4. *With the setup of Theorem 11.3.3, suppose that $A \leq HV \leq B$. Then*

$$\nabla V_1(x_1) = \frac{\int \nabla_{x_1} V(x_1,x_2) e^{-V(x_1,x_2)}\, dx_2}{\int e^{-V(x_1,x_2)}\, dx_2}.$$

In other words, if (X_1, X_2) is a random variable with the distribution μ, then $\nabla V_1(X_1)$ is given by the conditional expectation

$$\nabla V_1(X_1) = E[\nabla_{x_1} V(X_1, X_2) | X_1].$$

Sketch of proof. By Lemma 11.1.4, ∇V is Lipschitz, hence grows polynomially. Because $HV \geq A$ and A is positive definite, e^{-V} decays fast enough that we can justify differentiating under the integral sign:

$$\begin{aligned}\nabla[e^{-V_1(x_1)}] &= \nabla \left[\int e^{-V(x_1,x_2)}\, dx_2 \right] \\ &= -\int \nabla_{x_1} V(x_1, x_2) e^{-V(x_1,x_2)}\, dx_2.\end{aligned}$$

So $\nabla V_1 = -\nabla \log(e^{-V_1}) = -\nabla[e^{-V_1}]/e^{-V_1}$, which produces the asserted formula since

$$e^{-V_1(x_1)} = \int e^{-V(x_1,x_2)}\, dx_2.$$

Furthermore, it is clear that $e^{-V(x_1,x_2)}\, dx_2$ is the conditional distribution of X_2 given $X_1 = x_1$, so that the integral formula for ∇V_1 describes a conditional expectation. □

11.4 Log-Sobolev inequality and concentration

The measures μ given by a potential V with $HV \geq cI$ for some $c > 0$ exhibit concentration of measure in high dimensions. This means that if X is a random variable chosen according to the measure and f is a Lipschitz function, then $f(X)$ is close to $E[f(X)]$ with high probability, if d is large. This is a consequence of the log-Sobolev inequality, which was introduced by Gross [Gro75]. For further information, see also [AGZ09, §2.3.3 and 4.4.2] and [BL00].

Definition 11.4.1. We say that a measure μ on $\mathbb{R}^d$ satisfies the *log-Sobolev inequality with constant c* if for all C^1 functions f,

$$\int f^2 \log \frac{f^2}{\int f^2\, d\mu}\, d\mu \leq \frac{2}{c} \int \|\nabla f\|^2\, d\mu. \tag{11.1}$$

The following theorem is due to Bakry and Émery [BE85]. Proofs can also be found in [BL00, §3] and [AGZ09, §4.4.2].

Theorem 11.4.2. *Suppose that μ is the measure associated to a potential $V : \mathbb{R}^d \to \mathbb{R}$ with $HV \geq cI$ for some $c > 0$. Then μ satisfies the log-Sobolev inequality with constant c.*

One use of the log-Sobolev inequality is that it implies concentration of measure, which can be formulated more precisely as follows.

Definition 11.4.3. We say that a measure μ on $\mathbb{R}^d$ satisfies *Herbst's concentration inequality with constant* c if for all Lipschitz functions $f : \mathbb{R}^d \to \mathbb{R}$ and $\delta > 0$, we have $E|f(X)| < +\infty$ and

$$P(f(X) - E[f(X)] \geq \delta) \leq e^{-c\delta^2/2\|f\|_{\mathrm{Lip}}^2} \tag{11.2}$$

where X is a random variable distributed according to μ.

Note that by symmetry this implies

$$P(|f(X) - E[f(X)]| \geq \delta) \leq 2e^{-c\delta^2/2\|f\|_{\mathrm{Lip}}^2}. \tag{11.3}$$

Similarly, for complex-valued functions, we can apply the inequality to the real and imaginary parts with δ replaced by $\delta/\sqrt{2}$ and thus obtain

$$P(|f(X) - E[f(X)]| \geq \delta) \leq 4e^{-c\delta^2/4\|f\|_{\mathrm{Lip}}^2}. \tag{11.4}$$

The next theorem was originally proved in an unpublished letter of Herbst. A proof can be found in [AGZ09, Lemma 2.3.3].

Theorem 11.4.4. *If μ satisfies the log-Sobolev inequality with constant c, then it satisfies Herbst's concentration inequality with constant c.*

11.5 Application to random matrices

We will apply the above results to $M_n(\mathbb{C})_{\mathrm{sa}}^d$ rather than $\mathbb{R}^d$. We equip $M_n(\mathbb{C})_{\mathrm{sa}}^d$ with the inner product

$$\langle x, y \rangle_2 = \sum_{j=1}^d \tau_n(x_j y_j),$$

where x_j, y_j are the coordinates of x, y in $M_n(\mathbb{C})$ and τ_n is the normalized trace $(1/n)\,\mathrm{Tr}$. Since $\tau_n(x_j y_j)$ is real, we see that $M_n(\mathbb{C})_{\mathrm{sa}}^d$ is a real inner product space of dimension dn^2. Therefore, any choice of an orthonormal basis yields a linear isometry from $M_n(\mathbb{C})_{\mathrm{sa}}^d$ to $\mathbb{R}^{dn^2}$. For $u : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{R}$, we can define the gradient $\nabla u(x) \in M_n(\mathbb{C})_{\mathrm{sa}}^d$ and the Hessian $Hu(x) : M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})_{\mathrm{sa}}^d$ by means of this isometry with $\mathbb{R}^{dn^2}$, and the gradient and Hessian are independent of the particular choice of isometry. Or equivalently, if u is a C^2 function, the gradient and Hessian are the vector and linear transformation satisfying

$$u(x) = u(x_0) + \langle \nabla u(x_0), x - x_0 \rangle_2 + \langle Hu(x_0)(x - x_0), x - x_0 \rangle_2 + o(\|x - x_0\|_2^2).$$

Similarly, we define Lebesgue measure on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ as the push-forward of Lebesgue measure by a linear isometry $\mathbb{R}^{dn^2} \to M_n(\mathbb{C})_{\mathrm{sa}}^d$.

We will consider measures of the form $d\mu(x) = (1/\int e^{-n^2 V})e^{-n^2 V(x)}\,dx$, where dx denoted the Lebesgue measure described above and $V : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{R}$ is a given potential. The normalizing factor of n^2 (which is the same as the dimension of $M_n(\mathbb{C})$) is the right one because it makes all the theorems about the large-n limit work out correctly, as we will see later.

All the previous results in this chapter can be applied to such random matrix models. For example, if we assume that $cI \leq HV \leq CI$, then in the previous results we substitute n^2V for V, $n^2\nabla V$ for ∇V, $n^2 HV$ for HV, dn^2 for d, cn^2 for c, Cn^2 for C, and so forth. So for instance, in Lemma 11.2.5, we would get

$$\int \langle \nabla V(x), f(x) \rangle \, \mu(x) = \frac{1}{n^2} \int \nabla^\dagger f(x) \, d\mu(x),$$

and in Corollary 11.2.6, we would get

$$E\langle \nabla V(X), X \rangle = \frac{1}{n^2} dn^2 = d,$$

and hence

$$\frac{d}{C} \leq E\|X - E(X)\|_2^2 \leq \frac{d}{c}.$$

Of particular interest are the log-Sobolev and concentration inequalities for random matrix tuples. Such concentration inequalities were first used in the random matrix context by Guionnet and Zeitouni [GZ00] in their proof of the convergence of empirical spectral distributions.

Corollary 11.5.1. *Suppose that $V : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{R}$ satisfies $HV \geq cI$ for some $c > 0$, and let μ be the probability measure on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ given by the potential V. Then μ satisfies the log-Sobolev inequality and Herbst's concentration inequality with constant $n^2 c$.*

One application of the concentration inequality is the following a priori tail bound for the operator norms of random matrices, which will be used throughout the paper for truncation arguments (see e.g. the proofs of Theorems 15.1.5 and 15.1.7).

Lemma 11.5.2. *Suppose that μ is a probability measure on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ satisfying Herbst's concentration inequality with constant cn^2, and X is a random variable with distribution μ. Let $f : M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})$ be Lipschitz with respect to $\|\cdot\|_2$. Then we have*

$$P\Big(\|f(X) - E[f(X)]\|_\infty \geq c^{-1/2} \|f\|_{\mathrm{Lip}} (\Theta + \delta) \Big) \leq e^{-n\delta^2/2}. \tag{11.5}$$

where Θ is a universal constant (independent of n and c).

Proof. First, observe that $\|x\|_\infty \leq n^{1/2} \|x\|_2$ for $x \in M_n(\mathbb{C})_{\mathrm{sa}}^d$. In particular, $g(x) = \|f(x) - E[f(X)]\|_\infty$ is $n^{1/2}\|f\|_{\mathrm{Lip}}$-Lipschitz with respect to $\|\cdot\|_2$, and thus

$$P(g(X) \geq E[g(X)] + \delta) \leq e^{-cn\delta^2/2\|f\|_{\mathrm{Lip}}^2},$$

which implies after a change of variables for δ that

$$P(g(X) \geq E[g(X)] + c^{-1/2}\|f\|_{\mathrm{Lip}}\delta) \leq e^{-n\delta^2/2}.$$

Therefore, it suffices to show that for some constant Θ, we have

$$E[g(X)] = E[\|f(X) - E[f(X)]\|_\infty] \leq \Theta c^{-1/2}\|f\|_{\mathrm{Lip}}. \tag{11.6}$$

We may assume without loss of generality that f is self-adjoint since in the general case, $f = (1/2)(f+f^*)+i(1/2i)(f-f^*)$, and each of the terms on the right-hand side is Lipschitz. Thus, the self-adjoint case would imply the non-self-adjoint case at the cost of doubling the constant Θ. Now to prove self-adjoint case, we use an "ϵ-net argument" that is well-known in random matrix theory (see [Tao12, §2.3.1]). Fix n. Let $\{\eta_j\}_{j=1}^J$ be a maximal collection of unit vectors in $\mathbb{C}^n$ such that $\|\eta_i - \eta_j\| \geq 1/3$ for all $i \neq j$. Since this collection is maximal, for every unit vector η, there exists some η_j with $\|\eta - \eta_i\| < 1/3$. Now if $a \in M_n(\mathbb{C})_{\mathrm{sa}}$, then there is a unit vector with $\|a\|_\infty = |\langle \eta, a\eta\rangle|$. We may then choose η_j with $|\eta - \eta_j| < 1/3$

$$\begin{aligned}
\|a\|_\infty &= |\langle \eta, a\eta\rangle| \\
&= |\langle \eta_j, a\eta_j\rangle| + |\langle \eta_j, a(\eta - \eta_j)\rangle_2| + |\langle \eta - \eta_j, a\eta\rangle| \\
&\leq |\langle \eta_j, a\eta_j\rangle| + \frac{1}{3}\|a\|_\infty + \frac{1}{3}\|a\|_\infty,
\end{aligned}$$

so that

$$\|a\|_\infty \leq 3\max_j |\langle \eta_j, a\eta_j\rangle|.$$

Note that the balls $\{B(\eta_j, 1/6)\}_{j=1}^J$ in $\mathbb{C}^n$ are disjoint and contained in $B(0, 7/6)$. Hence, we can estimate the number of vectors by

$$J \leq \frac{|B(0,7/6)|}{|B(0,1/6)|} = 7^{2N}.$$

Let $K = \|f\|_{\mathrm{Lip}}$. For a matrix $a \in M_n(\mathbb{C})_{\mathrm{sa}}$, we have

$$|\langle \eta_i, a\eta_j\rangle| \leq \|a\|_\infty \leq n^{1/2}\|a\|_2.$$

This implies that $x \mapsto \langle \eta_j, f(x)\eta_j\rangle$ is $Kn^{1/2}$-Lipschitz with respect to $\|\cdot\|_2$ and hence

$$P\Big(|\langle \eta_j, (f(X) - E[f(X)])\eta_j\rangle| \geq \delta\Big) \leq 2e^{-cn\delta^2/2K^2}$$

Since $\|a\|_\infty \leq 3\max_j\langle \eta_j, a\eta_j\rangle$, we have

$$\begin{aligned}
P\Big(\|f(X) - E[f(X)]\|_\infty \geq 3\delta\Big) &\leq 2Je^{-cn\delta^2/2} \\
&\leq 2\cdot 7^{2n}e^{-cn\delta^2/2K^2}.
\end{aligned}$$

Thus, for any $t_0 > 0$, we have

$$\begin{aligned} E[\|f(X) - E[f(X)]\|_\infty] &= \int_0^\infty P(\|f(X) - E[f(X)]\|_\infty \geq t)\, dt \\ &\leq \int_0^{t_0} 1\, dt + \int_{t_0}^\infty 2 \cdot 7^{2n} e^{-cnt^2/18K^2}\, dt \\ &\leq t_0 + \int_{t_0}^\infty 2 \cdot 7^{2n} \frac{t}{t_0} e^{-cnt^2/18K^2}\, dt \\ &= t_0 + 2 \cdot 7^{2n} \frac{9K^2}{cnt_0} e^{-cnt_0^2/18K^2}. \end{aligned}$$

Now substitute $t_0 = 6c^{-1/2}K(\log 7)^{1/2}$ and obtain (11.6) with

$$\Theta = 6(\log 7)^{1/2} + \frac{3}{(\log 7)^{1/2}}.$$

(In fact, for a fixed n, we may use $\Theta_n = 6(\log 7)^{1/2} + 3/n(\log 7)^{1/2}$ in the self-adjoint case.) □

Corollary 11.5.3. *Let μ and X be as in the previous lemma. Then for any constant K,*

$$E\left[e^{K\|X\|_\infty} 1_{\|X\|_\infty > \|E(X)\|_\infty + c^{-1/2}\Theta}\right] \leq d\sqrt{\frac{2\pi}{nc}} e^{K^2/2nc} e^{\|E(X)\|_\infty + c^{-1/2}\Theta}.$$

Proof. Applying the previous lemma with $f(x) = x_j$, we obtain

$$P(\|X_j - E(X_j)\|_\infty \geq c^{-1/2}(\Theta + \delta)) \leq e^{-n\delta^2/2}.$$

By the triangle inequality and union bound,

$$P(\|X\|_\infty \geq \|E(X)\|_\infty + c^{-1/2}(\Theta + \delta)) \leq de^{-n\delta^2/2}.$$

Let $M = \|E(X)\|_\infty + c^{-1/2}\Theta$. By the layer-cake decomposition,

$$\begin{aligned} E[e^{K\|X\|_\infty} 1_{\|X\|_\infty > M}] &= d\int_0^\infty e^{K(M+t)} P(\|X\|_\infty \geq M + t)\, dt \\ &\leq de^{KM} \int_0^\infty e^{Kt} e^{-nct^2/2}\, dt \\ &= de^{KM + K^2/2nc} \int_0^\infty e^{-nc(t-K/nc)^2/2}\, dt \\ &\leq de^{KM + K^2/2nc} \int_{\mathbb{R}} e^{-nct^2/2}\, dt \\ &= d\sqrt{\frac{2\pi}{nc}}\, e^{KM + K^2/nc}. \end{aligned}$$

□

One perhaps surprising application of Lemma 11.5.2 is the following, which shows that $\|\cdot\|_2$-Lipschitz unitarily equivariant functions are $\|\cdot\|_\infty$ bounded on $\|\cdot\|_\infty$-balls.

Lemma 11.5.4. *Let $f : M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})$ be Lipschitz with respect to $\|\cdot\|_2$ and equivariant under unitary conjugation (that is, $f(ux_1u^*, \dots, ux_du^*) = uf(x_1, \dots, x_d)u^*$ for unitaries u). Let $\alpha_j \in \mathbb{R}$. Then*

$$\|f(x) - \tau_n(f(x))\|_\infty \leq 2\sqrt{2}d^{1/2}\Theta\|f\|_{\mathrm{Lip}} \max_j \|x_j - \alpha_j 1\|_\infty,$$

where Θ is the universal constant from the previous lemma.

The proof of the lemma relies on the fact that Herbst's concentration inequality holds for the Haar measure on the special unitary group $\mathbb{SU}(n)$. First, recall that $\mathbb{SU}(n)$ is a Riemannian manifold. The tangent space to $\mathbb{SU}(n)$ at a point u can be isometrically mapped onto the self-adjoint $n \times n$ matrices of trace zero as follows. An element $x \in M_n(\mathbb{C})_{\mathrm{sa}}$ of trace zero corresponds to the element of the tangent space given by path $\gamma(t) = e^{itx}u$ through the point u, and using the eigenvalue decomposition, it is easy to see that $\det(e^{itx}) = e^{it \operatorname{Tr}(x)} = 1$. Thus, we can equip $\mathbb{SU}(n)$ with the Riemannian metric induced by the inner product $\langle \cdot, \cdot \rangle_2$ on $M_n(\mathbb{C})_{\mathrm{sa}}$, or alternatively we can use the inner product associated to the unnormalized trace which is $\langle x, y \rangle_{\mathrm{Tr}} = \operatorname{Tr}(xy) = n\langle x, y \rangle_2$.

The log-Sobolev inequality makes sense for probability measures on a Riemannian manifold; we only have to interpret the term $\|\nabla f\|^2$ as the norm of the gradient of f with respect to the Riemannian metric. Furthermore, Herbst's concentration inequality makes sense for an arbitrary metric space, using the Lipschitz seminorm of a function on the metric space with respect to the distance function. On a Riemannian manifold, the log-Sobolev inequality with constant c implies Herbst's concentration inequality with constant c, where in Herbst's concentration inequality we use the distance function given by the infimum of the length of paths from one point to another, where path length is computed using the Riemannian metric.

It follows from the Bakry-Émery criterion (on manifolds) that the Haar measure on the special unitary group satisfies the log-Sobolev inequality with constant $n/2$ if we use the Riemann metric coming from Tr; this result is due to Gromov and can be found in [AGZ09, Theorem 4.4.7 and Appendix F.6] and [Mec19, Theorem 5.16]. This implies that the Haar measure satisfies the log-Sobolev inequality with constant $n^2/2$ if we use the Riemannian metric coming from τ_n. Hence, it satisfies Herbst's concentration inequality with the same constant.

Proof of Lemma 11.5.4. Fix $x \in M_n(\mathbb{C})_{\mathrm{sa}}^d$. Let U be a Haar random element of $\mathbb{SU}(n)$, and let $X = UxU^* = (Ux_1U^*, \dots, Ux_dU^*)$, which is a random d-tuple of self-adjoint matrices. We want to apply the previous lemma, and hence we want to show that X satisfies Herbst's concentration inequality with a certain constant.

To this end, we will estimate the Lipschitz constant of the function $g : \mathbb{SU}(n) \to M_n(\mathbb{C})_{\mathrm{sa}}^d$ given by $g(u) = uxu^*$. Let $\|g\|_{\mathrm{Lip}}$ be the Lipschitz norm of g with respect to the distance function from the Riemannian metric on the unitary group. Then $\|g\|_{\mathrm{Lip}} = \sup_u \|Jg(u)\|$, where $Jg(u)$ denote the Jacobian of g which maps the tangent space of $\mathbb{SU}(n)$ at u to the tangent space of $M_n(\mathbb{C})_{\mathrm{sa}}^d$ at $g(u)$. Recall we identified the tangent space to $\mathbb{SU}(n)$ with $M_n(\mathbb{C})_{\mathrm{sa}}$, and clearly the tangent space to $M_n(\mathbb{C})_{\mathrm{sa}}^d$ can be identified with $M_n(\mathbb{C})_{\mathrm{sa}}^d$. Pick $a \in M_n(\mathbb{C})_{\mathrm{sa}}$ with trace zero. Then

$$Jg(u)[a] = \frac{d}{dt}\Big|_{t=0} g(e^{ita}u) = \frac{d}{dt}\Big|_{t=0} [e^{ita}uxu^* e^{-ta}] = i[a, g(u)],$$

where $[a, g(u)] = ag(u) - g(u)a \in M_n(\mathbb{C})^d$ is the commutator. But note that

$$\|[a, ux_j u^*]\|_2 = \|[a, u(x_j - \alpha_j)u^*]\|_2 \leq 2\|a\|_2 \|x_j - \alpha_j\|_\infty.$$

Thus,

$$\|[a, g(u)]\|_2 = \left(\sum_{j=1}^d \|[a, ux_j u^*]\|_2^2 \right)^{1/2} \leq 2d^{1/2} \|a\|_2 \max_j \|x_j - \alpha_j\|_\infty.$$

Therefore, g is K-Lipschitz, where $K = 2d^{1/2} \max_j \|x_j - \tau(x_j)\|_\infty$.

Because U satisfies Herbst's concentration inequality with constant $n^2/2$ and g is K-Lipschitz, we know that $X = g(U)$ satisfies Herbst's concentration inequality with constant $n^2/2K^2$. The reason for this is that whenever $h : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{R}$ is Lipschitz, $h(X) = h \circ g(U)$, and $h \circ g$ is Lipschitz with constant $\|h\|_{\mathrm{Lip}} K$.

Therefore, by the previous lemma

$$P\Big(\|f(X) - E[f(X)]\|_\infty \geq \sqrt{2} K \|f\|_{\mathrm{Lip}} (\Theta + \delta) \Big) \leq e^{-n\delta^2/2}.$$

But note that $E[f(X)] = E[f(UxU^*)] = E[Uf(x)U^*] = \tau_n(f(x))$. Moreover, since $f(X) - E[f(X)] = U(f(x) - \tau_n(f(x)))U^*$, the random variable $f(X) - E[f(X)]$ is actually constant. Thus,

$$\|f(X) - E[f(X)]\|_\infty \leq \sqrt{2} K \Theta \|f\|_{\mathrm{Lip}} = 2\sqrt{2} d^{1/2} \Theta \|f\|_{\mathrm{Lip}} \max_j \|x_j - \alpha_j\|_\infty. \qquad \square$$

CHAPTER 12

Background: Entropy, Fisher's information, and transport

In this section, we review the entropy and Fisher's information of probability measures on $\mathbb{R}^d$, relative to Lebesgue measure and relative to Gaussian measure, as well as Talagrand's entropy-cost inequality. This background serves two main purposes in this paper. First, it motivates Voiculescu's two definitions of free entropy, paving the way for §16. Second, as motivation for our results on transport and their proofs, we will summarize Otto and Villani's approach to Talagrand's inequality.

12.1 Entropy and Fisher's information

Definition 12.1.1. The (classical, continuous) entropy of a measure $d\mu(x) = \rho(x)\,dx$ on $\mathbb{R}^d$ is defined as

$$h(\mu) := \int_{\mathbb{R}^n} -\rho \log \rho,$$

whenever the integral makes sense. If μ does not have a density, then we set $h(\mu) = -\infty$.

We will later use the following basic facts about the classical entropy, so for convenience we provide a proof.

Lemma 12.1.2. *Assume that μ is a probability measure on $\mathbb{R}^d$ with density ρ and that $\int \|x\|^2\,d\mu(x) < +\infty$.*

(1) The positive part of $-\rho\log\rho$ has finite integral with respect to Lebesgue measure and hence $\int -\rho \log \rho$ is well-defined in $[-\infty, +\infty)$.

(2) We have $h(\mu) \leq (d/2)\log 2\pi a e$, where $a = \int \|x\|^2\,d\mu(x)/d$, and equality is achieved in the case of a centered Gaussian with covariance matrix aI.

(3) Suppose $(\mu_k)_{k\in\mathbb{N}}$ is a sequence of probability measures with density ρ_k. Suppose $\rho_k \to \rho$ pointwise almost everywhere and that $\int \|x\|^2\,d\mu_k(x) \to \int \|x\|^2\,d\mu(x) < \infty$. Then $\limsup_{k\to\infty} h(\mu_k) \leq h(\mu)$.

(4) If ν is a probability measure with finite second moments, then $h(\mu * \nu) \geq h(\mu)$.

Proof. (1) Let $a = \int \|x\|^2 \, d\mu(x)/d$. Let $g(x) = (2\pi a)^{-d/2} e^{-\|x\|^2/2a}$ be the Gaussian measure with covariance matrix aI and let γ be the corresponding Gaussian measure. Let $\tilde{\rho} = \rho/g$ be the density of μ relative to the Gaussian. We write

$$\begin{aligned} -\rho(x) \log \rho(x) &= -\tilde{\rho}(x) \log \tilde{\rho}(x) \cdot g(x) - \tilde{\rho}(x) \log g(x) \cdot g(x) \\ &= -\tilde{\rho}(x) \log \tilde{\rho}(x) \cdot g(x) + \left(\frac{1}{2a} \|x\|^2 + \frac{d}{2} \log 2\pi a \right) \rho(x). \end{aligned}$$

The second term has a finite integral by assumption. The function $-t \log t$ is bounded above for $t \in \mathbb{R}$, and $g(x)$ is a probability density; thus, the positive part of $-\tilde{\rho} \log \tilde{\rho} \cdot g$ has finite integral. Hence, $\int -\rho \log \rho$ is well-defined in $[-\infty, \infty)$.

(2) The function $-t \log t$ is concave and its tangent line at $t = 0$ is $1 - t$ and hence $-t \log t \leq 1 - t$. Thus,

$$\int -\tilde{\rho} \log \tilde{\rho} \, d\gamma \leq \int (1 - \tilde{\rho}) \, d\gamma = 0,$$

so

$$h(\mu) \leq \int \left(\frac{1}{2a} \|x\|^2 + \frac{n}{2} \log 2\pi a \right) \rho(x) \, dx = \frac{n}{2} + \frac{n}{2} \log 2\pi a = \frac{n}{2} \log 2\pi e.$$

In the case where $\mu = \gamma$, we have $\tilde{\rho} = 1$ and hence $\int -\tilde{\rho} \log \tilde{\rho} = 0$.

(3) Let γ be the Gaussian measure with covariance matrix I, and g be its density. Let $\tilde{\rho}_k = \rho_k/g$. As before,

$$h(\mu_k) = \int -\tilde{\rho}_k \log \tilde{\rho}_k \, d\gamma + \int \left(\frac{1}{2} \|x\|^2 + \frac{n}{2} \log 2\pi \right) \, d\mu_k.$$

By assumption, the second term converges to $\int (\frac{1}{2} \|x\|^2 + \frac{n}{2} \log 2\pi) \, d\mu$ as $k \to \infty$. Since the function $-t \log t$ is bounded above and γ is a probability measure, the integral of the positive part of $-\tilde{\rho}_k \log \tilde{\rho}_k$ converges to the corresponding quantity for ρ. For the negative part, we can apply Fatou's lemma. This yields $\limsup_{k \to \infty} h(\mu_k) = h(\mu)$.

(4) We can assume without loss of generality that one of the measures, say μ, has finite entropy. Then $\mu * \nu$ has a density given almost everywhere by $\tilde{\rho}(x) = \int \rho(x - y) \, d\nu(y)$. Since $-t \log t$ is concave, Jensen's inequality implies that

$$-\tilde{\rho}(x) \log \tilde{\rho}(x) \geq \int -\rho(x - y) \log \rho(x - y) \, d\nu(y).$$

The right hand side is $\int \int -\rho(x-y) \log \rho(x-y) \, d\nu(y) \, dx = \int \int -\rho(x-y) \log \rho(x-y) \, dx \, d\nu(y) = h(\mu)$, where the exchange of order is justified because we know that $-\rho \log \rho$ is integrable since $h(\mu) > -\infty$. Therefore, $h(\mu * \nu) = \int -\tilde{\rho} \log \tilde{\rho} \geq h(\mu)$. □

The classical Fisher information of a probability measure μ on $\mathbb{R}^d$ describes how the entropy changes when μ is convolved with a Gaussian. Suppose μ is given by the smooth density $\rho > 0$ on $\mathbb{R}^d$, and let γ_t be the multivariable Gaussian measure on $\mathbb{R}^d$ with covariance matrix tI, which has the density $(2\pi)^{-d/2}e^{-\|x\|^2/2}$. Then the density ρ_t for $\mu_t = \mu * \gamma_t$ evolves according to the heat equation $\partial_t \rho_t = (1/2)\Delta\rho_t$. Integration by parts shows that $\partial_t h(\mu_t) = (1/2)\int \|\nabla\rho_t/\rho_t\|^2 d\mu_t$ (which we justify in more detail below).

The *Fisher information* of μ represents the derivative at time zero and it is defined as

$$\mathcal{I}(\mu) := \int \left\| \frac{\nabla\rho}{\rho} \right\|^2 d\mu.$$

The Fisher information is the $L^2(\mu)$ norm of the function $-\nabla\rho(x)/\rho(x)$, which is known as the *score function*.

If X is a random variable with smooth density ρ, then the $\mathbb{R}^d$-valued random variable $\Xi = -\nabla\rho(X)/\rho(X)$ satisfies the integration-by-parts relation

$$E[\Xi \cdot f(X)] = -\int \frac{\nabla\rho(x)}{\rho(x)} f(x)\rho(x)\,dx = \int \rho(x)\nabla f(x)\,dx = E[\nabla f(X)] \text{ for } f \in C_c^\infty(\mathbb{R}^d), \tag{12.1}$$

or equivalently $E[\Xi_j f(X)] = E[\partial_j f(X)]$ for each j. We remark that this equation also implies $E[\langle \Xi, f(X)\rangle] = E[\nabla^\dagger f(X)]$, which we already saw in a special case in Lemma 11.2.5.

In fact, the integration-by-parts relation $E[\Xi \cdot f(X)] = E[\nabla f(X)]$ makes sense even if we do not assume that X has a smooth density. If X is an $\mathbb{R}^d$-valued random variable on the probability space (Ω, P), we say that an $\mathbb{R}^d$-valued random variable $\Xi \in L^2(\Omega, P)$ is a *score function for* X if $E[\Xi \cdot f(X)] = E[\nabla f(X)]$ and if each Ξ_j is in the closure of $\{f(X) : f \in C_c^\infty(\mathbb{R}^d)\}$ in $L^2(\Omega, P)$.

In other words, this means that Ξ is a function of X (up to almost sure equivalence) and satisfies the integration-by-parts relation. Since the integration-by-parts relation uniquely determines the $L^2(\Omega, P)$ inner product of Ξ_j and $f(X)$ for all $f \in C_c^\infty(\mathbb{R}^d)$, it follows that the score function defined in this sense is unique (up to almost sure equivalence), and that it is given by $f(X)$ for some f that only depends on the law of X. Thus, we may unambiguously define the *Fisher information* $\mathcal{I}(\mu) = E[|\Xi|^2]$ if $X \sim \mu$ and Ξ is a score function for X, and $\mathcal{I}(\mu) = \infty$ if no score function exists.

The probabilistic viewpoint enables the use of conditional expectations to produce score functions and estimate Fisher information. See [Voi98a, Proposition 3.7] for the free case. This is also related to the conditional expectation formula Lemma 11.3.4 that we discussed for log-concave measures.

Lemma 12.1.3. *Suppose that X and Y are independent $\mathbb{R}^d$-valued random variables with $X \sim \mu$ and $Y \sim \nu$. If Ξ is a score function for X, then $E[\Xi | X + Y]$ is a score function for $X + Y$. In particular,*

$$\mathcal{I}(\mu * \nu) \leq \min(\mathcal{I}(\mu), \mathcal{I}(\nu)).$$

Proof. Because X and Y are independent, we have for $g \in C_c^\infty(\mathbb{R}^d \times \mathbb{R}^d)$ that $E[\Xi_j g(X, Y)] = E[\partial_{X_j} g(X, Y)]$. In particular, if $f \in C_c^\infty(\mathbb{R}^d)$, then

$$E[\Xi_j f(X + Y)] = E[\partial_{X_j}(f(X + Y))] = E[(\partial_j f)(X + Y)].$$

But $E[\Xi_j|X+Y]$ is the orthogonal projection onto the closed span of $\{f(X+Y) : f \in C_c^\infty(\mathbb{R}^d)\}$ and hence

$$E\left[E[\Xi_j|X+Y]f(X+Y)\right] = E[\partial_j f(X+Y)].$$

So $\mathcal{I}(\mu * \nu) = E[|E[\Xi|X+Y]|^2] \leq E[\|\Xi\|^2] = \mathcal{I}(\mu)$. By symmetry, $\mathcal{I}(\mu * \nu) \leq \mathcal{I}(\nu)$. □

The entropy of a measure μ can be recovered by integrating the Fisher information of $\mu * \gamma_t$. The following integral formula was the motivation for Voiculescu's definition of non-microstates free entropy χ^* (see §16). We include the proof for the reader's convenience. See also [Bar96, Lemma 1] and for the free case, see [Voi98a, Proposition 7.6].

Lemma 12.1.4. *Let μ be a probability measure on $\mathbb{R}^d$ with finite variance and with density ρ, and let γ_t be the Gaussian measure with covariance matrix tI. If $a = (1/d)\int \|x\|^2 \, d\mu(x)$, then for $t \geq 0$,*

$$\frac{d}{a+t} \leq \mathcal{I}(\mu * \gamma_t) \leq \min\left(\frac{d}{t}, \mathcal{I}(\mu)\right). \tag{12.2}$$

Moreover,

$$h(\mu * \gamma_t) - h(\mu) = \frac{1}{2}\int_0^t \mathcal{I}(\mu * \gamma_s)\, ds \tag{12.3}$$

and

$$h(\mu) = \frac{1}{2}\int_0^\infty \left(\frac{d}{1+s} - \mathcal{I}(\mu * \gamma_t)\right) ds + \frac{d}{2}\log 2\pi e. \tag{12.4}$$

Proof. To prove (12.2), suppose $t \geq 0$ and let X and Y be random variables with the laws μ and γ_t respectively. The lower bound is trivial if $\mathcal{I}(\mu * \gamma_t) = \infty$, so suppose that $X+Y$ has a score function Ξ. Then the integration-by-parts relation and independence of Y from Ξ shows that $E\langle \Xi, X+Y\rangle = d$. Thus,

$$E[\|\Xi\|^2] \geq \frac{\|E\langle \Xi, X+Y\rangle\|^2}{E\|X+Y\|^2} = \frac{d^2}{da+dt} = \frac{d}{a+t}.$$

The upper bound is trivial in the case where $t = 0$. If $t > 0$, then by the previous lemma $\mathcal{I}(\mu * \sigma_t) \leq \min(\mathcal{I}(\mu), \mathcal{I}(\gamma_t))$. Moreover, a direct computation shows that if $Y \sim \sigma_t^{(n)}$, then the score function is $(1/t)Y$ and the Fisher information is d/t.

Next, to prove (12.3), let $\mu_t := \mu * \gamma_t$. Then μ_t has a smooth density ρ_t. We claim that if $0 < \delta < t$, then

$$h(\mu_t) - h(\mu_\delta) = \frac{1}{2}\int_\delta^t \mathcal{I}(\mu_s)\, ds = \frac{1}{2}\int_\delta^t \int \frac{\|\nabla \rho_s(x)\|^2}{\rho_s(x)}\, dx\, ds. \tag{12.5}$$

This will follow from integration by parts, but to give a complete justification, we first introduce a smooth compactly supported "cutoff" function $\psi_R : \mathbb{R}^d \to \mathbb{R}$ such that $0 \leq \psi_R \leq 1$ and $\psi_R(x) = 1$ when $\|x\| \leq R$ and $\psi_R(x) = 0$ when $\|x\| \geq 2R$. By taking ψ_R to be the rescaling by R of some fixed function, we can arrange that $\|\nabla\psi_R(x)\|_2 \leq C/R$ for some constant C. Because $\partial_s\rho_s = (1/2)\Delta\rho_s$, we have

$$\begin{aligned}\frac{d}{dt}\left[-\int \psi_R\rho_s\log\rho_s\right] &= -\frac{1}{2}\int \psi_R\cdot(\Delta\rho_s\log\rho_s + \Delta\rho_s) \\ &= \frac{1}{2}\int \psi_R\frac{\|\nabla\rho_s\|^2}{\rho_s} + \frac{1}{2}\int \nabla\psi_R\cdot\nabla\rho_s\cdot(1+\log\rho_s),\end{aligned}$$

This implies

$$\begin{aligned}&-\int \psi_R\rho_t\log\rho_t + \int \psi_R\rho_\delta\log\rho_\delta \\ &\qquad = \frac{1}{2}\int_\delta^t\int \psi_R\|\nabla\rho_s/\rho_s\|^2\,d\mu_s - \frac{1}{2}\int_\delta^t\int(\nabla\psi_R\cdot\nabla\rho_s/\rho_s)(1+\log\rho_s)\,d\mu_s. \quad (12.6)\end{aligned}$$

We must now take the limit of each term as $R\to+\infty$. For the first term on the right hand side, the monotone convergence theorem yields

$$\lim_{R\to+\infty}\int_\delta^t\int \psi_R\|\nabla\rho_s/\rho_s\|^2\,d\mu_s\,ds = \int_\delta^t \mathcal{I}(\mu_s)\,ds.$$

The second term on the right hand side of (12.6) can be estimated as follows. Note that $\mu_s = \mu * \gamma_t$ and that γ_s has a density that is bounded uniformly for $s\in[\delta,t]$ and $x\in\mathbb{R}^d$. Therefore, ρ_s is uniformly bounded and hence $\log\rho_s$ is uniformly bounded above. To obtain a lower bound on $\log\rho_s$, first note that there is a $K>0$ such that

$$\mu(x : \|x\|\leq K)\geq 1/2.$$

Now if $x\in\mathbb{R}^d$ and $\|y\|\leq K$, then $\|x-y\|\geq\|x\|-K$ and hence $\|x-y\|^2\leq\|x\|^2-2K\|x\|+K^2\geq 2\|x\|^2+2K^2$, where the last inequality follows because $2K\|x\|\leq(1/2)\|x\|^2+2K^2$ by the arithmetic geometric mean inequality. Therefore, letting Z be the normalizing constant $(2\pi t)^{-d/2}$ for γ_t, we have

$$\begin{aligned}\rho_s(x) &= \frac{1}{Z}\int e^{-(1/2t)\|x-y\|^2}\,d\mu(y) \\ &\geq \frac{1}{Z}\int_{\|y\|\leq K}\int e^{-(1/2t)\|x-y\|^2}\,d\mu(y) \\ &\geq \frac{1}{Z}\int_{\|y\|\leq K} e^{-(1/t)(\|x\|^2+K^2)}\,d\mu(y) \\ &\geq \frac{e^{-NK^2/t}}{2Z}e^{-(1/t)\|x\|^2},\end{aligned}$$

so that $\log\rho_s\geq K'-(1/t)\|x\|^2$ for some constant K'. In particular, combining our upper and lower bounds, there is a constant α such that for sufficiently large x, we have $|1+\log\rho_s|\leq$

$\alpha\|x\|^2$. Recall that $\nabla\psi_R(x)$ is supported when $R \leq \|x\| \leq 2R$ and bounded by C/R and thus $|\nabla\psi_R(x)| \leq C/\|x\|$. Altogether we have $|\nabla\psi_R(1+\log\rho_s)| \leq \beta\|x\|$ for some constant β when $\|x\|$ is large enough. Thus, the second term on the right hand side of (12.6) is bounded by

$$\begin{aligned}\int_\delta^t \int \|(\nabla\psi_R \cdot \Xi_s)(1+\log\rho_s)\|\, d\mu_s\, ds &\leq \beta \int_\delta^t \int_{\|x\|\geq R} \|x\|\|\nabla\rho_s(x)/\rho_s(x)\|\, d\mu_s(x)\, ds \\ &\leq \frac{1}{2}\beta \int_\delta^t \int_{\|x\|\geq R} (\|x\|^2 + \|\nabla\rho_s(x)/\rho_s(x)\|^2)\, d\mu_s(x)\, ds.\end{aligned}$$

The right hand side is the tail of the convergent integral

$$\int_\delta^t \int (\|x\|^2 + \|\nabla\rho_s(x)/\rho_s(x)\|^2)\, d\mu_s(x)\, ds = \int_\delta^t [(a+ds) + \mathcal{I}(\mu_s)]\, ds < +\infty,$$

and therefore it goes to zero as $R \to +\infty$ by the dominated convergence theorem.

As for the left hand side of (12.6), we can apply the dominated convergence theorem to $-\int \psi_R\rho_t \log\rho_t$ and $-\int \psi_R\rho_\delta\log\rho_\delta$ given our earlier estimate that ρ_s is subquadratic for each s. Thus, after taking $R \to \infty$ in (12.6), we obtain (12.5).

To complete the proof of (12.3), we must take $\delta \searrow 0$ in (12.5). We can take the limit of the right hand side of (12.5) by the monotone convergence theorem. For the left hand side of (12.5), Lemma 12.1.2 (3) implies that $\limsup_{\delta\searrow 0} h(\mu_\delta) \leq h(\mu)$ because $\rho_\delta \to \rho$ almost everywhere by Lebesgue differentiation theory. On the other hand, $h(\mu_\delta) \geq h(\mu)$ by Lemma 12.1.2 (4), hence $h(\mu_\delta) \to h(\mu)$, so (12.3) is proved.

To prove (12.4), we follow [Voi98a, Proposition 7.6]. First, suppose that $h(\mu) > -\infty$. Note that

$$h(\mu) = \frac{1}{2}\int_0^t \left(\frac{d}{1+s} - \mathcal{I}(\mu_s)\right) ds - \frac{d}{2}\log(1+t) + h(\mu_t).$$

If $h(\mu) > -\infty$, then $\int_0^1 \left(\frac{d}{1+s} - \mathcal{I}(\mu_s)\right) ds$ is finite. In light of (12.2), the integral from 1 to ∞ is also finite and by the dominated convergence theorem

$$\lim_{t\to\infty} \frac{1}{2}\int_0^t \left(\frac{d}{1+s} - \mathcal{I}(\mu_s)\right) ds = \frac{1}{2}\int_0^\infty \left(\frac{d}{1+s} - \mathcal{I}(\mu_s)\right) ds.$$

It remains to understand the behavior of $h(\mu_t) - (d/2)\log(1+t)$. By Lemma 12.1.2 (4) and (2),

$$h(\mu_t) \geq h(\gamma_t) = \frac{d}{2}\log 2\pi e t = \frac{d}{2}\log 2\pi e + \frac{d}{2}\log t.$$

On the other hand, by Lemma 12.1.2 (2), since $\int \|x\|^2\, d\mu_t(x) = d(a+t)$, we have

$$h(\mu_t) \leq \frac{d}{2}\log 2\pi e(a+t) = \frac{d}{2}\log 2\pi e + \frac{d}{2}\log(a+t).$$

As $t \to \infty$, we have $\log(1+t) - \log(a+t) \to 0$ and $\log(1+t) - \log t \to 0$ and therefore

$$h(\mu_t) - \frac{d}{2}\log(1+t) \to \frac{d}{2}\log 2\pi e = \frac{d}{2}\log 2\pi e.$$

Hence,

$$h(\mu) = \frac{1}{2}\int_0^\infty \left(\frac{d}{1+s} - \mathcal{I}(\mu_s) \right) ds + \frac{d}{2}\log 2\pi e,$$

which is equivalent to the asserted formula (12.4). In the case where $h(\mu) = -\infty$, we also have $\int_0^1 \left(\frac{d}{1+s} - \mathcal{I}(\mu_s) \right) ds = -\infty$ by (12.3), but the integral from 1 to ∞ is finite as shown above. So both sides of (12.4) are $-\infty$. □

12.2 Relative entropy and Fisher information

Up to this point, we have discussed the entropy of measures on $\mathbb{R}^d$ using the density with respect to Lebesgue measure. But in fact, for any measure space $(\Omega, \mathcal{F}, m)$, there is a notion of entropy for measures on Ω.

Definition 12.2.1. If μ is a measure on Ω with $d\mu(x) = \rho(x)dm(x)$, then we set

$$h(\mu \| m) := \int -\rho \log \rho \, dm,$$

and we define $h(\mu \| m) = -\infty$ if there is no density. We call this the *entropy of μ relative to m*.

Thus, for instance,

(1) If Ω is $\mathbb{R}^d$ and m is Lebesgue measure, we obtain the continuous entropy discussed in the last section.

(2) If Ω is discrete and m is counting measure, then we obtain the discrete entropy of a probability mass function.

(3) In the case where m is counting measure, the entropy is nonnegative. To the contrary, in the case where m is a probability measure, the entropy is always ≤ 0 (this follows from Jensen's inequality).

In fact, many authors change the sign of $h(\mu \| \nu)$ when ν is a probability measure, in order to make it nonnegative. However, we will allow the sign to remain negative for consistency, because, for instance, we want to convert between Gaussian measure and Lebesgue measure as our background measures without sign changes.

If the background measure space Ω is $\mathbb{R}^d$ and the measure m is given by a probability density w, then there is a version of Fisher information just as in the case of Lebesgue measure.

Definition 12.2.2. Let m be a measure on $\mathbb{R}^d$, and let μ be a measure whose density with respect to m is ρ. Then we define the *score function of* μ *relative to* m as $-\nabla(\log\rho)$, and the *Fisher information* as

$$\mathcal{I}(\mu\|m) = \int \|-\nabla(\log\rho)\|^2 \, dm.$$

Suppose that m is given by a C^1 and strictly positive density w. Then the score function satisfies the integration-by-parts formula

$$\int -\nabla(\log\rho) f \, d\mu = \int \nabla f \, d\mu + \int f \nabla(\log w) \, d\mu \text{ for } f \in C_c^\infty(\mathbb{R}^d),$$

the analogue of (12.1). This implies also that for $f \in C_c^\infty(\mathbb{R}^d, \mathbb{R}^d)$,

$$\int \langle -\nabla(\log\rho), f\rangle \, d\mu = \int \nabla_w^\dagger f \, d\mu,$$

where

$$\nabla_w^\dagger f = \nabla^\dagger f + \langle \nabla(\log w), f\rangle,$$

which is formally the adjoint of ∇ on $L^2(w\,dx)$.

There is also an analogue of (12.3) in the relative context. As these results play only a heuristic role in this paper, the reader may assume in the following discussion that w and ρ have as much smoothness and decay at ∞ as needed, and we will not give the analytic justifications for the statements.

Let $\Delta_w = \nabla_w^\dagger \nabla$ be the Laplacian relative to w. Let ρ_t evolve according to the heat equation

$$\partial_t \rho_t = \Delta_w \rho_t = \Delta\rho_t + \langle \nabla(\log w), \nabla\rho_t\rangle$$

(compare [OV00, eq. (34)]), and let μ_t be the measure $d\mu_t = \rho_t \, w \, dx$. Then

$$\begin{aligned}
\partial_t h(\mu_t\|m) &= -\partial_t \int \rho_t (\log\rho_t) w \\
&= -\int (\partial_t\rho_t \log\rho_t + \rho_t(\partial_t\rho_t/\rho_t)) w \\
&= -\int (\Delta_w\rho_t \log\rho_t + \Delta_w\rho_t) w \\
&= \int \langle \nabla\rho_t, \nabla(1 + \log\rho_t)\rangle w \\
&= \int \|\nabla\rho_t/\rho_t\|^2 \rho_t w \\
&= \mathcal{I}(\mu_t\|m)
\end{aligned}$$

(compare [OV00, eq. (37)], but beware of the sign change in h). So $h(\mu_t\|m) - h(\mu\|m) = \int_0^t \mathcal{I}(\mu_s\|m)\,ds$, from which we conclude that $\lim_{t\to\infty} h(\mu_t\|m)$ exists and

$$-h(\mu\|m) = \int_0^\infty \mathcal{I}(\mu_t\|m) - \lim_{t\to\infty} h(\mu_t\|m).$$

It is often the case that $\lim_{t\to\infty} h(\mu_t\|m) = 0$; for instance, if m satisfies a log-Sobolev inequality, then by Lemma 12.2.3 below, $-h(\mu_t\|m) \le \mathrm{const}\,\mathcal{I}(\mu_t\|m)$, which goes to zero because $\int_0^\infty \mathcal{I}(\mu_t\|m)\,dt < \infty$. And if we know that $\lim_{t\to\infty} h(\mu_t\|m) = 0$, then we can conclude that $-h(\mu\|m) = \int_0^\infty \mathcal{I}(\mu_t\|m)$, which is the relative version of (12.4).

The following lemma is an information-theoretic interpretation of the log-Sobolev inequality 11.1 (compare [OV00, eq. (5)] for instance).

Lemma 12.2.3. *Let m be a probability measure on $\mathbb{R}^d$. Then m satisfies the log-Sobolev inequality with constant $c > 0$ if and only if*

$$-h(\mu\|m) \le \frac{1}{2c}\mathcal{I}(\mu\|m), \tag{12.7}$$

whenever $d\mu(x) = \rho(x)\,dm(x)$ for a C^1 density ρ.

Proof. ($\implies$) By approximation arguments, we can reduce to the case where $\rho > 0$ everywhere and ρ is bounded above. Let $f = \sqrt{\rho}$. Then

$$\int f^2 \log \frac{f^2}{\int f^2\,dm}\,dm = \int f^2 \log f^2\,dm = -h(\mu\|m).$$

By the log-Sobolev inequality, this is less than or equal to

$$\begin{aligned}\frac{2}{c}\int \|\nabla f\|^2\,dm &= \frac{2}{c}\int \|\nabla\sqrt{\rho}\|^2\,dm\\ &= \frac{1}{2c}\int \|\nabla\rho/\rho^{1/2}\|^2\,dm\\ &= \frac{1}{2c}\int \|\nabla\rho/\rho\|^2\rho\,dm\\ &= \frac{1}{2c}\mathcal{I}(\mu\|m).\end{aligned}$$

For the converse direction, given f and $\epsilon > 0$, we substitute $\rho = (\epsilon + f^2)/\int(\epsilon + f^2)\,dm$ into (12.7) and get

$$\int \frac{\epsilon + f^2}{\int(\epsilon+f^2)\,dm} \log \frac{\epsilon + f^2}{\int(\epsilon+f^2)\,dm}\,dm \le \frac{2}{c}\int \left\|\frac{f\nabla f}{\epsilon + f^2}\right\|^2 \frac{\epsilon + f^2}{\int(\epsilon+f^2)\,dm}\,dm.$$

Then we cancel out the factors of $1/\int(\epsilon + f^2)\,dm$ on both sides and take $\epsilon \searrow 0$ to get

$$\int f^2 \log \frac{f^2}{\int f^2\,dm}\,dm \le \frac{2}{c}\int_{f\neq 0} \|\nabla f\|^2\,dm \le \frac{2}{c}\int \|\nabla f\|^2\,dm. \qquad \square$$

Remark 12.2.4. There is not a direct relative analogue of Lemma 12.1.3 as that relies on translation invariance of Lebesgue measure.

12.3 Transportation and Talagrand's inequality

As in the previous section, let m be a probability measure on $\mathbb{R}^d$ given by a density w, let $d\mu(x) = \rho(x)\,dm(x)$ be another probability measure. Assume ρ and w have sufficient smoothness and decay at ∞, and let ρ_t evolve by the heat equation for Δ_w.

In their proof of Talagrand's inequality from the log-Sobolev inequality, Otto and Villani [OV00] used the following observation from the theory of transport equations.

Lemma 12.3.1. *Suppose that $\partial_t \rho_t = \Delta_w \rho_t$ for some $w > 0$. Let $V_t = -\log \rho_t$. Assume that ∇V_t is smooth in (x,t) and ∇V_t is globally Lipschitz in x uniformly for t in any compact time interval, and let $F_{s,t}(x)$ evolve according to the ODE*

$$\partial_s F_{s,t}(x) = \nabla V_s(x), \qquad F_{t,t}(x) = x.$$

Then $(F_{s,t})_* \mu_t = \mu_s$.

Proof. The Lipschitz assumption guarantees that the solution to the ODE is well-defined by the Picard-Lindelöf theorem. It is a well known and easy fact in ODE theory that $F_{s,t} \circ F_{t,u} = F_{s,u}$ and in particular $F_{s,t}^{-1} = F_{t,s}$, so that $F_{s,t}$ is a C^1 diffeomorphism.

Recall that μ_t has density $\rho_t w$ with respect to Lebesgue measure. Hence, by the change of variables formula for multiple integrals, the $(F_{s,t})_* \mu_t = (F_{t,s}^{-1})_* \mu_t$ has the density

$$(\rho_t w) \circ F_{t,s}\, |\det JF_{t,s}|,$$

where $JF_{t,s}$ denotes the Jacobian linear transformation. When $t = s$, this reduces to ρ_s, and hence to prove that $(F_{s,t})^* \mu_t = \mu_s$, it suffices to show that $(\rho_t w) \circ F_{t,s} |\det JF_{t,s}|$ is constant with respect to t.

To this end, we take the logarithm and then differentiate with respect to t. The logarithm is

$$(\log \rho_t) \circ F_{t,s} + (\log w) \circ F_{t,s} + \log |\det JF_{t,s}|. \tag{12.8}$$

Using the chain rule, the time-derivative of the first term is

$$\begin{aligned}
&\partial_t (\log \rho_t) \circ F_{t,s} + \langle \nabla (\log \rho_t) \circ F_{t,s}, \partial_t F_{t,s} \rangle \\
&= \frac{\partial_t \rho_t}{\rho_t} \circ F_{t,s} - \langle \nabla (\log \rho_t) \circ F_{t,s}, \nabla (\log \rho_t) \circ F_{t,s} \rangle \\
&= \frac{\Delta \rho_t + \langle \nabla (\log w), \rho_t \rangle}{\rho_t} \circ F_{t,s} + \|\nabla (\log \rho_t)\|^2 \circ F_{t,s} \\
&= \left(\frac{\Delta \rho_t}{\rho_t} + \frac{\langle \nabla (\log w), \nabla \rho_t \rangle}{\rho_t} - \frac{\|\nabla \rho_t\|^2}{\rho_t^2} \right) \circ F_{t,s}.
\end{aligned}$$

The derivative of the second term in (12.8) is

$$\langle \nabla (\log w) \circ F_{t,s}, \partial_t F_{t,s} \rangle = -\langle \nabla (\log w), \nabla (\log \rho_t) \rangle \circ F_{t,s} = -\frac{\langle \nabla (\log w), \nabla \rho_t \rangle}{\rho_t} \circ F_{t,s}.$$

This already cancels the middle term in our expression for the first term of (12.8). Now, for the third term in (12.8), observe that $F_{t+\epsilon,s} = F_{t+\epsilon,t} \circ F_{t,s}$, so that $JF_{t+\epsilon,s} = JF_{t+\epsilon,t} \circ F_{t,s} \cdot JF_{t,s}$, hence

$$\log|\det JF_{t+\epsilon,s}| = \log|\det JF_{t+\epsilon,t}| \circ F_{t,s} + \log|\det JF_{t,s}|.$$

Now let us differentiate in ϵ at $\epsilon = 0$. Since $F_{t,s}(x)$ is smooth in (x,t),

$$\frac{d}{d\epsilon}\Big|_{\epsilon=0} JF_{t+\epsilon,t} = J\frac{d}{d\epsilon}\Big|_{\epsilon=0} F_{t+\epsilon,t} = J(-\nabla(\log\rho_t)) = -J\left(\frac{\nabla\rho_t}{\rho_t}\right) = -\frac{H\rho_t}{\rho_t} + \frac{(\nabla\rho_t)(\nabla\rho_t)^*}{\rho_t^2}.$$

Since $F_{t+\epsilon,t}$ evaluates to the identity function when $\epsilon = 0$, we see that $\det F_{t+\epsilon,t} > 0$ for small ϵ, and moreover,

$$\log|\det JF_{t+\epsilon,s}| = \operatorname{Tr}\log JF_{t+\epsilon,s},$$

where the logarithm is evaluated by analytic functional calculus, that is, using the power series of log at 1. Differentiating at $\epsilon = 0$ results in

$$\operatorname{Tr}\left(-\frac{H\rho_t}{\rho_t} + \frac{(\nabla\rho_t)(\nabla\rho_t)^*}{\rho_t^2}\right) = -\frac{\Delta\rho_t}{\rho_t} + \frac{\|\nabla\rho_t\|^2}{\rho_t^2},$$

which implies that

$$\partial_t \log|\det JF_{t,s}| = \left(-\frac{\Delta\rho_t}{\rho_t} + \frac{\|\nabla\rho_t\|^2}{\rho_t^2}\right) \circ F_{t,s}.$$

This completes evaluating the third term of (12.8). When we add the three terms together, everything cancels, which proves that $(F_{s,t})_*\mu_t = \mu_s$ as desired. □

Remark 12.3.2. In Lemma 12.3.1, there is no reason that w has to be a probability density on $\mathbb{R}^d$; it works for any smooth positive function. In particular, it applies when $w = 1$, which is the case where m is Lebesgue measure.

Let us now explain how Otto and Villani used this construction of transport to prove that the log-Sobolev inequality implies Talagrand's inequality in [OV00, Thm. 1].

Definition 12.3.3. For two probability measures μ and ν on $\mathbb{R}^d$, we define the L^2 *Wasserstein distance* $W_2(\mu,\nu)$ to be the infimum of $\|X - Y\|_{L^2} := (E\|X-Y\|^2)^{1/2}$ where X and Y are random variables on the same probability space with $X \sim \mu$ and $Y \sim \nu$.

Definition 12.3.4. We say that a probability measure m on $\mathbb{R}^d$ satisfies *Talagrand's inequality with constant* c if

$$W_2(\mu,m)^2 \leq -\frac{2}{c}h(\mu\|m)$$

for all probability measures m.

Theorem 12.3.5 ([OV00, Thm. 1])**.** *Let m is given by a C^2 density w with $H(\log w) \leq KI$ for some constant K, and suppose m has finite variance. If m satisfies the log-Sobolev inequality with constant c, then it satisfies Talagrand's inequality with constant c.*

Sketch of proof. Let's only explain the case with some extra smoothness assumptions, since the general case can be handled with approximation arguments. Assume w is smooth and μ is given by a smooth density ρ with respect to w. Let ρ_t evolve according to the heat equation as above and let μ_t be the corresponding measure. By regularity theory for PDE, sufficient smoothness and decay conditions on w and ρ will guarantee that ρ_t is smooth.

Let X be a random variable with distribution μ, and let $X_t = F_{t,0}(X)$, so that $X_t \sim \mu_t$ by the previous lemma. Note that $F_{s,t}(X_t) = X_s$. Now observe that for $s \leq t$,

$$\begin{aligned}\|X_t - X_s\|_{L^2} &= \|F_{t,s}(X_s) - X_s\|_{L^2} \\ &= \left\| \int_s^t \nabla(\log \rho_u) \circ F_{u,s}(X_s)\, du \right\| \\ &\leq \int_s^t \|\nabla(\log \rho_u)(X_u)\|_{L^2}\, ds.\end{aligned}$$

But note that

$$\|\nabla(\log \rho_u)(X_u)\|_{L^2}^2 = \mathcal{I}(\mu_u\|m).$$

By the log-Sobolev inequality of Lemma 12.2.3, we get

$$\mathcal{I}(\mu_u\|m)^{1/2} \leq \frac{\mathcal{I}(\mu_u\|m)}{\sqrt{-2ch(\mu_u\|m)}} = -\frac{d}{du}\sqrt{\frac{2}{c}h(\mu_u\|m)}.$$

Therefore,

$$\|X_t - X_s\|_{L^2} \leq \sqrt{-\frac{2}{c}h(\mu_s\|m)} - \sqrt{-\frac{2}{c}h(\mu_t\|m)}.$$

As remarked in the last section, since m satisfies the log-Sobolev inequality, we have $h(\mu_t\|m) \to 0$ as $t \to \infty$ since $\mathcal{I}(\mu_t\|m) \to 0$. But this implies that $(X_t)_{t\geq\infty}$ is Cauchy in L^2 as $t \to +\infty$. So there exists some variable X_∞ with $X_t \to X_\infty$ in L^2.

Now we claim that $X_\infty \sim m$. Using the Csiszár-Kullback-Pinsker inequality (see Lemma 12.3.6 below),

$$\|\rho_t - 1\|_{L^1(m)} \leq \sqrt{-2h(\mu_t\|m)}.$$

So $\rho_t \to 1$ in $L^1(m)$. Thus, for any $f \in C_c^\infty(\mathbb{R}^d)$, we have

$$E[f(X_\infty)] = \lim_{t\to+\infty} E[f(X_t)] = \lim_{t\to+\infty} \int f \rho_t\, dm = \int f\, dm.$$

Therefore, $X_\infty \sim m$, and hence

$$W_2(\mu, m) \leq \|X_0 - X_\infty\|_{L^2} \leq \sqrt{-\frac{2}{c}h(\mu\|m)}. \qquad \square$$

Lemma 12.3.6 (Csiszár-Kullback-Pinsker)**.** *Let m be a probability measure on $\mathbb{R}^d$ and let $d\mu = \rho\, dm$. Then*

$$\|\rho - 1\|_{L^1(m)} \leq 2\sqrt{-h(\mu\|m)}$$

Proof. Because $\int \rho\, dm = 1$, we have

$$\int_{\rho\le 1} (\rho - 1)\, dm = -\int_{\rho\ge 1} (\rho - 1)\, dm,$$

and hence

$$\|\rho - 1\|_{L^1(m)} = 2\int_{\rho\le 1} (1-\rho)\, dm \le 2\left(\int_{\rho\le 1} (1-\rho)^2\, dm\right)^{1/2}.$$

Note that $f(\rho) = \rho\log\rho$ is convex with $f'(1) = 1$ and $f''(\rho) = 1/\rho$. Hence, for $\rho \le 1$, we have $f''(\rho) \ge 1$, so that

$$\rho\log\rho \ge (\rho - 1) + (\rho - 1)^2.$$

Thus,

$$\begin{aligned}
\int_{\rho\le 1} (\rho-1)^2\, dm &\le \int_{\rho\le 1} [\rho\log\rho - (\rho - 1)]\, dm \\
&\le \int [\rho\log\rho - (\rho - 1)]\, dm \\
&= \int \rho\log\rho\, dm \\
&= -h(\mu\|m),
\end{aligned}$$

since $\rho\log\rho - (\rho - 1) \ge 0$ by convexity and $\int (\rho - 1)\, dm = 0$. $\square$

12.4 Conditional entropy, Fisher's information, and transport

All the concepts and results we discussed so far in the chapter have analogues in the conditional setting. Suppose that m is a probability measure on $\mathbb{R}^{d_1+d_2}$ which has a disintegration

$$dm(x_1, x_2) = dm_2(x_2|x_1)dm_1(x_1)$$

where $m_2(\cdot|x_1)$ is a Borel measure for each x_1 which depends Borel-measurably on x_1. Given a probability measure $d\mu = \rho\, dm$, the density can be expressed as

$$\rho(x_1, x_2) = \rho_2(x_2|x_1)\rho_1(x_1),$$

where $\rho_1(x_1) = \int \rho(x_1, x_2)\, dm_2(x_2|x_1)$, which is finite almost everywhere. Then writing $\log\rho(x_1, x_2) = \log\rho_2(x_2|x_1) + \log\rho_1(x_1)$, we obtain the relation

$$-\int \rho\log\rho\, dm = -\int \rho_1\log\rho_1\, dm_1 - \int \left(\int \rho_2(x_2|x_1)\log\rho_2(x_2|x_1)\, dm_2(x_2|x_1)\right)\rho_1(x_1)\, dm_1(x_1).$$

In other words, if we let $d\mu_1 = \rho_1\, dm_1$ and $d\mu_2(x_2|x_1) = \rho_2(x_2|x_1)\, dm_2(x_2|x_1)$, then we have

$$h(\mu\|m) = h(\mu_1\|m_1) + \int h\Big(\mu_2(\cdot|x_1)\|m_2(\cdot|x_1)\Big)\, d\mu_1(x_1),$$

which is known as *additivity of entropy under conditioning.*

Furthermore, if for each x_1, the measures $\mu_2(\cdot|x_1)$ and $m_2(\cdot|x_1)$ are given by densities satisfying sufficient smoothness and decay conditions, then we can define measures $\mu_{2,t}(\cdot|x_1)$ for $t \geq 0$ that evolve according to the heat equation with respect to the density of $m_2(\cdot|x_1)$. We will then obtain for $s \leq t$,

$$\int h\Big(\mu_{2,t}(\cdot|x_1)\|m_2(\cdot|x_1)\Big)\,d\mu_1(x_1) - h\Big(\mu_{2,s}(\cdot|x_1)\|m_2(\cdot|x_1)\Big)\,d\mu_1(x_1)$$
$$= \int_s^t \int \mathcal{I}\Big(\mu_{2,u}(\cdot|x_1)\|m_2(\cdot|x_1)\Big)\,d\mu_1(x_1)\,du,$$

or roughly speaking, conditional entropy is the integral of conditional Fisher information.

Let $d\mu_t(x_1,x_2) = d\mu_{2,t}(x_2|x_1)\,d\mu_1(x_1)$. Lemma 12.3.1 adapts to the conditional setting as follows. Let $\rho_{2,t}(x_2|x_1)$ be the conditional density of $\mu_{2,t}(\cdot|x_1)$. Let $V_{2,t}(x_1,x_2) = -\log \rho_{2,t}(x_2|x_1)$. Then we can define $F_{s,t} : \mathbb{R}^{d_1+d_2} \to \mathbb{R}^{d_2}$ by

$$\partial_s F_{s,t}(x_1,x_2) = \nabla_{x_2} V_{2,s}(x_1, F_{s,t}(x_1,x_2)), \qquad F_{t,t}(x_1,x_2) = x_2.$$

In fact, letting $V_t(x_1,x_2) = -\log \rho_{2,t}(x_2|x_1) - \log \rho_1(x_1)$ be minus the log-density of μ_t, we have $\nabla_{x_2} V_{2,t}(x_1,x_2) = \nabla_{x_2} V_t(x_1,x_2)$, and thus the above ODE can be expressed purely in terms of the overall density ρ_t as

$$\partial_s F_{s,t}(x_1,x_2) = \nabla_{x_2} V_s(x_1, F_{s,t}(x_1,x_2)), \qquad F_{t,t}(x_1,x_2) = x_2.$$

If we apply Lemma 12.3.1 to $\mu_{2,t}(\cdot|x_1)$ for each x_1, we see that the function $F_{s,t}(x_1,\cdot)$ pushes forward $\mu_{2,t}(\cdot|x_1)$ to $\mu_{2,s}(\cdot|x_1)$ for each x_1. Therefore, when we consider the fibers all together, the function $(x_1, F_{s,t}(x_1,x_2))$ pushes forward μ_t to μ_s.

To frame the whole discussion in more probabilistic terms, let $(X_1, X_{2,t}) \sim \mu_t$. Then $(X_1, F_{s,t}(X_1, X_{2,t})) \sim (X_1, X_{2,s})$. Thus, $F_{s,t}$ can be viewed as a *transport from* $X_{2,t}$ *to* $X_{2,s}$ *conditioned on* X_1.

The above considerations naturally lead to conditional versions of the log-Sobolev and Talagrand inequalities, as well as conditional score functions. But we will not discuss any of these in detail except in the case where the measure m is Lebesgue measure or Gaussian.

For future reference, let us give the precise formulations of conditional entropy and Fisher information when the background measure is Lebesgue measure.

Definition 12.4.1. Let (X,Y) be a random variable in $\mathbb{R}^{d_1+d_2}$. Suppose that the distribution of (X,Y) has a disintegration as $\rho(x|y)\,dx\,d\nu(y)$ for some probability measure ν. Then we define the *conditional entropy*

$$h(X|Y) = -\int \rho(x|y)\log\rho(x|y)\,dx\,d\nu(y).$$

If the law of (X,Y) does not admit such a disintegration, then we define $h(X|Y) = -\infty$.

Definition 12.4.2. If (X, Y) is a random variable in $\mathbb{R}^{d_1+d_2}$, then a d_1-tuple Ξ is a *score function for X given Y* if for any $f \in C_c^\infty(\mathbb{R}^{d_1+d_2})$, we have

$$E[\Xi f(X, Y)] = E[\nabla_x f(X, Y)].$$

If a score function for X given Y exists, then we define the *conditional Fisher information* by

$$\mathcal{I}(X|Y) = \|\Xi\|_{L^2}^2,$$

and otherwise the Fisher information is defined to be ∞.

Lemma 12.4.3. *Let Y be a random variable in $\mathbb{R}^{d_2}$, and let X and Z be random variables in $\mathbb{R}^{d_1}$ which are conditionally independent given Y. If Ξ is a score function for X given Y, then $E[\Xi|X + Z, Y]$ is a score function for $X + Z$ given Y.*

The proof is the same as that of Lemma 12.1.3, so we leave it as an exercise. The next lemma is the conditional version of Lemma 12.1.4.

Lemma 12.4.4. *Let (X, Y) be a random variable in $\mathbb{R}^{d_1+d_2}$ whose distribution disintegrates as $\rho(x|y)\,dx\,d\nu(y)$, and such that $a := (1/d_1)E\|X\|^2 < \infty$. Let Z_t be an independent Gaussian $\mathbb{R}^{d_1}$-tuple with covariance matrix tI. Then*

$$\frac{d_1}{a+t} \leq \mathcal{I}(X + Z_t|Y) \leq \min\left(\frac{d_1}{t}, \mathcal{I}(X|Y)\right)$$

and

$$h(X|Y) = \frac{1}{2}\int_0^\infty \left(\frac{d_1}{1+t} - \mathcal{I}(X + Z_t|Y)\right) dt + \frac{d_1}{2}\log 2\pi e.$$

Proof. The first claim follows from the same argument as in Lemma 12.1.4.

To prove the second claim, recall that for $t > 0$, the variable Z_t/t is a score function for Z_t given Y can be defined by Z_t/t. Therefore, by the previous lemma, the score function Ξ_t for $X + Z_t$ given Y is $E[Z_t/t|X + Z_t, Y]$. This conditional expectation can be evaluated as $-\nabla_x \log \rho_t(x|y)$, where $\rho_t(x|y)$ evolves from $\rho(x|y)$ according to the heat equation. Indeed,

$$-\nabla_x \log \rho_t(x|y) = \frac{\int (z/t)\rho(x - z|y)e^{-\|z\|^2/2t}\,dz}{\int \rho(x - z|y)e^{-\|z\|^2/2t}\,dz},$$

which is exactly the conditional expectation since the density of $(X + Z_t, Z_t)$ given Y is $\rho(x - z, z)e^{-\|z\|^2/2t}$. By fixing y and applying Lemma 12.1.4, we obtain

$$-\int \rho_t(x|y)\log \rho_t(x|y)\,dx + \int \rho(x|y)\log \rho(x|y)\,dx = \int_0^t \int \|\nabla_x \log \rho_s(x, y)\|^2\,dx\,ds.$$

By Lemma 12.1.2, the first integral is bounded below by $h(Z_t|Y) = h(Z_t)$. Hence, if we integrate with respect to y, the first term on the left-hand side is a bounded function of y

and the second term is bounded above by $E[X^2|Y]$. Everything on the right-hand side is nonnegative. Thus, we may integrate the above equation with respect to Y and obtain

$$h(X + Z_t|Y) - h(X|Y) = \int_0^t \mathcal{I}(X + Z_s|Y)\, ds.$$

Then to derive the formula for $h(X|Y)$ we use the same reasoning as in Lemma 12.1.4. For instance, to control the asymptotic behavior of $h(X+Z_t|Y)$ as $t \to \infty$, we apply Lemma 12.1.2 to the conditional distribution of $X + Z_t$ given Y, which results in

$$\begin{aligned} h(Z_t) = h(Z_t|Y) &\leq h(X + Z_t|Y) \\ &\leq E\left(\frac{d_1}{2} \log \frac{2\pi e E[(X + Z_t)^2|Y]}{d_1}\right) \\ &\leq \frac{d_1}{2} \log \frac{2\pi e E[(X + Z_t)^2]}{d_1}, \end{aligned}$$

where the last line follows from Jensen's inequality and concavity of the logarithm function. We leave the rest as an exercise. □

12.5 Entropy and transport relative to Gaussian measure

There is a special relationship between Lebesgue measure and Gaussian measure in this whole story of entropy and transport. The entropy, Fisher information, heat evolution, and transport functions associated to the Gaussian measure can all be obtained from those associated to Lebesgue measure by a relatively simple change of variables, which will be used in §17 and which we will now explain.

Let μ be a measure on $\mathbb{R}^d$ with density ρ. Let γ_t be the measure on $\mathbb{R}^d$ with Gaussian density $(2\pi t)^{-d/2}e^{-\|x\|^2/2t}$. Then the density of μ with respect to $\gamma_{1/d}$ is of course $\tilde{\rho}(x) = \rho(x)(2\pi)^{d/2}e^{d\|x\|^2/2}$.

Let us denote the entropy with respect to the Gaussian measure as

$$h_g(\mu) := h(\mu\|\gamma_1).$$

Furthermore, to prepare for the more probabilistic notation we will use in the paper, if X is a random variable $\sim \mu$, we denote $h(X) := h(\mu)$ and $h_g(X) := h_g(\mu)$.

Then we have

$$\begin{aligned} h_g(X) &= -\int \tilde{\rho} \log \tilde{\rho} \, d\gamma_1 \\ &= -\int \rho(x)(2\pi)^{d/2} e^{\|x\|^2/2} \log \left(\rho(x)(2\pi)^{d/2} e^{\|x\|^2/2} \right) (2\pi)^{-d/2} e^{-\|x\|^2/2} \, dx \\ &= -\int \rho(x) \log \rho(x) \, dx - \int \left(\frac{\|x\|^2}{2} + \frac{d}{2} \log 2\pi \right) \rho(x) \, dx \\ &= h(X) - \frac{1}{2} E\|X\|^2 - \frac{d}{2} \log 2\pi. \end{aligned}$$

Furthermore, the Fisher information with respect to Lebesgue measure and Gaussian measure are related as follows. Let

$$\mathcal{I}_g(X) := \mathcal{I}_g(\mu) := \int \|\nabla \tilde{\rho} / \tilde{\rho}\|^2 \, d\gamma_1,$$

and denote $\mathcal{I}(X) := \mathcal{I}(\mu)$. Then

$$-\nabla \log \tilde{\rho}(x) = -\nabla \log \rho(x) + \nabla(\|x\|^2/2) = -\nabla(\log \rho(x)) - x.$$

Hence,

$$\begin{aligned} \mathcal{I}_g(X) &= \int \|\nabla(\log \rho(x)) + x\|^2 \, d\mu(x) \\ &= \int \|\nabla(\log \rho(x))\|^2 \, d\mu(x) + 2 \int \langle \nabla(\log \rho(x)), x \rangle \, d\mu(x) + \int \|x\|^2 \, d\mu(x) \\ &= \mathcal{I}(X) - 2d + E\|X\|^2, \end{aligned}$$

where the middle term has been evaluated using integration by parts.

Now let ρ_t evolve according to the flat heat equation $\partial_t \rho_t = (1/2)\Delta \rho_t$, and let μ_t be the corresponding measure. Recall that if $X \sim \mu$ and if $Z_t \sim \gamma_t$ is independent of X, then $X + Z_t \sim \mu_t$. (Here we include a factor of 1/2 because it is convenient for the probabilistic viewpoint. It does not matter for our present purpose how the Z_t's relate to each other for different values of t, but one could for instance take them to be a Brownian motion.)

Now let $\tilde{\mu}_t$ be the law of $e^{-t/2} X + e^{-t/2} Z_{e^t - 1} \sim e^{-t/2} X + (1 - e^{-t})^{1/2} Z_1$. Thus, the density of $\tilde{\mu}_t$ with respect to Lebesgue measure is

$$e^{td/2} \rho_{e^t - 1}(e^{t/2} x),$$

and hence the density with respect to γ_1 is

$$\tilde{\rho}_t(x) := (2\pi)^{d/2} e^{\|x\|^2/2} e^{td/2} \rho_{e^t - 1}(e^{t/2} x).$$

Then a direct computation shows that $\tilde{\rho}_t$ evolves according to the heat equation relative to the Gaussian density, which is

$$\partial_t \tilde{\rho}_t(x) = \frac{1}{2}(\Delta \tilde{\rho}_t(x) - \langle x, \nabla \tilde{\rho}_t(x) \rangle), \tag{12.9}$$

where we have observed that the gradient of the log-density of the Gaussian is x. The normalizing factor of $1/2$ comes from our choice to use $(1/2)\Delta$ in the flat heat equation, which amounts to a reparametrization of time by a factor of $1/2$. To check that (12.9) holds, observe that

$$\partial_t \tilde{\rho}_t(x) = (2\pi)^{d/2} e^{\|x\|^2/2} \left(\frac{d}{2} e^{td/2} \rho_{e^t-1}(e^{t/2}x) + e^{td/2} e^t \frac{1}{2} \Delta \rho_t(e^{t/2}x) + e^{td/2} \frac{1}{2} e^{t/2} \langle x, \nabla \rho_{e^t-1}(e^{t/2}x) \rangle \right).$$

Meanwhile,

$$\nabla \tilde{\rho}_t(x) = (2\pi)^{d/2} e^{td/2} \left(x e^{\|x\|^2/2} \rho_{e^t-1}(e^{t/2}x) + e^{\|x\|^2/2} e^{t/2} \nabla \rho_{e^t-1}(e^{t/2}x) \right)$$

and

$$\begin{aligned} \frac{1}{2}\Delta \tilde{\rho}_t(x) = \frac{1}{2}(2\pi)^{d/2} e^{td/2} \Big(& d e^{\|x\|^2/2} \rho_{e^t-1}(e^{t/2}(x)) + \|x\|^2 e^{\|x\|^2/2} \rho_{e^t-1}(e^{t/2}x) \\ & + 2\langle x e^{\|x\|^2/2}, e^{t/2} \nabla \rho_{e^t-1}(e^{t/2}x) \rangle + e^{\|x\|^2/2} e^t \Delta \rho_{e^t-1}(e^{t/2}x) \Big). \end{aligned}$$

By comparing all the terms together, (12.9) is verified. In short, the heat semigroup for the Laplacian with respect to Gaussian measure is just a reparametrization of the flat heat semigroup.

This reparametrization explains how the expressions for entropy in terms of Fisher's information for the Lebesgue and Gaussian measures are related. Recall that

$$\begin{aligned} h_g(X) &= h(X) - E\|X\|^2 - \frac{d}{2} \log 2\pi \\ &= \frac{1}{2} \int_0^\infty \left(\frac{d}{1+s} - \mathcal{I}(X + Z_s) \right) ds - \|X\|^2 + \frac{d}{2}. \end{aligned}$$

But we also know that

$$h_g(X) = -\frac{1}{2} \int_0^\infty \mathcal{I}_g(e^{-u/2} X + e^{-u/2} Z_{e^u-1}) \, du.$$

By substituting $u = e^s - 1$ in the integral and using the scaling behavior of score functions, this expression for $h_g(X)$ can be shown to be equivalent to the first one. We leave the details as an exercise.

The same change of variables relates the transport functions given by Lemma 12.3.1 for the Lebesgue measure and for the Gaussian measure. Indeed, let $F_{s,t}$ be the transport function that pushes forward μ_t to μ_s, given by

$$\partial_s F_{s,t}(x) = -\frac{1}{2}\nabla(\log \rho_s) \circ F_{s,t}(x), \qquad F_{t,t}(x) = x.$$

Let

$$\tilde{F}_{s,t}(x) = e^{-s/2} F_{e^s-1,e^t-1}(e^{t/2}x).$$

Using our random variables X and Z_t again, we have $\tilde{F}_{s,t}(e^{-t/2}X + e^{-t/2}Z_{e^t-1}) \sim e^{-s/2}X + e^{-s/2}Z_{e^s-1}$, which shows that $\tilde{F}_{s,t}$ pushes forward $\tilde{\mu}_t$ to $\tilde{\mu}_s$. In fact, $\tilde{F}_{s,t}$ is exactly the same transport function as constructed by Lemma 12.3.1 for the Gaussian measure, since

$$\begin{aligned}\partial_s \tilde{F}_{s,t}(x) &= \frac{1}{2}e^{-s/2} \cdot e^s(-\nabla(\log \rho_{e^s-1})) \circ F_{e^s-1,e^t-1}(e^{t/2}x) - \frac{1}{2}e^{-s/2}F_{e^s-1,e^t-1}(e^{t/2}x) \\ &= (-\nabla(\log \tilde{\rho}_s)) \circ \tilde{F}_{s,t},\end{aligned}$$

since

$$-\nabla \log \tilde{\rho}_s(x) = -e^{s/2}\nabla(\log \rho_{e^s-1})(e^{s/2}x) - x.$$

Furthermore, these lemmas adapt easily to the conditional setting.

CHAPTER 13

Tools: Functions of non-commuting real variables

This chapter reviews the properties of the algebra TrP_d of trace polynomials in d self-adjoint variables. We then define a certain completion of TrP_d, which we denote by $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. An element of $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ represents a function f that can be evaluated on any tuple $x = (x_1, \dots, x_d)$ of self-adjoint elements in a tracial W^*-algebra such that λ_x is approximated by the non-commutative laws of matrix tuples (or equivalently, such that $\mathrm{W}^*(x)$ embeds into the ultraproduct of the hyperfinite II_1 factor). The output $f(x)$ will be an element of $L^2(\mathrm{W}^*(x))$, and amazingly every element of $L^2(\mathrm{W}^*(x))$ can be realized in this fashion (see Proposition 13.6.6). The functions from $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ are closed under certain algebraic and composition operations. Moreover, they are a natural tool to describe the large-n limit of functions on $M_n(\mathbb{C})_{\mathrm{sa}}^d$, which we will apply in the rest of the paper. This chapter is based on [Jek19, §3] and also draws on [HJN19, §2].

13.1 The $*$-algebra of trace polynomials

Trace polynomials have been used by many previous authors. They were first studied from an algebraic viewpoint since the give all the unitarily invariant polynomials over $n \times n$ matrices for every n [Raz74, Pro76, Ler76, Raz87]. Their applications to non-commutative probability, and in particular their relationship with matrix Brownian motion, have been studied in [Rai97, Sen08, Ceb13, DHK13, Kem16, Kem17, DGS16].

Roughly speaking, we think of a trace polynomial as a function obtained by multiplying non-commutative polynomials and traces of non-commutative polynomials. However, let us first give a formal algebraic definition.

Definition 13.1.1. Let $\mathbb{C}\langle X_1, \dots, X_d\rangle$ be the non-commutative polynomial algebra in indeterminates $X_1, \dots, X_d$. We equip $\mathbb{C}\langle X_1, \dots, X_d\rangle$ with the unique $*$-algebra structure such that $X_j = X_j^*$. Then $\mathbb{C}\langle X_1, \dots, X_d\rangle$ has a vector space basis consisting of the non-commutative monomials $X_{i(1)} \dots X_{i(k)}$ for $k \geq 0$ and $i(1), \dots, i(k) \in \{1, \dots, d\}$, and we call k the *degree* of the monomial. We say that two degree-k mononomials $X_{i(1)}, \dots, X_{i(k)}$ and $X_{i'(1)}, \dots, X_{i'(k)}$ are *cyclically equivalent* if we have $i'(j) = i(\sigma(j))$ for some permutation σ in the group generated by the cyclic permutation $1 \mapsto 2 \mapsto \dots \mapsto k \mapsto 1$.

Definition 13.1.2. We define TrP_d^0 as the commutative polynomial algebra in the indeterminates $\mathrm{tr}(p)$ for a non-commutative mononomials $p \in \mathbb{C}\langle X_1, \dots, X_d\rangle$ of degree $k \geq 1$, modulo the ideal generated by $\mathrm{tr}(p) - \mathrm{tr}(q)$ for p and q that are cyclically equivalent. We equip TrP_d^0 with the unique $*$-structure satisfying $\mathrm{tr}(p)^* = \mathrm{tr}(p^*)$ for a monomial p, which is well-defined because if p and q are cyclically equivalent, then so are p^* and q^*.

Clearly, an equivalent definition would be as follows. Let $\mathcal{P}$ be a collection of non-commutative monomials, one from each cyclic equivalence class. Then TrP_d^0 is isomorphic to the commutative polynomial algebra in indeterminates $(\mathrm{tr}(p))_{p\in\mathcal{P}}$.

Here tr is merely a formal symbol used as notation for the variables $\mathrm{tr}(p)$ generating the TrP_d^0. Although we initially only used the notation $\mathrm{tr}(p)$ when p is a monomial of degree at least one, the map $p \mapsto \mathrm{tr}(p)$ extends to a unique linear map $\mathbb{C}\langle X_1, \dots, X_d\rangle \to \mathrm{TrP}_d^0$ such that $1 \mapsto 1$. For a general polynomial p, we denote the application of this map by $\mathrm{tr}(p)$. Thus, $\mathrm{tr}(p+q) = \mathrm{tr}(p) + \mathrm{tr}(q)$ and $\mathrm{tr}(p^*) = \mathrm{tr}(p)^*$ for general non-commutative polynomials p and q.

Definition 13.1.3. TrP_d is defined to be $\mathrm{TrP}_d^0 \otimes \mathbb{C}\langle X_1, \dots, X_d\rangle$, as a tensor product of $*$-algebras. Furthermore, to simplify notation, we identify TrP_d^0 and $\mathbb{C}\langle X_1, \dots, X_d\rangle$ with the subalgebra $\mathrm{TrP}_d^0 \otimes 1$ and $1 \otimes \mathbb{C}\langle X_1, \dots, X_d\rangle$ respectively. Thus, for instance, we may express TrP_d as the span of terms of the form $\mathrm{tr}(p_1) \dots \mathrm{tr}(p_\ell) p_0$ where $\ell \geq 0$ and p_1, …, p_ℓ are non-commutative monomials of degree at least one, and p_0 is another non-commutative monomial.

Definition 13.1.4. We equip TrP_d^0 and TrP_d with gradings as follows. Let $\mathbb{C}\langle X_1, \dots, X_d\rangle_k$ denote the span of the monomials of degree k, and let $\mathrm{tr}(\mathbb{C}\langle X_1, \dots, X_d\rangle_k)$ denote their images in TrP_d^0. The 0th graded component of TrP_d^0 is $\mathbb{C}$. For $k \geq 1$, we define the kth graded component of TrP_d^0 as

$$\sum_{\ell=1}^{k} \sum_{\substack{k_1,\dots,k_\ell \geq 1: \\ k_1+\dots+k_\ell = k}} \mathrm{tr}(\mathbb{C}\langle X_1, \dots, X_d\rangle_{k_1}) \dots \mathrm{tr}(\mathbb{C}\langle X_1, \dots, X_d\rangle_{k_\ell}).$$

Similarly, TrP_d is equipped with the tensor product grading from TrP_d^0 and $\mathbb{C}\langle X_1, \dots, X_d\rangle$, which implies that the kth graded component is

$$\sum_{\ell=1}^{k} \sum_{\substack{k_0 \geq 0; k_1,\dots,k_\ell \geq 1: \\ k_0+k_1+\dots+k_\ell = k}} \mathrm{tr}(\mathbb{C}\langle X_1, \dots, X_d\rangle_{k_1}) \dots \mathrm{tr}(\mathbb{C}\langle X_1, \dots, X_d\rangle_{k_\ell}) \mathbb{C}\langle X_1, \dots, X_d\rangle_{k_0}.$$

Note that the $*$-operation respects both these gradings. We refer to these gradings as the (total) *degree*, and the *degree* of a trace polynomial is defined to be the highest graded component in which this polynomial has a non-zero component.

Definition 13.1.5. We define the *trace map* $\mathrm{TrP}_d \to \mathrm{TrP}_d^0 \subseteq \mathrm{TrP}_d$ as the unique linear map T such that

$$T(\mathrm{tr}(p_1) \dots \mathrm{tr}(p_\ell) p_0) = \mathrm{tr}(p_1) \dots \mathrm{tr}(p_\ell) \, \mathrm{tr}(p_0)$$

for monomials p_0, …, p_ℓ. We also use the notation $\mathrm{tr}(f) = T(f)$. Note that T respects the $*$-operation and the grading.

Now let us define the evaluation of trace polynomials on matrix tuples and more generally self-adjoint tuples from a W*-algebra. As an example, let $x = (x_1, \dots, x_d)$ be a d-tuple of self-adjoint $n \times n$ matrices. If $f = \operatorname{tr}(p_1) \dots \operatorname{tr}(p_\ell) p_0$ is an element of TrP_d, then we define

$$f(x) = \tau_n(p_1(x)) \dots \tau_n(p_\ell(x)) p_0(x).$$

In other words, $f(x)$ is defined by substituting $x_1, \dots, x_d$ for the formal variables $X_1, \dots, X_d$ and the matrix trace τ_n for the formal symbol tr. If $p_0, \dots, p_\ell$ are monomials such that f has total degree k, then $f(tx) = t^k f(x)$. Thus, $x \mapsto f(x)$ can be expressed as a polynomial of degree k in the matrix entries $((x_j)_{k,\ell} : j = 1, \dots, d; k, \ell = 1, \dots, n)$. The general definition of the evaluation map is as follows.

Definition 13.1.6. Let $(\mathcal{M}, \tau)$ be a tracial W*-algebra, and let $x = (x_1, \dots, x_d) \in \mathcal{M}_{\mathrm{sa}}^d$. Then we define $\operatorname{ev}_{(\mathcal{M},\tau),x} : \operatorname{TrP}_d \to \mathcal{M}$ as the unique $*$-homomorphism satisfying

$$\operatorname{ev}_{\mathcal{M},\tau,x}(p) = p(x)$$
$$\operatorname{ev}_{\mathcal{M},\tau,x}(\operatorname{tr}(p)) = \tau(p(x))1.$$

To see that this is well-defined, note that if p and q are cyclically equivalent monomials, then $\tau(p(x)) = \tau(q(x))$. Also, $\tau(p^*(x)) = \overline{\tau(p(x))}$. By the universal property of polynomial algebras, there is a unique $*$-homomorphism $\operatorname{TrP}_d^0 \to \mathbb{C}$ such that $\operatorname{tr}(p)$ is mapped to $\tau(p(x))$. Similarly, there is a unique $*$-homomorphism $\mathbb{C}\langle X_1, \dots, X_d\rangle \to \mathcal{M}$ sending X_j to x_j. Finally, we can take the tensor product of these two maps to obtain a map $\operatorname{TrP}_d^0 \otimes \mathbb{C}\langle X_1, \dots, X_d\rangle \to \mathbb{C} \otimes \mathcal{M} = \mathcal{M}$.

For the most part, we will use more compact (but less precise) notation and write

$$f(x) := \operatorname{ev}_{(\mathcal{M},\tau),x}(f) \text{ for } f \in \operatorname{TrP}_d.$$

The ambient W*-algebra $\mathcal{M}$ and its trace τ are suppressed in the notation. It is clear that $f(x)$ only depends on x and the restriction of τ to the algebra generated by x (which is equivalent to the non-commutative law λ_x). Thus, the ambient trace *does* matter for the definition of $f(x)$ even though $\mathcal{M}$ itself is not important. Thus, the reader must bear in mind that the notation $f(x)$ implicitly assumes a particular law for x.

13.2 Construction of $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$

From an analytic viewpoint, we prefer to work with certain separation-completion of TrP_d called $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ (the rationale for this notation will be explained later). In [Jek18, §8.1] and [Jek19], we sketched several equivalent ways of defining these separation-completions. Here we take a direct and elementary approach.

Definition 13.2.1. For $f \in \operatorname{TrP}_d$, we define

$$\|f\|_{2,R} = \sup_n \sup\{\|\operatorname{ev}_{M_n(\mathbb{C}),\tau_n,x}(f)\|_2 : x \in M_n(\mathbb{C})_{\mathrm{sa}}^d, \|x\|_\infty \le R\}.$$

Definition 13.2.2. Define $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ to be the separation-completion of TrP_d with respect to the family of seminorms $(\|\cdot\|_{2,R})_{R>0}$.

The separation-completion is a standard construction in topological vector spaces, but let us explain it in more detail. The separation-completion of TrP_d is the set of equivalence classes of sequences $(f_k)_{k\in\mathbb{N}}$ that are Cauchy with respect to each of these seminorms, where $(f_k)_{k\in\mathbb{N}} \sim (g_k)_{k\in\mathbb{N}}$ if $\|f_k - g_k\|_{2,R} \to 0$ for every R. The separation-completion has a vector space structure and $\|\cdot\|_{2,R}$ is defined on the separation-completion as the limit of $\|f_k\|_{2,R}$ over the Cauchy sequence $(f_k)_{k\in\mathbb{N}}$.

There is a canonical map $\mathrm{TrP}_d \to C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ sending f to the constant sequence $(f)_{k\in\mathbb{N}}$, and this map is linear and isometric with respect to $\|\cdot\|_{2,R}$. It is true, but irrelevant for our purposes, that this map is injective; this follows from [Pro76, Corollary 4.4].

Furthermore, $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ has a natural topology generated by the family of seminorms $(\|\cdot\|_{2,R})_{R>0}$. Since $\|f\|_{2,R}$ is increasing in R, we can obtain the topology equivalently by only using the seminorms associated to $R \in \mathbb{N}$. Since the topology is generated by a countable family of seminorms, this makes $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ a *Fréchet space*; for background on Fréchet spaces, see e.g. [Fol99, §5.4].

Finally, it is clear that the $*$-operation and trace map on TrP_d pass to well-defined maps on $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. This is because $|\tau(x)| \leq \|x\|_2$ and $\|x^*\|_2 = \|x\|_2$ for an operator X in a tracial W*-algebra, and hence $\|\mathrm{tr}(f)\|_{2,R} \leq \|f\|_{2,R} = \|f^*\|_{2,R}$. Our goal is to extend the evaluation maps $\mathrm{ev}_{\mathcal{M},\tau,x}$ from TrP_d to $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$, whenever x is a self-adjoint d-tuple that can be approximated in non-commutative law by matrix tuples.

Definition 13.2.3. We define

$$\Sigma_{d,R}^{\mathrm{app}} = \overline{\bigcup_{n\in\mathbb{N}} \{\lambda_x : x \in M_n(\mathbb{C})_{\mathrm{sa}}^d, \|x\|_\infty \leq R\}} \subseteq \Sigma_{d,R}.$$

Remark 13.2.4. It is a standard fact in the theory of tracial von Neumann algebras that $\lambda_x \in \Sigma_{d,R}^{\mathrm{app}}$ if and only if $\mathrm{W}^*(X)$ embeds (in a trace-preserving way) into the ultraproduct $\prod_{n\to\omega} M_n(\mathbb{C})$, and this occurs if and only if $\mathrm{W}^*(X)$ embeds into the ultrapower $\mathcal{R}^\omega$, where $\mathcal{R}$ is the hyperfinite II_1 factor. We will not pursue the ultraproduct viewpoint in this exposition, but the reader may refer to [Cap10] for background, and [Jek19, §3] for explanation of these function spaces in terms of ultraproducts.

Lemma 13.2.5. *Suppose that $x^{(k)}$ is a self-adjoint d-tuple from $(\mathcal{M}^{(k)}, \tau^{(k)})$ with $\|x^{(k)}\|_\infty \leq R$, and suppose that x is a self-adjoint d-tuple from $(\mathcal{M}, \tau)$ with $\|x\|_\infty \leq R$. The following are equivalent.*

(1) $x^{(k)}$ converges in non-commutative law to x.

(2) $f(x^{(k)}) \to f(x)$ for every $f \in \mathrm{TrP}_d^0$.

(3) $\langle f(x^{(k)}), g(x^{(k)})\rangle_2 \to \langle f(x), g(x)\rangle_2$ for all $f, g \in \mathrm{TrP}_d$.

Proof. (1) $\implies$ (2). It suffices to prove the claim when $f = \operatorname{tr}(p_1)\dots\operatorname{tr}(p_\ell)$ for monomials $p_1, \dots, p_\ell$. But convergence in law implies by definition that $\tau^{(k)}(p_j(x^{(k)})) \to \tau(p(x))$ as $k \to \infty$.

(2) $\implies$ (3). By linearity, we may reduce to the case where

$$f = \operatorname{tr}(p_1)\dots\operatorname{tr}(p_\ell)p_0, \qquad g = \operatorname{tr}(q_1)\dots\operatorname{tr}(q_m)q_0.$$

If we evaluate f and g on a d-tuple X and take the inner product, that is the same as evaluating the function

$$h = \operatorname{tr}(p_1^*)\dots\operatorname{tr}(p_\ell^*)\operatorname{tr}(p_0^*q_0)\operatorname{tr}(q_1)\dots\operatorname{tr}(q_m).$$

But $h(x^{(k)}) \to h(x)$ by (2).

(3) $\implies$ (1). If (3) holds, then for every non-commutative polynomial p, we have $\tau^{(k)}(p(x^{(k)})) = \langle 1, p(x^{(k)})\rangle_2 \to \langle 1, p(x)\rangle_2 = \tau(p(X))$, so $x^{(k)}$ converges in law to x. $\square$

Lemma 13.2.6. *If $f \in \operatorname{TrP}_d$, then*

$$\|f\|_{2,R} = \sup\{\|f(x)\|_2 : \lambda_x \in \Sigma_{d,R}^{\mathrm{app}}\},$$

where the set on the right-hand side is shorthand for the set of $\|f(x)\|_2$ for any tracial W^*-*algebra $(\mathcal{M}, \tau)$ and self-adjoint tuple x from $\mathcal{M}$ with $\lambda_x \in \Sigma_{d,R}^{\mathrm{app}}$ (which in particular requires that $\|x\|_\infty \le R$).*

Proof. The inequality $\le$ is easy because the law of any matrix tuple with $\|x\|_\infty \le R$ is included on the right-hand side. The inequality $\ge$ holds because for any x on the right-hand side, there is a sequence of matrix tuples that is bounded by R in $\|\cdot\|_\infty$ and converges in law to x. $\square$

Lemma 13.2.7. *Let x be a self-adjoint tuple from $(\mathcal{M}, \tau)$ with $\lambda_x \in \Sigma_{d,R}^{\mathrm{app}}$ for some $R > 0$. Then the map $\operatorname{ev}_{\mathcal{M},\tau,x} : \operatorname{TrP}_d \to \mathcal{M}$ extends uniquely to a continuous linear map $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2) \to L^2(\mathcal{M}, \tau)$. Moreover, the image is contained in $L^2(\mathrm{W}^*(x))$. Also,*

$$\operatorname{ev}_{\mathcal{M},\tau,x}(\operatorname{tr}(f)) = \tau(\operatorname{ev}_{\mathcal{M},\tau,x}(f)), \qquad \operatorname{ev}_{\mathcal{M},\tau,x}(f^*) = \operatorname{ev}_{\mathcal{M},\tau,x}(f)^*.$$

Proof. If $(f_k)_{k\in\mathbb{N}}$ is Cauchy with respect to $\|\cdot\|_{2,R}$, then by the previous corollary, $\operatorname{ev}_{\mathcal{M},\tau,x}(f_k)$ is Cauchy in $L^2(\mathcal{M}, \tau)$. Moreover, if two Cauchy sequences are equivalent, then the limits in $L^2(\mathcal{M}, \tau)$ are equal.

If $f \in \operatorname{TrP}_d$, then it is clear that $f(x)$ is in the algebra generated by x. Thus, if f is in the separation-completion, then $f(x)$ is in the $\|\cdot\|_2$-closure of the algebra generated by x, which is $L^2(\mathrm{W}^*(x))$.

The last two equalities obviously hold for trace polynomials and hence hold for all functions by approximation. $\square$

We will continue to abbreviate $\mathrm{ev}_{\mathcal{M},\tau,x}(f)$ as $f(x)$.

Lemma 13.2.8. *Let $x^{(k)}$ and x be d-tuples from tracial* W^**-algebras $(\mathcal{M}^{(k)}, \tau^{(k)})$ and $(\mathcal{M}, \tau)$ respectively whose laws are in $\Sigma_{d,R}^{\mathrm{app}}$. The following are equivalent:*

(1) $x^{(k)}$ converges to x in non-commutative law.

(2) $\tau^{(k)}(f(x^{(k)})) \to \tau(f(x))$ for every $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$.

(3) $\tau^{(k)}(f(x^{(k)})^* g(x^{(k)})) \to \tau(f(x)^* g(x))$ for every $f, g \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$.

Proof. (3) $\implies$ (2) is trivial because we can substitute 1 as one of the functions. (2) $\implies$ (1) follows from Lemma 13.2.5. Finally, if (1) holds, then (3) holds by trace polynomials by Lemma 13.2.5, and then (3) extends to all of $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ by a straightforward approximation argument. □

The notation $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ does not have a literal meaning, but rather it is based on analogy with the commutative setting. By the Stone-Weierstrass theorem, any continuous function on $\mathbb{R}^d$ can be expressed as the limit of a sequence of polynomials p_n which converge uniformly on $[-R, R]^d$ for every R. The norm $\|\cdot\|_{2,R}$ that we defined is loosely analogous to taking the supremum of the values of polynomial over $x \in [-R, R]^d$. Thus, the space that we defined is something like a non-commutative analogue of $C(\mathbb{R}^d)$. Thus, we write $\mathbb{R}^{*d}$ to evoke the idea of a "free product of d copies of $\mathbb{R}$."

We have added the term $\|\cdot\|_2$ to the notation to emphasize that the norm being used for the output $f(X)$ is the $\|\cdot\|_2$ rather than $\|\cdot\|_\infty$. We have added the subscript tr to denote the fact that we have taken the completion of trace polynomials rather than non-commutative polynomials. We have written the subscript app because we restrict our attention to tuples that can be approximated in law by matrix tuples, rather than all non-commutative laws.

Of course, the construction would still make perfect sense if we instead took the supremum of $\|f(x)\|_2$ over all d-tuples with $\|x\|_\infty \leq R$ rather than only those approximable in non-commutative law by matrices. However, in §13.5, we will use functions in $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ to describe the large-n limit of sequences of functions on matrices, and thus we want f to be uniquely determined by its evaluation on matrix tuples. It is also technically convenient in a number of places that to define a function f it suffices to give the values on matrices, but this is not essential.

13.3 Continuity properties and functional calculus

This section explores various continuity and boundedness properties of functions from the space $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. Then we explain how the space $C_{\mathrm{tr,app}}(\mathbb{R}^{*1}, \|\cdot\|_2)$ subsumes the one-variable functional calculus.

Definition 13.3.1. For $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$, we define

$$\|f\|_{\infty,R} = \sup_n \sup\{\|f(x)\|_\infty : x \in M_n(\mathbb{C})_{\mathrm{sa}}^d\}.$$

We say that f is $\|\cdot\|_\infty$ *bounded on* $\|\cdot\|_\infty$*-balls* if $\|f\|_{\infty,R} < \infty$ for every $R > 0$. We say that f is *globally* $\|\cdot\|_\infty$*-bounded* if $\sup_R \|f\|_{\infty,R} < \infty$.

Definition 13.3.2. Let $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. We say that f is $\|\cdot\|_2$*-uniformly continuous on the* $\|\cdot\|_2$*-ball of radius* R if for every $\epsilon > 0$, there exists a $\delta > 0$ such that for all n and all $x, y \in M_n(\mathbb{C})_{\mathrm{sa}}^d$ with $\|x\|_2, \|y\|_2 \leq R$, if $\|x - y\|_2 < \delta$, then $\|f(x) - f(y)\|_2 < \epsilon$.

We say that f is $\|\cdot\|_2$*-uniformly continuous on* $\|\cdot\|_2$*-balls* if this holds for every R.

We make the same definitions with $\|\cdot\|_2$-balls replaced by $\|\cdot\|_\infty$-balls, where we only replace $\|x\|_2, \|y\|_2 \leq R$ with $\|x\|_\infty, \|y\|_\infty \leq R$.

Finally, we say *globally* $\|\cdot\|_2$*-uniformly continuous* if the same $\delta(\epsilon)$ works for every R.

Definition 13.3.3. A function $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ is said to be *globally* $\|\cdot\|_2$*-Lipschitz* if there is a constant K such that $\|f(x) - f(y)\|_2 \leq K\|x - y\|_2$ for all $x, y \in M_n(\mathbb{C})_{\mathrm{sa}}^d$ for all n, and we write

$$\|f\|_{\mathrm{Lip}} = \sup_n \sup \left\{ \frac{\|f(x) - f(y)\|_2}{\|x - y\|_2} : x, y \in M_n(\mathbb{C})_{\mathrm{sa}}^d \right\}.$$

Moreover, we define *Lipschitz on* $\|\cdot\|_2$ *balls* and *Lipschitz on* $\|\cdot\|_\infty$*-balls* similarly to the previous definition. We denote by $\|f\|_{\mathrm{Lip},R}$ the Lipschitz constant on the $\|\cdot\|_\infty$-ball of radius R.

Lemma 13.3.4.

(1) Every $f \in \mathrm{TrP}_d$ *is* $\|\cdot\|_\infty$*-bounded and* $\|\cdot\|_2$*-Lipschitz on operator norm balls.*

(2) Every $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ *is* $\|\cdot\|_2$*-uniformly continuous on* $\|\cdot\|_\infty$*-balls.*

(3) If $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ *is* $\|\cdot\|_2$*-Lipschitz on* $\|\cdot\|_\infty$*-balls, then* f *is* $\|\cdot\|_\infty$*-bounded on* $\|\cdot\|_\infty$*-balls with*

$$\|f\|_{R,\infty} \leq \|f\|_{2,R} + 2\sqrt{2}d^{1/2}\Theta R\|f\|_{\mathrm{Lip},R},$$

where Θ *is the constant from Lemma 11.5.2.*

Proof. (1) By linearity, it suffices to consider the case where

$$f = \mathrm{tr}(p_1) \dots \mathrm{tr}(p_\ell) p_0,$$

where the p_j's are monomials. The $\|\cdot\|_\infty$-boundedness is clear since each term is $\|\cdot\|_\infty$ bounded. Then we note that for $\|x\|_\infty, \|y\|_\infty \leq R$,

$$\|f(x) - f(y)\|_2 \leq \|p_0\|_{\infty,R} \sum_{j=1}^{\ell} |\operatorname{tr}(p_j(x)) - \operatorname{tr}(p_j(y))| \prod_{i \neq j} \|\mathrm{tr}(p_i)\|_{\infty,R} + \|p_0(x) - p_0(y)\|_2 \prod_{i=1}^{\ell} \|\mathrm{tr}(p_j)\|_{\infty,R},$$

so it suffices to prove that $\operatorname{tr}(p_i)$ and p_0 are $\|\cdot\|_2$-Lipschitz on operator norm balls. But $\|\operatorname{tr}(p_i(x))1 - \operatorname{tr}(p_i(y))1\|_2 \leq \|p_i(x) - p_i(y)\|_2$, so it suffices to show that any monomial p is Lipschitz on an operator norm ball. We can write $p(x) = x_{i_1} \dots x_{i_\ell}$, and then use the same type of product argument as above to see that p is $\ell R^{\ell-1}$-Lipschitz on the operator norm ball of radius R.

(2) By definition, every $f \in C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ is a limit of trace polynomials with respect to the seminorms $\|\cdot\|_{2,R}$, and uniform continuity is preserved under uniform limits.

(3) Note that every $f \in \operatorname{TrP}_d$ and hence every $f \in C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ defines a unitarily equivariant function on $M_n(\mathbb{C})_{\text{sa}}^d$ for every n. By Lemma 11.5.4, for any n and any $x \in M_n(\mathbb{C})_{\text{sa}}^d$,

$$\|f(x) - \tau_n(f(x))\|_\infty \leq 2\sqrt{2} d^{1/2} \Theta \|f\|_{\text{Lip},R} \max_j \|x_j\|_\infty.$$

Since $|\tau_n(f(X))| \leq \|f\|_{2,R}$, we conclude that $\|f(X)\|_\infty$ is $\|\cdot\|_\infty$-bounded on the operator norm ball of radius R with the asserted estimate.

Actually, Lemma 11.5.4 as stated assumed that f was globally Lipschitz. However, as the function UxU^* used in the proof remains inside the $\|\cdot\|_\infty$-ball of radius R when $\|x\|_\infty \leq R$, the lemma could be sharpened to only use the Lipschitz norm on an R-ball. □

The space $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ subsumes the continuous functional calculus for self-adjoint operators in the following sense.

Lemma 13.3.5. *Let $\phi : \mathbb{R} \to \mathbb{C}$ be continuous.*

(1) There exists a unique $f \in C_{\text{tr,app}}(\mathbb{R}^{*1}, \|\cdot\|_2)$ such that $f(x) = \phi(x)$ for every self-adjoint matrix x. Moreover, this equality also holds if x is a d-tuple from a tracial W**-algebras with $\lambda_x \in \Sigma_{d,R}^{\text{app}}$.*

(2) This f is $\|\cdot\|_\infty$-bounded on $\|\cdot\|_\infty$-balls with

$$\|f\|_{\infty,R} = \sup\{|\phi(t)| : t \in [-R, R]\}.$$

(3) Let $\widehat{\phi}$ be the Fourier transform $\widehat{\phi}(t) = \int_{\mathbb{R}} e^{-2\pi i t s}\, ds$. If $(1 + |t|)\widehat{\phi}(t)$ is continuous and integrable, then f is globally $\|\cdot\|_2$-Lipschitz and

$$\|f\|_{\text{Lip}} \leq 2\pi \int_{\mathbb{R}} |t \widehat{\phi}(t)|\, dt.$$

Proof. First, suppose that ϕ is a polynomial and x is a self-adjoint matrix with $\|x\|_\infty \leq R$. Then $\|\phi(x)\|_\infty \leq \sup\{|\phi(t)| : t \in [-R, R]\}$ by the spectral mapping theorem, so that $\|\phi\|_{\infty,R} \leq \sup\{|\phi(t)| : t \in [-R, R]\}$. Actually, equality is achieved since a real number is a 1×1 matrix and thus participates in the supremum on the left-hand side.

If $\phi : \mathbb{R} \to \mathbb{C}$ is continuous, then by the Stone-Weierstrass theorem, there is a sequence of polynomials $(\phi_k)_{k \in \mathbb{N}}$ such that $\phi_k \to \phi$ uniformly on $[-R, R]$ for every R. Thus, by the claim we just proved, ϕ_k defines a Cauchy sequence with respect to $\|\cdot\|_{R,\infty}$, which must converge to some element $f \in C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. By straightforward limiting arguments, $f(x) = \phi(x)$ for self-adjoint d-tuples x with $\lambda_x \in \Sigma_{d,R}^{\text{app}}$, and (1) and (2) hold for ϕ and f since they hold for polynomials. Uniqueness in (1) holds because f is uniquely determined by its evaluation on matrices.

We defer the proof of (3) to §18.1; see Remark 18.1.8. □

The previous lemma enables to perform "cut-off tricks," which is one ingredient in the following lemma. This lemma shows that the continuity and boundedness properties discussed above extend automatically from matrix tuples to tuples of self-adjoint operators. The proof is more-or-less standard approximation arguments, but (2) is more subtle than it might first appear.

Lemma 13.3.6. *Let $f \in C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$.*

(1) Suppose x is a self-adjoint d-tuple from a tracial W^*-algebra $(\mathcal{M}, \tau)$ and $\|x\|_\infty \leq R$. If $\lambda_x \in \Sigma_{d,R_1}^{\text{app}}$ for some $R_1 \geq R$, then $\lambda_x \in \Sigma_{d,R}^{\text{app}}$.

(2) If $x^{(k)}$ and x are self-adjoint d-tuples from $(\mathcal{M}^{(k)}, \tau_k)$ and $(\mathcal{M}, \tau)$ with $\lambda_{x^{(k)}}, \lambda_x \in \Sigma_{d,R}^{\text{app}}$, and $x^{(k)}$ converges in non-commutative law to x, then

$$\|f(x)\|_\infty \leq \liminf_{k \to \infty} \|f(x^{(k)})\|_\infty.$$

(3) If f is $\|\cdot\|_\infty$-bounded on the $\|\cdot\|_\infty$-ball of radius R, and if x is a self-adjoint tuple from $(\mathcal{M}, \tau)$ with $\lambda_x \in \Sigma_{d,R}^{\text{app}}$, then

$$\|f(x)\|_\infty \leq \|f\|_{\infty,R}.$$

(4) Suppose that $\omega : [0, \infty) \to [0, \infty)$ is a modulus of continuity for f on the $\|\cdot\|_\infty$-ball of radius R, that is, ω is an increasing continuous function with $\omega(0) = 0$ and

$$\|f(x) - f(y)\|_2 \leq \omega(\|x - y\|_2) \text{ for } x, y \in M_n(\mathbb{C})_{\text{sa}}^d \text{ for } n \in \mathbb{N}.$$

Then the same statement holds when x and y are self-adjoint tuples in $(\mathcal{M}, \tau)$ with $\lambda_{x,y} \in \Sigma_{2d,R}^{\text{app}}$.

(5) The same claim as (4) holds except with the modulus of continuity on a $\|\cdot\|_2$-ball. (However, x and y are still assumed to be d-tuples of bounded operators.)

Proof. (1) This fact is well-known, but we will give the proof for completeness anyway. Let $(x^{(k)})_{k \in \mathbb{N}}$ be a sequence of matrix tuples that converges in law to x such that $\|x^{(k)}\|_\infty$ is uniformly bounded by some R_1. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a continuous function such $\phi(t) = t$ for $|t| \leq R$ and $|\phi(t)| \leq R$ everywhere. Let $y^{(k)}$ be the tuple $(\phi(x_1^{(k)}), \ldots, \phi(x_d^{(k)}))$. Since $\|x^{(k)}\|_\infty \leq R_1$ for all k, it is straightforward to check that $y^{(k)}$ still converges in non-commutative law to

x. The idea is that this is true when ϕ is a polynomial, and hence also true when ϕ is a continuous function.

(2) By Lemma 13.2.8, for any non-commutative polynomials p and q, we have

$$\begin{aligned}\tau_k(p(x^{(k)})^* f(x^{(k)}) q(x^{(k)})) &= \tau_k(q(x^{(k)}) p(x^{(k)})^* f(x^{(k)})) \\ &\to \tau(q(x)p(x)^* f(x)) \\ &= \tau(p(x)^* f(x) q(x)).\end{aligned}$$

Note that

$$|\tau(p(x^{(k)})^* f(x^{(k)}) q(x^{(k)}))| \leq \|f(x^{(k)})\|_\infty \|p(x^{(k)})\|_2 \|q(x^{(k)})\|_2.$$

This in the limit as $k \to \infty$, we have

$$|\tau(p(x)^* f(x) q(x))| \leq \left(\liminf_{k\to\infty} \|f(x^{(k)})\|_\infty\right) \|p(x)\|_2 \|q(x)\|_2;$$

Since this holds for all p and q, it follows that $f(x)$ agrees with a bounded operator on $L^2(\mathcal{M})$ with $\|f(x)\|_\infty \leq \liminf_{k\to\infty} \|f(x^{(k)})\|_\infty$.

(3) follows from (2) by taking $x^{(k)}$ to be a sequence of matrix approximations as in (1).

(4) If $\lambda_{(x,y)} \in \Sigma_{d,R}^{\mathrm{app}}$, then there is a sequence of matrix tuples $(x^{(k)}, y^{(k)})$ also with $\|\cdot\|_\infty$ bounded by R that converge in non-commutative law to (x, y). By Lemma 13.2.8, we have $\|x^{(k)} - y^{(k)}\|_2 \to \|x - y\|_2$ and $\|f(x^{(k)}) - f(y^{(k)})\|_2 \to \|f(x) - f(y)\|_2$, hence the inequality in (4) is satisfied.

(5) The argument is similar to (4). □

Remark 13.3.7. Suppose that f is $\|\cdot\|_2$-uniformly continuous on $\|\cdot\|_2$-balls. Suppose that $(\mathcal{M}, \tau)$ is a tracial W*-algebra which is embeddable into $\mathcal{R}^\omega$. Then we can define $f(x)$ for $x \in L^2(\mathcal{M}, \tau)_{\mathrm{sa}}^d$. Indeed, if $x^{(k)}$ is a sequence of tuples from $\mathcal{M}_{\mathrm{sa}}^d$ converging to x in $L^2(\mathcal{M}, \tau)^d$, then because f is uniformly continuous on $\|\cdot\|_2$-balls, the sequence $(f(x_k))_{k\in\mathbb{N}}$ is Cauchy and hence converges in $L^2(\mathcal{M}, \tau)$, and we may call this limit $f(x)$ since it is independent of the choice of approximating sequence. Then claim (3) of the last lemma generalizes to tuples from L^2.

13.4 Composition and algebraic operations

Next, we want to consider composition and algebraic operations for functions in $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. To start off, let us discuss composition of trace polynomials.

Definition 13.4.1. Let $f = (f_1, \dots, f_m) \in (\mathrm{TrP}_d)_{\mathrm{sa}}^m$. We define a map $\mathrm{TrP}_m \to \mathrm{TrP}_d$, denoted $g \mapsto g \circ f$, as follows.

First, there is a unique map $\Phi_f : \mathbb{C}\langle X_1, \dots, X_k\rangle \to \mathrm{TrP}_d$ sending X_j to $f_j(X)$. Then by the universal property of polynomial algebras, there is a unique $*$-homomorphism $\Psi_f : \mathrm{TrP}_m^0 \to \mathrm{TrP}_d$ sending $\mathrm{tr}(p)$ to $T(\Phi_f(p))$ for every monomial p. Then we define the composition map $g \mapsto g \circ f$ as $\Psi_f \otimes \Phi_f$.

The following is straightforward to verify.

Observation 13.4.2. *Let $f \in (\mathrm{TrP}_d)_{\mathrm{sa}}^m$ and $g \in \mathrm{TrP}_m$. If x is a self-adjoint d-tuple from a tracial* W^**-algebra $(\mathcal{M}, \tau)$, then $(g \circ f)(x) = g(f(x))$.*

Now we explain under what conditions we can perform composition for elements in the separation-completion.

Proposition 13.4.3. *Let $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\mathrm{sa}}^m$ and let $g \in C_{\mathrm{tr,app}}(\mathbb{R}^{*m}, \|\cdot\|_2)$. Suppose that either*

(A) f is $\|\cdot\|_\infty$-bounded on $\|\cdot\|_\infty$-balls, or

(B) g is $\|\cdot\|_2$-uniformly continuous on $\|\cdot\|_2$-balls.

Then there exists a unique function $g \circ f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ such that $(g \circ f)(x) = g(f(x))$ for all $x \in M_n(\mathbb{C})_{\mathrm{sa}}^d$ for all n. Moreover, this equality also holds for self-adjoint d-tuples x with $\lambda_x \in \Sigma_{d,R}$ for some R (where in Case (B), $g(f(x))$ is defined by Remark 13.3.7).

Proof. Observe that to prove the first conclusion of the proposition for a particular function g, it suffices to show that for every $R > 0$ and $\epsilon > 0$, there exists $h \in \mathrm{TrP}_d$ such that $\|g(f(x)) - h(x)\|_2 \leq \epsilon$ for all $x \in M_n(\mathbb{C})_{\mathrm{sa}}^d$ with $\|x\|_\infty \leq R$, for all n. Indeed, this claim would imply by diagonalization that we can choose a sequence $(h_k)_{k \in \mathbb{N}}$ from TrP_d which is Cauchy in $\|\cdot\|_{2,R}$ for every R with $h_k(x) \to g(f(x))$ for every matrix d-tuple x.

Case (A): Let $\mathcal{A}$ be the set of functions g such that the first conclusion of the lemma holds. We claim that if $g_1, g_2 \in \mathcal{A}$ and g_2 is $\|\cdot\|_\infty$-bounded on $\|\cdot\|_\infty$-balls, then $g_1 g_2 \in \mathcal{A}$. In light of the discussion above, fix $R > 0$ and $\epsilon > 0$. Note that $g_2 \circ f$ is $\|\cdot\|_\infty$-bounded on $\|\cdot\|_\infty$-balls, so let $R_1 > \|g_2\|_{\infty,R}$. Since $g_1 \in \mathcal{A}$, there exists $h_1 \in \mathrm{TrP}_d$ such that $\|g_1(f(x)) - h_1(x)\|_2 \leq \epsilon/2R_1$ whenever $x \in M_n(\mathbb{C})_{\mathrm{sa}}^d$ with $\|x\|_\infty \leq R$, for any n. Since h_1 is a trace polynomial, we have $\|h_1\|_{R,\infty} < \infty$. Let $R_2 > \|h_1\|_{R,\infty}$. Then there exist a trace polynomial h_2 such that $\|g_2(f(x)) - h_2(x)\|_2 \leq \epsilon/2R_2$ for $\|x\|_\infty \leq R$. Therefore, overall, for $\|x\|_\infty \leq R$,

$$
\begin{aligned}
&\|g_1(f(x))g_2(f(x)) - h_1(x)h_2(x)\|_2 \\
&\quad \leq \|(g_1(f(x)) - h_1(x))g_2(f(x))\|_2 + \|h_1(X)(g_2(f(x)) - h_2(x))\|_2 \\
&\quad \leq \|g_1(f(x)) - h_1(x)\|_2 \|g_2(f(x))\|_\infty + \|h_1(x)\|_\infty \|g_2(f(x)) - h_2(x)\|_2 \\
&\quad \leq \frac{\epsilon}{2R_2} R_2 + \frac{\epsilon}{2R_1} R_1 = \epsilon.
\end{aligned}
$$

Therefore, $g_1 g_2 \in \mathcal{A}$ as desired.

Now $\mathcal{A}$ clearly contains the coordinate functions X_j. So the above claim shows that $\mathcal{A}$ contains all non-commutative monomials. Hence, by linearity it contains $\mathbb{C}\langle X_1, \dots, X_k \rangle$.

Next, note that if $g \in \mathcal{A}$, then $\mathrm{tr}(g) \in \mathcal{A}$. This implies that $\mathcal{A}$ contains the traces of non-commutative monomials. Since $\mathrm{tr}(p)$ always evaluates to a scalar multiple of the identity matrix, it is $\|\cdot\|_\infty$ bounded on $\|\cdot\|_\infty$-balls. Therefore, by our claim about multiplication again, $\mathcal{A}$ contains $\operatorname{tr}(p_1)\dots\operatorname{tr}(p_\ell)p_0$ for any non-commutative monomials $p_0, \dots, p_\ell$.

Therefore, $\mathcal{A}$ contains TrP_m. So it suffices to show that $\mathcal{A}$ is closed. Suppose that $g_k \in \mathcal{A}$ and $g_k \to g$. If $R > 0$, let $R_1 = \|f\|_{\infty,R}$. Then $\|g_k \circ f - g_j \circ f\|_{2,R} \leq \|g_k - g_j\|_{2,R_1}$, and hence $(g_k \circ f)_{k\in\mathbb{N}}$ is Cauchy and thus converges to some function $g \circ f$.

Next, let $\mathcal{B}$ be the set of g such that the equality $(g \circ f)(x) = g(f(x))$ holds for d-tuples from tracial W*-algebras with $\lambda_x \in \Sigma_{d,R}^{\mathrm{app}}$. The same approximation arguments that we used for $\mathcal{A}$ also apply to $\mathcal{B}$ (relying on Lemma 13.3.6), which proves the second conclusion of the proposition.

Case (B): Let $(f_k)_{k\in\mathbb{N}}$ be a sequence of trace polynomials converging to f, and assume without loss of generality that f_k is self-adjoint (since we can replace it with $(f_k + f_k^*)/2$. Then $g \circ f_k$ is well-defined by Case (A). Let $R > 0$ and $\epsilon > 0$. Note that $R_1 := \sup_k \|f_k\|_{2,R} < \infty$. Because g is $\|\cdot\|_2$-uniformly continuous on the $\|\cdot\|_2$-ball of radius R_1, there exists a δ such that $\|x - y\|_2 < \delta$ and $\|x\|_2, \|y\|_2 \leq R_1$ implies $\|g(x) - g(y)\|_2 < \epsilon$. Therefore, whenever $\|f_j - f_k\|_{2,R_1} < \delta$, we have $\|g \circ f_j - g \circ f_k\|_{2,R} \leq \epsilon$. Hence, $(g \circ f_k)_{k\in\mathbb{N}}$ is Cauchy in $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$, and one sees easily that the limiting function satisfies the desired conclusion. □

Corollary 13.4.4. *Let $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\mathrm{sa}}^m$ be $\|\cdot\|_\infty$-bounded on $\|\cdot\|_\infty$-balls. Let $(x^{(k)})$ and x be a d-tuples from $(\mathcal{M}^{(k)}, \tau^{(k)})$ and $(\mathcal{M}, \tau)$ respectively with $\|x^{(k)}\|_\infty \leq R$ and $\|x\|_\infty \leq R$. If $x^{(k)}$ converges to x in law, then $f(x^{(k)})$ converges to $f(x)$ in law.*

Proof. If $g \in C_{\mathrm{tr,app}}(\mathbb{R}^{*m}, \|\cdot\|_2)$, then by the proposition, $g \circ f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. So by Lemma 13.2.8, $\tau^{(k)}(g \circ f(x^{(k)})) \to \tau(g \circ f(x))$. Thus, $f(x^{(k)})$ converges to $f(x)$ in law by the same lemma. □

Remark 13.4.5. As corollaries to Proposition 13.4.3, we have the following properties:

(1) The set of functions in $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ which are $\|\cdot\|_\infty$-bounded on $\|\cdot\|_\infty$-balls forms an algebra. This is because if f_1 and f_2 are self-adjoint, then $f_1 f_2 = g(f_1, f_2)$, where $g(X_1, X_2) = X_1 X_2$. And general products can be expressed in terms of the products of self-adjoint elements by decomposing into real and imaginary parts.

(2) If $f, g \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$, then $\operatorname{tr}(fg)$ is a well-defined element of $\mathbb{C}_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. This is because the function $\operatorname{tr}(X_1 X_2)$ is $\|\cdot\|_2$-uniformly continuous on $\|\cdot\|_2$-balls.

(3) It is also true that if f_1 and $f_2 \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ and if f_2 is $\|\cdot\|_\infty$-bounded on $\|\cdot\|_\infty$-balls, then $f_1 f_2$ is well-defined. This does not follow immediately from the proposition. However, it is true by the same argument that we used in Case (A) of the proof.

13.5 Asymptotic Approximation for Functions of Matrices

In this section, we explain how the elements of $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ represent the large-n limit of sequences of functions $(f^{(n)})_{n\in\mathbb{N}}$ on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ which are asymptotically approximable by trace polynomials, a certain notion of "good asymptotic behavior" from the viewpoint of free probability.

Definition 13.5.1. Let $f^{(n)} : M_N(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})$ for each n. We say that $(f^{(n)})_{n\in\mathbb{N}}$ is *asymptotically approximable by trace polynomials* if for every $R > 0$ and $\epsilon > 0$, there exists $f \in \mathrm{TrP}_d$ such that

$$\limsup_{n\to\infty} \sup_{\substack{x\in M_n(\mathbb{C})_{\mathrm{sa}}^d \\ \|x\|_\infty \le R}} \|f^{(n)}(x) - f(x)\|_2 \le \epsilon.$$

Definition 13.5.2. Let $f^{(n)} : M_N(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})$ for each n, and let $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. We say that $(f^{(n)})_{n\in\mathbb{N}}$ is *asymptotic to* f, or $f^{(n)} \rightsquigarrow f$, if for every $R > 0$,

$$\limsup_{n\to\infty} \sup_{\substack{x\in M_n(\mathbb{C})_{\mathrm{sa}}^d \\ \|x\|_\infty \le R}} \left\|f^{(n)}(x) - f(x)\right\|_2 = 0.$$

Lemma 13.5.3. *Let $f^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})$ for each n. Then $(f^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials if and only if there exists $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ such that $f^{(n)} \rightsquigarrow f$. Moreover, $\|f\|_{2,R} = \lim_{n\to\infty} \|f^{(n)}\|_{2,R}^{(n)}$ for each n.*

For the proof, it will be convenient to use the following notation and observation.

Definition 13.5.4. If $f : M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})$ or if $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$, then we write

$$\|f\|_{2,R}^{(n)} = \sup_{\substack{x\in M_n(\mathbb{C})_{\mathrm{sa}}^d \\ \|x\|_\infty \le R}} \|f(x)\|_2.$$

and

$$\|f\|_{\infty,R}^{(n)} = \sup_{\substack{x\in M_n(\mathbb{C})_{\mathrm{sa}}^d \\ \|x\|_\infty \le R}} \|f(x)\|_\infty.$$

Lemma 13.5.5. *Let $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. Then*

$$\lim_{n\to\infty} \|f\|_{2,R}^{(n)} = \|f\|_{2,R}.$$

Proof. It is clear that $\limsup_{n\to\infty} \|f\|_{2,R}^{(n)} \le \|f\|_{2,R}$. For the reverse direction, let $x \in M_\ell(\mathbb{C})_{\mathrm{sa}}^d$ with $\|x\|_\infty \le R$. For each $n \in \mathbb{N}$, let us write $n = \ell q + r$ where $0 \le r < \ell$ using division. Then let $x^{(n)} = x^{\oplus q} \oplus 0_r$, that is, a block diagonal matrix of q copies of x then an $r \times r$ zero block. For any non-commutative polynomial p, we have

$$\tau_n(p(x^{(n)})) = \frac{\ell q}{n} \tau_q(p(x)),$$

which converges to $\tau_q(p(x))$ as $n \to \infty$. Hence, $x^{(n)}$ converges to x in non-commutative law, by Lemma 13.2.8, $\|f(x^{(n)})\|_2 \to \|f(x)\|_2$, hence

$$\|f(x)\|_2 \le \liminf_{n\to\infty} \|f\|_{2,R}^{(n)}.$$

Since x and ℓ were arbitrary, $\|f\|_{2,R} \le \liminf_{n\to\infty} \|f\|_{2,R}^{(n)}$ as desired. □

Proof of Lemma 13.5.3. Suppose that $(f^{(n)})_{n \in \mathbb{N}}$ is asymptotically approximable by trace polynomials. For each $k \in \mathbb{N}$, choose a trace polynomial f_k such that

$$\limsup_{n \to \infty} \left\| f^{(n)} - f_k \right\|_{2,k}^{(n)} \leq 2^{-k}.$$

This implies that for $j, k \geq R$,

$$\|f_j - f_k\|_{2,R} = \lim_{n \to \infty} \|f_j - f_k\|_{2,R}^{(n)} \leq 2^{-j} + 2^{-k}.$$

Thus, $(f_k)_{k \in \mathbb{N}}$ is Cauchy in $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$, and hence it converges to some f. Then observe that

$$\left\| f^{(n)} - f \right\|_{2,R}^{(n)} \leq \left\| f^{(n)} - f_k \right\|_{2,R}^{(n)} + \|f - f_k\|_{2,R}.$$

and hence $\|f^{(n)} - f\|_{2,R}^{(n)} \to 0$ as $n \to \infty$, so $f^{(n)} \rightsquigarrow f$.

Conversely, suppose that $f^{(n)} \rightsquigarrow f$. Then for every $\epsilon > 0$ and $R > 0$, there is a trace polynomial g such that $\|f - g\|_{2,R} \leq \epsilon$. Then

$$\limsup_{n \to \infty} \left\| f^{(n)} - g \right\|_{2,R}^{(n)} \leq \limsup_{n \to \infty} \left\| f^{(n)} - f \right\|_{2,R}^{(n)} + \|f - g\|_{2,R} \leq \epsilon,$$

so $f^{(n)}$ is asymptotically approximable by trace polynomials as desired.

Finally, to show that $\|f\|_{2,R} = \lim_{n \to \infty} \|f^{(n)}\|_{2,R}^{(n)}$, note that $\|f\|_{2,R} = \lim_{n \to \infty} \|f\|_{2,R}^{(n)}$ by the previous lemma, but

$$\left| \|f\|_{2,R}^{(n)} - \|f^{(n)}\|_{2,R}^{(n)} \right| \leq \|f - f^{(n)}\|_{2,R}^{(n)} \to 0. \qquad \square$$

Most of the results from the last two sections have asymptotic versions that are proved in more or less the same way. Rather than prove them all exhaustively, we will focus on those which will be used in our treatment of random matrix models.

Lemma 13.5.6. *Let $f^{(n)}, g^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})$ such that $f^{(n)} \rightsquigarrow f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ and $g^{(n)} \rightsquigarrow g \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. Let $x^{(n)} \in M_n(\mathbb{C})_{\mathrm{sa}}^d$ such that $\|x^{(n)}\|_\infty \leq R$ and $x^{(n)}$ converges in non-commutative law to a self-adjoint d-tuple x from $(\mathcal{M}, \tau)$. Then*

$$\lim_{n \to \infty} \tau_n \left(f^{(n)}(x^{(n)})^* g^{(n)}(x^{(n)}) \right) = \tau(f(x)^* g(x)).$$

Proof. By Lemma 13.2.8,

$$\lim_{n \to \infty} \tau_n(f(x^{(n)})^* g(x^{(n)})) = \tau(f(x)^* g(x)).$$

But note that by Cauchy-Schwarz and the triangle inequality

$$\begin{aligned} &|\tau_n(f^{(n)}(x^{(n)})^* g^{(n)}(x^{(n)})) - \tau_n(f(x^{(n)})^* g(x^{(n)}))| \\ &\qquad \leq \left\| f^{(n)} - f \right\|_{2,R}^{(n)} \left\| g^{(n)} \right\|_{2,R}^{(n)} + \left\| g^{(n)} - g \right\|_{2,R}^{(n)} \|f\|_{2,R} \to 0. \qquad \square \end{aligned}$$

Definition 13.5.7. Let $f^{(n)} : M_n(\mathbb{C})_{\text{sa}}^d \to M_n(\mathbb{C})$. We say that $(f_{n\in\mathbb{N}}^{(n)}$ is $\|\cdot\|_2$*-uniformly equicontinuous* if there is exists a continuous increasing function $\omega : [0,\infty) \to [0,\infty)$ with $\omega(0) = 0$ (called a *modulus of continuity*) such that for every n and every x, $y \in M_n(\mathbb{C})_{\text{sa}}^d$, we have

$$\left\|f^{(n)}(x) - f^{(n)}(y)\right\|_2 \leq \omega\left(\|x-y\|_2\right).$$

Lemma 13.5.8. *Let $(f^{(n)})_{n\in\mathbb{N}}$ be $\|\cdot\|_2$-uniformly equicontinuous and suppose that $f^{(n)} \rightsquigarrow f \in C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. Then f is $\|\cdot\|_2$-uniformly continuous with respect to the same modulus of continuity. In particular, if $f^{(n)}$ is K-Lipschitz for every n, then f is also K-Lipschitz.*

Proof. If x and $y \in M_\ell(\mathbb{C})_{\text{sa}}^d$ with $\|x\|_\infty$, $\|y\|_\infty \leq R$, then by the same reasoning as in Lemma 13.5.5, there exist $x^{(n)}, y^{(n)} \in M_n(\mathbb{C})_{\text{sa}}^d$ with $\|\cdot\|_\infty$ bounded by R such that $(x^{(n)}, y^{(n)})$ converges in non-commutative law (X,Y). By Lemma 13.5.6, $\|f^{(n)}(x^{(n)}) - f^{(n)}(y^{(n)})\|_2 \to \|f(x) - g(x)\|_2$ and $\|x^{(n)} - y^{(n)}\|_2 \to \|x-y\|_2$. Hence, $\|f(x) - f(y)\|_2 \leq \omega(\|x-y\|_2)$. □

Remark 13.5.9. The previous definition and lemma also adapt for $\|\cdot\|_2$-uniform equicontinuity on $\|\cdot\|_2$-balls or $\|\cdot\|_\infty$ balls.

Definition 13.5.10. Let $f^{(n)} : M_n(\mathbb{C})_{\text{sa}}^d \to M_n(\mathbb{C})$. We say that $(f^{(n)})_{n\in\mathbb{N}}$ is *uniformly* $\|\cdot\|_\infty$*-bounded on* $\|\cdot\|_\infty$ *balls* if $\sup_n \|f^{(n)}\|_{R,\infty}^{(n)} < \infty$ for each $R > 0$.

Lemma 13.5.11. *Let $(f^{(n)})_{n\in\mathbb{N}}$ be uniformly $\|\cdot\|_\infty$-bounded on $\|\cdot\|_\infty$-balls and suppose $f^{(n)} \rightsquigarrow f \in C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. Then f is $\|\cdot\|_\infty$-bounded on $\|\cdot\|_\infty$-balls with*

$$\|f\|_{R,\infty} \leq \liminf_{n\to\infty} \left\|f^{(n)}\right\|_{R,\infty}^{(n)}.$$

Proof. Let $x \in M_\ell(\mathbb{C})_{\text{sa}}^d$ be bounded by R in $\|\cdot\|_\infty$. Choose $x^{(n)} \in M_n(\mathbb{C})_{\text{sa}}^d$ also bounded by R that converges in non-commutative law to x. Proceeding as in the proof of Lemma 13.3.6 (2), we obtain

$$\|f(x)\|_\infty \leq \liminf_{n\to\infty} \left\|f^{(n)}(x^{(n)})\right\|_\infty \leq \liminf_{n\to\infty} \left\|f^{(n)}\right\|_{R,\infty}^{(n)}.$$ □

Our random matrix results rely heavily on asymptotic approximation for composition of functions. In the following, if $f^{(n)} = (f_1^{(n)}, \dots, f_m^{(n)}) : M_n(\mathbb{C})_{\text{sa}}^d \to M_n(\mathbb{C})_{\text{sa}}^m$, then we say that $(f^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials, or respectively asymptotic to $f = (f_1, \dots, f_m) \in C_{\text{tr,app}}(R^{*d}, \|\cdot\|_2)^m$, if the corresponding statement holds for each of the coordinate functions $f_j^{(n)}$ and f_j.

Proposition 13.5.12. *Let $f^{(n)} : M_n(\mathbb{C})_{\text{sa}}^d \to M_n(\mathbb{C})_{\text{sa}}$ with $f^{(n)} \rightsquigarrow f \in C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\text{sa}}^m$, let $g^{(n)} : M_n(\mathbb{C})_{\text{sa}}^m \to M_n(\mathbb{C})$ with $g^{(n)} \rightsquigarrow g \in C_{\text{tr,app}}(\mathbb{R}^{*m}, \|\cdot\|_2)$. Suppose that either*

(A) $(f^{(n)})_{n\in\mathbb{N}}$ is uniformly $\|\cdot\|_\infty$-bounded on $\|\cdot\|_\infty$-balls, or

(B) $(g^{(n)})_{n\in\mathbb{N}}$ is $\|\cdot\|_2$-uniformly equicontinuous on $\|\cdot\|_2$-balls.

Then $g^{(n)} \circ f^{(n)} \rightsquigarrow g \circ f$*.*

Proof. Case (A): Fix $R > 0$. By assumption, there exists $R_1 > 0$ such that $\|f^{(n)}\|_{\infty,R}^{(n)} \leq R_1$ for all n, and hence $\|f\|_{\infty,R} \leq R_1$ also by Lemma 13.5.11. Then

$$\left\|g^{(n)} \circ f^{(n)} - g \circ f^{(n)}\right\|_{2,R}^{(n)} \leq \left\|g^{(n)} - g\right\|_{2,R_1}^{(n)} \to 0.$$

By Lemma 13.3.4 (1), g is $\|\cdot\|_2$-uniformly continuous on the $\|\cdot\|_\infty$-ball of radius R_1, so let ω be a modulus of continuity for f on this ball. Then

$$\left\|g \circ f^{(n)} - g \circ f\right\|_{2,R}^{(n)} \leq \omega\left(\|f^{(n)} - f\|_{2,R}^{(n)}\right) \to 0.$$

Therefore, $\|g^{(n)} \circ f^{(n)} - g \circ f\|_{2,R}^{(n)} \to 0$ as desired.

Case (B): Fix $R > 0$ and $\epsilon > 0$. Let h be a d-tuple of self-adjoint trace polynomials with $\|h - f\|_{2,R} \leq \epsilon$. Since $\|f^{(n)} - f\|_{2,R}^{(n)} \to 0$, there exists $R_1 > 0$ such that $\|f^{(n)}\|_{2,R}^{(n)} \leq R_1$ for all n and hence also

$$\limsup_{n\to\infty} \|f^{(n)} - h\|_{2,R}^{(n)} \leq \epsilon, \qquad \|h\|_{2,R} \leq R_1 + \epsilon$$

Let ω be a modulus of continuity for $(g^{(n)})_{n\in\mathbb{N}}$ on the $\|\cdot\|_2$-ball of radius of $R_1 + \epsilon$. Then

$$\limsup_{n\to\infty} \left\|g^{(n)} \circ f^{(n)} - g^{(n)} \circ h\right\|_{2,R}^{(n)} \leq \limsup_{n\to\infty} \omega\left(\|f^{(n)} - h\|_{2,R}^{(n)}\right) \leq \omega(\epsilon)$$
$$\|g \circ f - g \circ h\|_{2,R} \leq \omega\left(\|f - g\|_{2,R}\right) \leq \omega(\epsilon).$$

Let $R_2 = \|h\|_{R,\infty}$, which is finite because h is a trace polynomial. Then

$$\limsup_{n\to\infty} \|g^{(n)} \circ h - g \circ h\|_{2,R} \leq \limsup_{n\to\infty} \|g^{(n)} - g\|_{2,R_1} = 0.$$

Therefore, by the triangle inequality,

$$\limsup_{n\to\infty} \|g^{(n)} \circ f^{(n)} - g \circ f\|_{2,R} \leq 2\omega(\epsilon).$$

Since ϵ and R were arbitrary, we are done. □

Finally, asymptotically approximable sequences are closed under limits in an appropriate sense.

Lemma 13.5.13. *For* $k, n \in \mathbb{N}$*, let* $f_k^{(n)} : M_n(\mathbb{C})_{\text{sa}}^d \to M_n(\mathbb{C})$*. Suppose that for each* k*, we have* $f_k^{(n)} \rightsquigarrow f_k \in C_{\text{tr,app}}(R^{*d}, \|\cdot\|_2)$*. Suppose further that* $f^{(n)} : M_n(\mathbb{C})_{\text{sa}}^d \to M_n(\mathbb{C})$ *such that for each* $R > 0$*,*

$$\lim_{k\to\infty} \limsup_{n\to\infty} \left\|f_k^{(n)} - f^{(n)}\right\|_{2,R}^{(n)} = 0.$$

Then f_k *converges in* $C_{\text{tr,app}}(R^{*d}, \|\cdot\|_2)$ *to some* f*, and we have* $f^{(n)} \rightsquigarrow f$*.*

Proof. Let
$$\epsilon_{k,R} = \limsup_{n\to\infty} \left\| f_k^{(n)} - f^{(n)} \right\|_{2,R}^{(n)}.$$
By Lemma 13.5.3 and the triangle inequality
$$\|f_j - f_k\|_{2,R} = \lim_{n\to\infty} \|f_j - f_k\|_{2,R}^{(n)} \leq \epsilon_{j,R} + \epsilon_{k,R}.$$

Since $\epsilon_{k,R} \to 0$ as $k \to \infty$, it follows that $(f_k)_{k\in\mathbb{N}}$ is Cauchy in $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ and hence converges to some f, and moreover $\|f_k - f\|_{2,R} \leq \epsilon_{k,R}$. Note that
$$\left\| f^{(n)} - f \right\|_{2,R}^{(n)} \leq \|f^{(n)} - f_k^{(n)}\|_{2,R}^{(n)} + \|f_k^{(n)} - f_k\|_{2,R}^{(n)} + \|f_k - f\|_{2,R}.$$

Hence,
$$\limsup_{n\to\infty} \left\| f^{(n)} - f \right\|_{2,R}^{(n)} \leq \epsilon_{k,R} + 0 + \epsilon_{k,R} = 2\epsilon_{k,R}.$$
Taking $k \to \infty$, we obtain $\lim_{n\to\infty} \|f^{(n)} - f\|_{2,R} = 0$, so that $f^{(n)} \rightsquigarrow f$ as desired. □

13.6 $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ as a vector bundle

In this section, we continue the themes suggested by Remark 13.4.5. This section is not needed for any of the proofs of our main results. Rather, it is intended to bring conceptual clarity to the definition of $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. Thus, our discussion will be more colloquial and assume wider background knowledge.

We introduce the following temporary notation:

- We denote $\mathcal{H}_d = C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$.
- $\mathcal{F}_d$ will be the set of elements in $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ that are $\|\cdot\|_\infty$-bounded on $\|\cdot\|_\infty$-balls.
- $\mathcal{F}_d^0$ will denote the closure of TrP_d^0 in $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. We refer to $\mathcal{F}_d^0$ as the *scalar-valued part* of $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. Note that $\mathcal{F}_d^0 \subseteq \mathcal{F}_d$.

In light of Remark 13.4.5, we have the following properties.

- $\mathcal{F}_d$ is a $*$-algebra and $\mathcal{F}_d^0$ is a $*$-subalgebra contained in the center.
- $\mathcal{H}_d$ is a bimodule over $\mathcal{F}_d$.
- There is a map $\langle \cdot, \cdot \rangle_d : \mathcal{H}_d \times \mathcal{H}_d \to \mathcal{F}_d^0$ given by $(f, g) \mapsto \operatorname{tr}(f^* g)$. This is map is right- (and left-) $\mathcal{F}_d^0$-linear in the second coordinate and thus is an "$\mathcal{F}_d^0$-valued inner product."
- The left and right actions of $\mathcal{F}_d$ on $\mathcal{H}_d$ are $*$-representations.

- The map $\operatorname{tr} : \mathcal{F}_d \to \mathcal{F}_d^0$ is an "$\mathcal{F}_d^0$-valued trace."

Proposition 13.6.1. *For $d \geq 2$, the center of $\mathcal{F}_d$ is $\mathcal{F}_d^0$.*

Proof. Let f be in the center of $\mathcal{F}_d$. Note that the evaluation maps $\operatorname{ev}_{\mathcal{M},\tau,x}$ are $*$-homomorphisms $\mathcal{F}_d \to \mathrm{W}^*(x)$ by Proposition 13.4.3. Thus, for every self-adjoint tuple x with $\lambda_x \in \Sigma_{d,R}$, the function $f(x)$ is in the center of $\mathrm{W}^*(x)$.

It is well known that for each n, the set of self-adjoint pairs (x_1, x_2) that generate $M_n(\mathbb{C})$ as an algebra is generic. Hence, for generic $x \in M_n(\mathbb{C})_{\mathrm{sa}}^d$, $f(x)$ must be in the center of $M_n(\mathbb{C})$, which is $\mathbb{C}$. Therefore, $f(x) = \operatorname{tr}(f(x))$. By approximation, we have $f(x) = \operatorname{tr}(f(x))$ for all x, and hence $f \in \mathcal{F}_d^0$. $\square$

Remark 13.6.2. Alternatively, we could use the fact that factor traces are dense in $\Sigma_{d,R}^{\mathrm{app}}$. That proof does not depend on matrix approximation, and hence would work for the versions of $\mathcal{F}_d$ and $\mathcal{F}_d^0$ over $\Sigma_{d,R}$ rather than $\Sigma_{d,R}^{\mathrm{app}}$.

Proposition 13.6.3. *$\mathcal{F}_d^0$ is isomorphic as a $*$-algebra to the projective limit of the algebras $C(\Sigma_{d,R}^{\mathrm{app}})$ over the system of restriction maps $C(\Sigma_{d,R_1}^{\mathrm{app}}) \to C(\Sigma_{d,R_2}^{\mathrm{app}})$ for $R_1 \geq R_2$.*

Proof. By Lemma 13.2.8, $f \in \mathcal{F}_d^0$ and if $x^{(k)}$ and x are d-tuples bounded in $\|\cdot\|_\infty$ by R, approximable in non-commutative law by matrices, such that $x^{(k)}$ converges in law to x, then $f(x^{(k)}) \to f(x)$. In particular, $f(x)$ only depends on the law of x, so there exists a function $\phi_R(f) : \Sigma_{d,R}^{\mathrm{app}} \to \mathbb{C}$ such that $f(x) = \phi_R(f)(\lambda_x)$ whenever $\lambda_x \in \Sigma_{d,R}^{\mathrm{app}}$. Moreover, this $\phi_R(f)$ is continuous. It is immediate that $\phi_{R_1}(f)|_{\Sigma_{d,R_2}} = \phi_{R_2}(f)$ for $R_1 \geq R_2$.

Hence, we get a well-defined $*$-homomorphism from $\mathcal{F}_d^0$ into the projective limit, sending f to the element ϕ which is the projective limit of the maps ϕ_R associated f. This $*$-homomorphism is injective since $f(x) = \phi_R(\lambda_x)$. Note also that $\|f\|_{R,\infty} = \|\phi_R\|_{C(\Sigma_{d,R}^{\mathrm{app}})}$. It remains to prove surjectivity of our $*$-homomorphism.

We claim that if $\psi \in C(\Sigma_{d,R})$ and $\epsilon > 0$, then there exists $f \in \mathrm{TrP}_d^0$ with $\|\phi_R(f) - \psi\|_{C(\Sigma_{d,R}^{\mathrm{app}})} < \epsilon$. To prove this, note that $\mathcal{A} = \{\phi_R(f) : f \in \mathrm{TrP}_d^0\}$ is a $*$-subalgebra of $\Sigma_{d,R}$ that contains the function 1. Also, $\mathcal{A}$ separates points because if two laws λ_1 and λ_2 are distinct, then by definition, there is a non-commutative polynomial p such that $\lambda_1(p) \neq \lambda_2(p)$, which means that $\phi_R(\operatorname{tr}(p))(\lambda_1) \neq \phi_R(\operatorname{tr}(p))(\lambda_2)$. Thus, our claim follows from the Stone-Weierstrass theorem.

Let $(\psi_R)_{R>0}$ represent an element ϕ of the inductive limit. For $k \in \mathbb{N}$ choose a trace polynomial $f_k \in \mathrm{TrP}_d^0$ such that $\|\psi_k - \phi_k(f_k)\|_{C(\Sigma_{d,k})} < 1/k$. Then for each $R > 0$, the sequence $(\phi_R(f_k))_{k \in \mathbb{N}}$ converges to ϕ_R. Hence, $(f_k)_{k \in \mathbb{N}}$ is Cauchy with respect to $\|\cdot\|_{2,R}$ and hence f_k converges to some $f \in \mathcal{F}_d^0$. $\square$

Let $\Sigma_d^{\mathrm{app}} = \bigcup_{R>0} \Sigma_{d,R}$. Consider a vector bundle over Σ_d^{app}, where the fiber at each point λ is the Hilbert space $L^2(\lambda)$ obtained from the GNS construction. Each $f \in \mathcal{H}_d$ thus represents a section of this vector bundle, which assigns to a law λ the element $f(x_\lambda) \in L^2(\lambda)$, where x_λ is the canonical tuple of operators on $L^2(\lambda)$ given by the GNS construction. The next proposition characterizes the sections that arise from $\mathcal{H}_d$ as precisely the sections that are continuous in a certain sense.

Proposition 13.6.4. *Let F be a map that assigns to each $\lambda \in \Sigma_{d,R}$ a vector $F(\lambda) \in L^2(\lambda)$. Then the following are equivalent:*

(1) $F(\lambda) = f(x_\lambda)$ for some $f \in \mathcal{H}_d$.

(2) For every non-commutative polynomial p, the map $\lambda \mapsto \|p(X_\lambda) - F(\lambda)\|_2^2$ is continuous on $\Sigma_{d,R}$ for every $R > 0$.

Proof. (1) $\implies$ (2). It follows from Lemma 13.2.8 that if $f \in \mathcal{H}_d$, then $\lambda \mapsto \|p(x_\lambda) - f(x_\lambda)\|_2^2$ is continuous on $\Sigma_{d,R}^{\text{app}}$ for each R.

(2) $\implies$ (1). It suffices to show that for each $R > 0$ and $\epsilon > 0$, there exists $f \in \operatorname{TrP}_d$ such that $\|f(x_\lambda) - F(\lambda)\|_2 < \epsilon$ for all $\lambda \in \Sigma_{d,R}^{\text{app}}$. Indeed, if we prove this claim, then we can take a sequence of trace polynomials f_k associated to $\epsilon_k \to 0$ and $R_k \to \infty$, then $(f_k)_{k\in\mathbb{N}}$ will be Cauchy in $\|\cdot\|_{R,2}$ for each $R > 0$, and the limit f will satisfy (1).

So pick $R > 0$ and $\epsilon > 0$. For each $\lambda \in \Sigma_{d,R}$, since $F(\lambda) \in L^2(\lambda)$, there exists a non-commutative polynomial p_λ such that $\|p_\lambda(X_\lambda) - F(\lambda)\|_2 < \epsilon/2$. Because the map $\mu \mapsto \|p_\lambda(X_\mu) - F(\mu)\|_2$ is continuous on $\Sigma_{d,R}^{\text{app}}$, the set

$$\mathcal{U}_\lambda = \{\mu \in \Sigma_{d,R}^{\text{app}} : \|p_\lambda(X_\mu) - F(\mu)\|_2 < \epsilon\}$$

is open and contains λ. So $(\mathcal{U}_\lambda)_{\lambda\in\Sigma_{d,R}^{\text{app}}}$ is an open cover of the compact space $\Sigma_{d,R}^{\text{app}}$.

Therefore, there exists a finite cover $(\mathcal{U}_{\lambda_j})_{j=1}^k$ and a partition of unity $(\psi_j)_{j=1}^k$ in $C(\Sigma_{d,R}^{\text{app}})$ subordinated to the cover $(\mathcal{U}_{\lambda_j})_{j=1}^k$. By the Stone-Weierstrass theorem, there exists for each j a trace polynomial f_j such that $\|\phi_R(f_j) - \psi\|_{2,R}\|p_{\lambda_j}\|_{2,R} < \epsilon/2k$, where ϕ_R is the map from $\mathcal{F}_d^0$ to $C(\Sigma_{d,R})$ in the previous proposition.

Let $f = \sum_{j=1}^k f_j p_{\lambda_j}$. Then $\mu \in \Sigma_{d,R}^{\text{app}}$,

$$\begin{aligned}
\|f(x_\mu) - F(\mu)\|_2 &= \left\| \sum_{j=1}^k f_j(x_\mu)p_{\lambda_j}(x_\mu) - \phi_j(\mu)F(\mu) \right\|_2 \\
&\le \sum_{j=1}^k \|(f_j(x_\mu) - \phi_j(\mu))p_{\lambda_j}(x_\mu)\|_2 + \sum_{j=1}^k \|\phi_j(\mu)(p_{\lambda_j}(x_\mu) - F(\mu))\|_2 \\
&\le \sum_{j=1}^k \|\phi_R(f_j) - \psi_j\|_{C(\Sigma_{d,R}^{\text{app}}}\|p_{\lambda_j}\|_{2,R} + \sum_{j=1}^k \psi_j(\mu)\frac{\epsilon}{2} \\
&\le \frac{\epsilon}{2} + \frac{\epsilon}{2},
\end{aligned}$$

where the inequality $\|\psi_j(\mu)(p_{\lambda_j})(x_\mu) - F(\mu))\|_2 \le \psi_j(\mu)(\epsilon/2)$ follows because $\operatorname{supp}(\psi_j) \subseteq \mathcal{U}_{\lambda_j}$. □

Remark 13.6.5. In a similar way, each element of $\mathcal{F}_d$ represents a section of the vector bundle over Σ_d^{app} that assigns $\mathrm{W}^*(\lambda)$ to the point λ. The sections that arise from $\mathcal{F}_d$ are precisely those which are, for every $R > 0$, uniformly bounded in operator norm and continuous in the sense of the previous proposition on $\Sigma_{d,R}$.

The idea of the next proposition is that for any $\lambda_0 \in \Sigma_d^{\mathrm{app}}$ and $\zeta_0 \in L^2(\lambda_0)$, there exists a continuous section of the vector bundle $\lambda \mapsto L^2(\lambda)$ with $F(\lambda_0) = \zeta_0$. Or in more pedestrian terms, every element of $L^2(\mathrm{W}^*(x))$ can be realized as $f(x)$ for some $f \in \mathcal{H}_d$.

Proposition 13.6.6. *Let x be a d-tuple of self-adjoint operators in a tracial W^*-algebra and let $\mathcal{M} = \mathrm{W}^*(x)$.*

(1) For every $z \in L^2(\mathcal{M})$, there exists $f \in \mathcal{H}_d = C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ such that $f(x) = z$.

(2) We can choose f to be globally $\|\cdot\|_2$-bounded and $\|\cdot\|_2$-uniformly continuous.

(3) In fact, we can choose f to be globally $\|\cdot\|_2$-uniform limit of globally $\|\cdot\|_2$-Lipschitz functions which are globally $\|\cdot\|_\infty$-bounded.

(4) If $z \in \mathcal{M}$, we can choose f to also be globally $\|\cdot\|_\infty$-bounded by $\|z\|_\infty$ while simultaneously satisfying (3).

Proof. (1) Let $R > \|x\|_\infty$. Let $(p_k)_{k\in\mathbb{N}}$ be a sequence of non-commutative polynomials such that $p_k(x) \to z$ in $L^2(\mathcal{M})$. By passing to a subsequence, we may assume without loss of generality that $\|p_{k+1}(x) - p_k(x)\|_2 < 2^{-k}$. Let

$$\mathcal{U}_k = \{\lambda \in \Sigma_{d,R}^{\mathrm{app}} : \lambda((p_{k+1} - p_k)^*(p_{k+1} - p_k)) < (2^{-k})^2\}.$$

Because $\mathcal{U}_k$ is open and contains λ_x, by Urysohn's lemma there exists $\psi_k \in C(\Sigma_{d,R}^{\mathrm{app}})$ such that $0 \le \psi_k \le 1$ and $\mathrm{supp}(\psi_k) \subseteq \mathcal{U}_k$ and $\psi_k(\lambda_x) = 1$. By the Stone-Weierstrass theorem, there exists $f_k \in \mathrm{TrP}_d^0$ such that

$$\|\phi_R(f_k) - \psi_k\|_{C(\Sigma_{d,R}^{\mathrm{app}})} \|p_{k+1} - p_k\|_{2,R} \le 2^{-k}.$$

By shifting f_k by an additive constant, we can arrange at the same time that $f_k(x) = 1$.

By our choice of $\mathcal{U}_k$, we have for every $\lambda \in \Sigma_{d,R}^{\mathrm{app}}$ that

$$\|\psi_k(\lambda)(p_{k+1}(x_\lambda) - p_k(x_\lambda))\|_2 \le 2^{-k}.$$

Also,

$$\|(\psi_k(\lambda) - f_k(x_\lambda))(p_{k+1}(x_\lambda) - p_k(x_\lambda))\|_2 \le 2^{-k}.$$

Hence,

$$\|f_k \cdot (p_{k+1} - p_k)\|_{2,R} \le 2^{-k+1}.$$

Let $h \in C_c^\infty(\mathbb{R}, \mathbb{R})$ such that $h(t) = t$ for $|t| \le \|x\|_\infty$ and $|h(t)| \le R$ for all $t \in \mathbb{R}$. For a self-adjoint d-tuple $y = (y_1, \dots, y_d)$, let us write $h(y) := (h(y_1), \dots, h(y_d))$. Then

$[f_k(p_{k+1} - p_k)] \circ h \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ by Lemma 13.3.5 and Proposition 13.4.3. Also, for every R', we have

$$\|[f_k \cdot (p_{k+1} - p_k)] \circ h\|_{2,R'} \leq \|f_k \cdot (p_{k+1} - p_k)\|_{2,R} \leq 2^{-k+1}$$

because h maps into the $\|\cdot\|_\infty$-ball of radius R. Therefore, the series

$$f := p_1 \circ h + \sum_{k=1}^{\infty} [f_k \cdot (p_{k+1} - p_k)] \circ h$$

converges in $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$. And note that

$$\begin{aligned} f(x) &= p_1(h(x)) + \sum_{k=1}^{\infty} [f_k \cdot (p_{k+1} - p_k)] \circ h(x) \\ &= p_1(x) + \sum_{k=1}^{\infty} f_k(x)(p_{k+1}(x) - p_k(x)) \\ &= p_1(x) + \sum_{k=1}^{\infty} (p_{k+1}(x) - p_k(x)) \\ &= z. \end{aligned}$$

Therefore, (1) is proved.

(2) follows immediately from (3). To prove (3), continuing with the notation from (1), note that by Lemma 13.3.5, h is globally $\|\cdot\|_2$-Lipschitz and $\|\cdot\|_\infty$-bounded. Also, $f_k \cdot (p_{k+1} - p_k)$ is a trace polynomial, and hence it is $\|\cdot\|_2$-Lipschitz and $\|\cdot\|_\infty$-bounded on the $\|\cdot\|_\infty$-ball of radius R. Thus, $[f_k \cdot (p_{k+1} - p_k)] \circ h$ is globally $\|\cdot\|_2$-Lipschitz and globally $\|\cdot\|_\infty$ bounded. The same holds for $p_1 \circ h$. Thus, the terms in the series defining f are globally $\|\cdot\|_2$-Lipschitz and $\|\cdot\|_\infty$-bounded and the series converges $\|\cdot\|_2$-uniformly everywhere as shown above.

(4) By polar decomposition, we can write $z = uy$ where u is unitary and y is positive. Let $h \in C_c^\infty(\mathbb{R}, \mathbb{R})$ be a function that maps $[0, 1]$ onto $[0, \|z\|_\infty]$ in a strictly increasing fashion and is globally bounded by $\|z\|_\infty$, which is possible by standard bump function constructions. By spectral theory, there exist self-adjoints z_1 and z_2 in $\mathcal{M}$ such that $u = e^{iz_1}$ and $y = h(z_2)$ (here $z_2 = (h|_{[0,1]})^{-1}(y)$).

By (1), there exist $f_1, f_2 \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ such that $z_1 = f_1(x)$ and $z_2 = f_2(x)$. We can assume that f_1 and f_2 are self-adjoint (by replacing f_j with $(f_j + f_j^*)/2$). By Lemma 13.3.5 (3), h is $\|\cdot\|_2$-Lipschitz; also, e^{iX} is globally $\|\cdot\|_2$-Lipschitz (see Lemma 18.1.4 and Corollary 18.1.7). Thus, by Proposition 13.4.3 (B), $e^{if_1} \cdot (h \circ f_2)$ is well defined, and

$$e^{if_1(x)} h(f_2(x)) = e^{iz_1} h(z_2) = z.$$

Now e^{if_1} is bounded in $\|\cdot\|_\infty$ by 1 and $h \circ f_2$ is bounded in $\|\cdot\|_\infty$ by $\|z\|_\infty$.

Finally, by (3) we can choose sequences of self-adjoint, globally $\|\cdot\|_2$-Lipschitz, globally $\|\cdot\|_\infty$-bounded functions $f_1^{(k)}$ and $f_2^{(k)}$ that converge uniformly $\|\cdot\|_2$ to f_1 and f_2. Then $e^{if_1^{(k)}} \cdot (h \circ f_2^{(k)})$ is also globally $\|\cdot\|_2$-Lipschitz and globally $\|\cdot\|_\infty$-bounded, and it converges $\|\cdot\|_2$-uniformly to $e^{if_1} \cdot (h \circ f_2)$. □

13.7 Model-theoretic viewpoint

Like the last section, this section is intended purely as motivation for studying the space $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$, as well as making conceptual connections for future research. The goal is to explain how the functions in $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ has a natural interpretation in the model theory of tracial von Neumann algebras introduced in [FHS13, FHS14a, FHS14b], namely as certain quantifier-free definable functions.

For the sake of making the connection clearer, we will actually be looking at a variant of this space. First of all, instead of using self-adjoint variables $X_1, \dots, X_d$, consider variables $Z_1, \dots, Z_d$ representing arbitrary bounded operators. Let $\mathbb{C}\langle Z_1, \dots, Z_d, Z_1^*, \dots, Z_d^* \rangle$ be the universal $*$-algebra generated by d-variables; we call its elements *non-commutative $*$-polynomials*. A *non-commutative $*$-law* is a linear map $\lambda : \mathbb{C}\langle Z_1, \dots, Z_d, Z_1^*, \dots, Z_d * \rangle \to \mathbb{C}$ which is unital, positive, exponentially bounded, and tracial. We denote the space of such $*$-laws with exponential bound R by $\Sigma_{d,R}^*$. Of course, the $*$-polynomials in the variables $Z_1, \dots Z_d$ are equivalent to non-commutative polynomials in the $2d$ self-adjoint variables $\mathrm{Re}(Z_j) = (Z_j + Z_j^*)/2$ and $\mathrm{Im}(Z_j) = (Z_j - Z_j^*)/2i$.

The algebra TrP_d^* of trace $*$-polynomials is defined similarly to TrP_d except that the non-commutative polynomials both inside and outside the trace symbol are now taken from $\mathbb{C}\langle Z_1, \dots, Z_d, Z_1^*, \dots, Z_d^* \rangle$. Moreover, for each tuple $z \in \mathcal{M}^d$, where $(\mathcal{M}, \tau)$ is a tracial W*-algebra, there is an evaluation $*$-homomorphism $\mathrm{ev}_{\mathcal{M},\tau,z} : \mathrm{TrP}_d^* \to \mathcal{M}$ sending $p \in \mathbb{C}\langle Z_1, \dots, Z_d, Z_1^*, \dots, Z_d^* \rangle$ to $p(z)$ and $\mathrm{tr}(p)$ to $\tau(p(z))$.

Next, for $f \in \mathrm{TrP}_d^*$, we define

$$\|f\|_{2,R} = \sup\{\|f(z)\|_2 : z \in \mathcal{M}^d, \|z\|_\infty \leq R, (\mathcal{M}, \tau) \text{ tracial W}^*\text{-algebra}\}.$$

In contrast to our previous discussion, we consider arbitrary tuples z rather than only matrix tuples. This means that $\|f\|_{2,R}$ tests all $*$-laws in $\Sigma_{d,R}^*$, not only those in $\Sigma_{d,R}^{*,\mathrm{app}}$ (meaning the ones that can be approximated by matrices). We denote by $C_{\mathrm{tr}}((\mathbb{R} * i\mathbb{R})^{*d}, \|\cdot\|_2)$ the Fréchet space obtained as the separation-completion of TrP_d^* with respect to the seminorms $\|\cdot\|_{2,R}$ for $R > 0$.

Note that the evaluation map $\mathrm{ev}_{\mathcal{M},\tau,Z}$ extends to a well-defined map $C_{\mathrm{tr}}((\mathbb{R}*i\mathbb{R})^{*d}, \|\cdot\|_2) \to L^2(\mathcal{M}, \tau)$ for each $(\mathcal{M}, \tau)$ and $Z \in \mathcal{M}^d$. In particular, for each $f \in C_{\mathrm{tr}}((\mathbb{R} * i\mathbb{R})^{*d}, \|\cdot\|_2)$, there is a well-defined function

$$f^{\mathcal{M}} : \mathcal{M}^d \to L^2(\mathcal{M}, \tau) : z \mapsto \mathrm{ev}_{\mathcal{M},\tau,z}(f).$$

The space $C_{\mathrm{tr}}((\mathbb{R} * i\mathbb{R})^{*d}, \|\cdot\|_2)$ is the one we will interpret in the model-theoretic framework.

Next, let us sketch the setup of continuous model theory, or model theory for metric structures [BBH08, BU10]. We will follow the treatment in [FHS14a] which introduces "domains of quantification" to obviate some of the technical annoyances related to sorts.

A *language* $\mathcal{L}$ consists of:

- A set $\mathcal{S}$ whose elements are called *sorts*.
- For each $S \in \mathcal{S}$, a privileged relation symbol d_S (which will represent a metric) and a set $\mathcal{D}_S$ whose elements are called *domains of quantification for* S.
- A countably infinite set of *variable symbols* for each sort S.
- A set of *function symbols*.
- For each function symbol f, an assigned tuple $(S_1, \ldots, S_n)$ of sorts called the *domain*, another sort S called the *codomain*. We call n the *arity of* f.
- For each function symbol f with domain $(S_1, \ldots, S_n)$ and codomain S, and for every $\mathbf{D} = (D_1, \ldots, D_n) \in \mathcal{D}_{S_1} \times \cdots \times \mathcal{D}_{S_n}$, there is an assigned $D^f_{\mathbf{D}} \in \mathcal{D}_S$ (representing a range bound), and assigned moduli of continuity $\omega^f_{\mathbf{D},1}, \ldots, \omega^f_{\mathcal{D},n}$. (Here "modulus of coninuity" means a continuous increasing, zero-preserving function $[0, \infty) \to [0, \infty)$).
- A set of *relation symbols*.
- For each relation symbol R, an assigned domain $(S_1, \ldots, S_n)$ as in the case of function symbols.
- For each relation symbol R and for every $\mathbf{D} = (D_1, \ldots, D_n) \in \mathcal{D}_{S_1} \times \cdots \times \mathcal{D}_{S_n}$, there is an assigned $N^R_{\mathbf{D}} \in [0, \infty)$ and assigned moduli of continuity $\omega^R_{\mathbf{D},1}, \ldots, \omega^R_{\mathcal{D},n}$.

Given a language $\mathcal{L}$, an $\mathcal{L}$-*structure* assigns an object to each symbol in $\mathcal{L}$, called the *interpretation* of that symbol, in the following manner:

- Each sort $S \in \mathcal{S}$ is assigned a metric space $M(S)$, and the symbol d_S is interpreted as the metric on S.
- Each domain of quantification $D \in \mathcal{D}_S$ is assigned a subset $M(D) \subseteq M(S)$, such that $M(D)$ is complete for each D, and $M(S) = \bigcup_{D \in \mathcal{D}_S} M(D)$.
- Each function symbol f with domain $(S_1, \ldots, S_n)$ and codomain S is interpreted as a function $f^M : M(S_1) \times \cdots \times M(S_n) \to M(S)$. Moreover, for each $\mathbf{D} = (D_1, \ldots, D_n) \in \mathcal{D}_{S_1} \times \cdots \times \mathcal{D}_{S_n}$, the function f^M maps $M(D_1) \times \cdots \times M(D_n)$ into $M(D^f_{\mathbf{D}})$. Finally, f^M restricted to $D_1 \times \cdots \times D_n$ is uniformly continuous in the ith variable with modulus of continuity of $\omega^f_{\mathbf{D},i}$.

- Each relation symbol R with domain $(S_1, \dots, S_n)$ is interpreted as a function $R^M : M(S_1) \times \cdots \times M(S_n) \to \mathbb{R}$. Moreover, for each $\mathbf{D} = (D_1, \dots, D_n) \in \mathcal{D}_{S_1} \times \cdots \times \mathcal{D}_{S_n}$, f^M is bounded by $N_{\mathbf{D}}^R$ on $M(D_1) \times \cdots \times M(D_n)$ and uniformly continuous in the ith argument with modulus of continuity of $\omega_{\mathbf{D},i}^f$.

In particular, the *language* $\mathcal{L}_{\mathrm{Tr}}$ *of tracial* W**-algebra* has the following symbols:

- A single sort S, representing the W*-algebra.
- Domains of quantification $\{D_n\}_{n \in \mathbb{N}}$, representing the operator norm balls of radius $n \in \mathbb{N}$.
- The metric symbol d, representing the metric induced by $\|\cdot\|_2$.
- A binary function symbol $+$, representing addition.
- A binary function symbol $\cdot$, representing multiplication.
- A unary function symbol $*$, representing the adjoint operation.
- For each $\lambda \in \mathbb{C}$, a unary function symbol, representing multiplication by λ.
- A function symbol 1 of arity 0, representing the identity element.
- Two relation symbols $\operatorname{Re}\operatorname{tr}$ and $\operatorname{Im}\operatorname{tr}$, representing the real and imaginary parts of the trace.

Each function and relation symbol is assigned range bounds and moduli of continuity that one would expect, e.g. multiplication is supposed to map $D_m \times D_n$ into D_{mn} with $\omega^{\cdot}_{(D_m,D_n),1}(t) = nt$ and $\omega^{\cdot}_{(D_m,D_n),2} = mt$. A tracial W*-algebra represents a structure in $\mathcal{L}_{\mathrm{Tr}}$, where the symbols have the interpretations explained above. (Although not every $\mathcal{L}_{\mathrm{Tr}}$-structure comes from a tracial W*-algebra, one can formulate axioms in the language such that any structure satisfying these axioms comes from a tracial W*-algebra; see [FHS14a, §3.2].)

Terms in a language $\mathcal{L}$ are expressions obtained by iteratively composing the function symbols and variables. For example, if Z_1, Z_2, ... are variables in a sort S and $f : S \times S \to S$ and $g : S \times S \to S$ are function symbols, then $f(g(Z_1, Z_2), Z_1)$ is a term. Each term has assigned range bounds and moduli of continuity in each variable which are the obvious ones computed from those of the individual function symbols making up the composition. Any term can be interpreted in an $\mathcal{L}$-structure as a function. For example, in the language $\mathcal{L}_{\mathrm{Tr}}$, the terms are expressions obtained from iterating scalar multiplication, addition, multiplication, and the $*$-operation on variables and the unit symbol 1. If $(\mathcal{M}, \tau)$ is a tracial W*-algebra, then the interpretation of the term in $\mathcal{M}$ is a function represented by a $*$-polynomial.

Basic formulas in a language are obtained by evaluating relation symbols on terms. The basic formulas have assigned range bounds and moduli of continuity similar to the function symbols. In an $\mathcal{L}$-structure, a basic formula is interpreted as a real-valued function. In $\mathcal{L}_{\mathrm{Tr}}$, a basic formula can take the form $\operatorname{Im}\operatorname{tr}(f)$ or $\operatorname{Im}\operatorname{tr}(f)$ where f is an expression obtained by iterating the algebraic operations. Thus, when evaluated in a tracial W*-algebra, it corresponds to the real or imaginary part of the trace of a non-commutative $*$-polynomial.

Formulas are obtained from basic formulas by iterating several operations. First, given a formulas ϕ_1, ..., ϕ_n and $F : \mathbb{R}^n \to \mathbb{R}$ continuous, $F(\phi_1, \dots, \phi_n)$ is a formula. Second, if ϕ is a formula, D is a domain of quantification for some sort S, and x is a variable in S, then $\inf_{x \in D} \phi$ and $\sup_{x \in D} \phi$ are formulas. Each formula has an interpretation in every $\mathcal{L}$-structure $\mathcal{M}$, defined by induction on the complexity of the formula. If $\phi = F(\phi_1, \dots, \phi_n)$, then $\phi^M = F(\phi_1^M, \dots, \phi_n^M)$. The operations $\sup_{X \in D}$ or $\inf_{X \in D}$ will only change the interpretation of ϕ if ϕ has some occurrence of the variable X that is not already "bound" to some previous quantifier sup or inf. In that case, we take the sup or inf of the formula as X (in each of its "free" occurrences) ranges over $M(D)$. All these formulas also have assigned range bounds and moduli of continuity.

Our main concern at this point is the *quantifier-free formulas*, those obtained from the basic formulas by applying continuous functions $F : \mathbb{R}^n \to \mathbb{R}$ without using the quantifiers sup and inf. In the case of tracial W*-algebras, the real and imaginary parts of any trace polynomial $\phi \in \mathrm{TrP}_d^{*,0}$ can be represented as a quantifier-free formula in the variable symbols Z_1, ..., Z_d, since ϕ is obtained by adding and multiplying (which is, in particular, applying continuous functions $\mathbb{R}^n \to \mathbb{R}$) the real and imaginary parts of traces of non-commutative polynomials (which are basic formulas). Conversely, if we take $F(\operatorname{tr}(p_1), \dots, \operatorname{tr}(p_n))$, for some $F : \mathbb{R}^n \to \mathbb{R}$ continuous, this is an element of $C_{\mathrm{tr}}((\mathbb{R} * i\mathbb{R})^{*d}, \|\cdot\|_2)$ since F can be approximated by polynomials uniformly on bounded subsets of $\mathbb{R}^n$.

Similar to the way that we took the completion of $\mathrm{TrP}_d^{0,*}$, in continuous model theory, one often works with a completion of the space of formulas. Suppose that ϕ_k is a sequence of formulas depending on free variables $(Z_i)_{i \in \mathcal{I}}$ with $Z_i \in S_i$ and that $\phi_k^{\mathcal{M}}$ converges to some function $\mathcal{M}^I \to \mathbb{R}$, uniformly on $\prod_{i \in \mathcal{I}} \mathcal{M}(D_i)$ for each choice of domains of quantification $D_i \in \mathcal{D}_{S_i}$, with the same rate of convergence for every $\mathcal{M}$ in a certain class of $\mathcal{L}$-structures. Then ϕ is said to be a *definable relation* over this class. Similarly, if the ϕ_k's are quantifier-free, then ϕ is a *quantifier-free definable relation*.

For instance, if a sequence of trace polynomials in $\mathrm{TrP}_d^{0,*}$ converges in $\|\cdot\|_{2,R}$ for every $R > 0$, then the limiting object is a quantifier-free definable relation over the class of tracial W*-algebras. This leads to the following observation.

Observation 13.7.1. *The scalar-valued part of* $C_{\mathrm{tr}}((\mathbb{R} * i\mathbb{R})^{*d}, \|\cdot\|_2)$ *is equivalent to the space of definable relations in variables* Z_1, ..., Z_d *over the class of tracial* W**-algebras.*

A definable relation ϕ is always equipped with moduli of continuity just like the formulas are. One possible choice for the modulus of continuity on a domain $\mathbf{D}$ can be obtained by looking at how close ϕ_k uniformly approximates ϕ on $\mathbf{D}$ and the modulus of continuity of ϕ_k on this domain (the standard argument that uniform continuity is preserved under uniform limits). In our particular case of quantifier-free formulas, this is just the statement that any scalar-valued element of $C_{\mathrm{tr}}((\mathbb{R} * i\mathbb{R})^{*d}, \|\cdot\|_2)$ is $\|\cdot\|_2$-uniformly continuous on $\|\cdot\|_\infty$-balls, which we proved in Lemma 13.3.4 (2).

Furthermore, it is a basic fact in continuous model theory that if a relation is definable over one model $\mathcal{M}$, then you can create a definable relation over a larger class $\mathcal{C}$ by modifying the approximating sequence ϕ_k so that it converges everywhere. This is known as a *forced limit* (see [BU10, §3.2]). Our proof of Proposition 13.6.6 (1) is a loose analogue of this technique, except that it is done for operator-valued rather than scalar-valued functions.

The non-commutative law of a tuple $(z_1, \dots, z_d)$ in $(\mathcal{M}, \tau)$ also has an interpretation in model theory. It is equivalent information to knowing $\phi^{\mathcal{M}}(z_1, \dots, z_d)$ for every quantifier-free definable relation in the variables $Z_1, \dots, Z_d$. In model theory, this is called the *quantifier-free type* of $(z_1, \dots, z_d)$ in $(\mathcal{M}, \tau)$. (The *type* would of course encode the values of all formulas where $Z_1, \dots, Z_d$ are the only free variables, which would involve looking at sup's and inf's over some other variables.) It is well known in continuous model theory that definable relations are equivalent to continuous functions on the type space. Similarly, the quantifier-free definable relations are equivalent to continuous functions on the space of quantifier-free types. But in our example of tracial W^*-algebras, this is exactly the content of Proposition 13.6.3 (modulo replacing matrix-approximable laws with all laws).

Now that we have explained the scalar-valued part of $C_{\mathrm{tr}}((\mathbb{R} * i\mathbb{R})^{*d}, \|\cdot\|_2)$ corresponds to quantifier-free definable relations in variables $Z_1, \dots, Z_d$, let us move on to relate $C_{\mathrm{tr}}((\mathbb{R} * i\mathbb{R})^{*d}, \|\cdot\|_2)$ to quantifier-free definable functions.

Suppose that for each $\mathcal{L}$-structure $\mathcal{M}$ in a certain class $\mathcal{C}$, we are given a function $f^{\mathcal{M}} : \mathcal{M}(S_1) \times \dots \times \mathcal{M}(S_n) \to \mathcal{M}(S)$ (which has range bounds and uniform continuity properties similar to the terms). Then we say that f is a *definable function* over the class $\mathcal{C}$, if the function $g^{\mathcal{M}}(X_1, \dots, X_n, Y) = d_S(f^{\mathcal{M}}(X_1, \dots, X_n), Y)$ is equal to $\phi^{\mathcal{M}}$ for all $(\mathcal{M}, \tau)$, for some definable relation ϕ over $\mathcal{C}$. Similarly, f is a quantifier-free definable function if this occurs with ϕ a quantifier-free definable relation.

Proposition 13.7.2. *For each tracial* W^**-algebra* $(\mathcal{M}, \tau)$*, let* $f^{\mathcal{M},\tau} : \mathcal{M}^d \to \mathcal{M}$ *be a function that is* $\|\cdot\|_\infty$*-bounded on* $\|\cdot\|_\infty$*-balls. Then* $f^{\mathcal{M},\tau}$ *is a quantifier-free definable function over the class of tracial* W^**-algebras if and only if* f *is given by a function in* $C_{\mathrm{tr}}((\mathbb{R} * i\mathbb{R})^{*d}, \|\cdot\|_2)$*.*

Proof. Suppose that $f^{\mathcal{M},\tau}$ is a quantifier-free definable function, so that $\|f^{\mathcal{M},\tau}(z) - y\|_2$ is given by a definable relation in the variables $Z_1, \dots, Z_d$ and Y evaluated at (z, y). It follows from the composition properties of formulas that for each non-commutative $*$-polynomial p, the function $\|f^{\mathcal{M},\tau}(z) - p(z)\|_2$ is also a quantifier-free definable relation.

Now we claim that $f^{\mathcal{M},\tau}(z)$ is always in $\mathrm{W}^*(z)$. Fix $\mathcal{M}$ and z, and let $\mathcal{N} = \mathrm{W}^*(z)$. Because $\|f^{\mathcal{M},\tau}(z) - p(z)\|_2$ is a quantifier-free definable relation, it only depends on the non-commutative law of z, and hence $\|f^{\mathcal{M},\tau}(z) - p(z)\|_2 = \|f^{\mathcal{N},\tau}(z) - p(z)\|_2$. Since $\mathcal{N} = \mathrm{W}^*(Z)$, there is a sequence of non-commutative polynomials p_k such that $p_k(z)$ converges to $f^{\mathcal{N},\tau}(z)$, which implies that $p_k(z)$ also converges to $f^{\mathcal{M},\tau}(z)$, hence $f^{\mathcal{M},\tau}(z) = f^{\mathcal{N},\tau}(z)$.

Next, $\|f^{\mathrm{W}^*(z),\tau}(z) - p(z)\|_2$ depends continuously on λ_z, so by Proposition 13.6.4 (or more precisely its analogue for $C_{\mathrm{tr}}((\mathbb{R} * i\mathbb{R})^{*d}, \|\cdot\|_2))$, $f^{\mathrm{W}^*(Z),\tau}(Z)$ is given by $g(z)$ for some fixed $g \in C_{\mathrm{tr}}((\mathbb{R} * i\mathbb{R})^{*d}, \|\cdot\|_2)$. But we showed above that $f^{\mathcal{M},\tau}(z) = f^{\mathrm{W}^*(Z),\tau}(z)$ for each z, hence $f^{\mathcal{M},\tau}(z) = g(z)$.

Conversely, if $f \in C_{\mathrm{tr}}((\mathbb{R} * i\mathbb{R})^{*d}, \|\cdot\|_2)$, then so is $\|f(Z) - Y\|_2$ by Proposition 13.4.3, and hence it is a quantifier-free definable relation. □

There are several main points we wish to emphasize with these remarks on model theory. First, although the space $C_{\mathrm{tr}}((\mathbb{R} * i\mathbb{R})^{*d}, \|\cdot\|_2)$ has not been precisely discussed in prior work in free probability, it is a natural object of study which model theorists have in some sense already defined. Second, this is an example where the two different viewpoints (in this case, model theory and "non-commutative real analysis") converge on the same object, which ought to inspire future collaboration on this topic.

Third, we stress that the functions considered in this paper are all *quantifier-free*, which provides a heuristic reason for why their large-n behavior on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ is easy to understand. In our treatment of the solutions to certain differential equations in §14 and §15, we will not use any PDE tools that require taking suprema and infima, but rather we will construct the solutions iteratively using only "quantifier-free" operations. If we defined functions on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ using formulas with suprema and infima, then a priori there would be no way to know that the large-n behavior of $f^{(n)}(x^{(n)})$ only depends on the large-n behavior of the non-commutative law $\lambda_{x^{(n)}}$. We would need to know the asymptotic behavior of the *type* of $x^{(n)}$ in $M_n(\mathbb{C})_{\mathrm{sa}}^d$, not merely the *quantifier-free type*, and this relates to deep questions about the model theory of matrix algebras; see [FHS14a, §6.4].

CHAPTER 14

Tools: Basic constructions in differential equations

Our results in the next few chapters are based on the analysis of differential equations for functions on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ in the large-n limit. In preparation, this chapter will describe how trace polynomials behave under several basic operations in the theory of differential equations — differentiation, the heat equation, and the solution of the ODE. First, we will explain how to compute the Jacobian and Laplacian of trace polynomials. Next, we will show that the heat semigroup (or equivalently convolution with the Gaussian measure) on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ preserves asymptotic approximability by trace polynomials. Finally, we will show that if a sequence of vector fields is asymptotically approximable, then so is the sequence of flows generated by these vector fields.

14.1 Differentiation of trace polynomials

In this section, we compute first and second derivatives for trace polynomials as functions $M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})$. These computations are more or less contained in [Rai97], [Ceb13], [DHK13, §3].

We preface this with an explanation of our notation and conventions regarding classical derivatives of functions on $M_n(\mathbb{C})_{\mathrm{sa}}^d$. Recall that we equip $M_n(\mathbb{C})^d$ with the inner product $\langle \cdot, \cdot \rangle_2$ associated to the normalized trace, that is, $\langle x, y \rangle_2 = \sum_{j=1}^d \tau_n(x_j^* y_j)$. If we restrict to $M_n(\mathbb{C})_{\mathrm{sa}}^d$, this becomes a real inner product. Thus, we can identify $M_n(\mathbb{C})_{\mathrm{sa}}^d$ isometrically with $\mathbb{R}^{dn^2}$ by choosing some orthonormal basis. Furthermore, since every $z \in M_n(\mathbb{C})$ can be uniquely written as $x+iy$ for some self-adjoint x and y, we view $M_n(\mathbb{C})$ as the complexification of $M_n(\mathbb{C})_{\mathrm{sa}}$, and hence $M_n(\mathbb{C})^d$ as the complexification of $M_n(\mathbb{C})_{\mathrm{sa}}^d$.

(1) If $f : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{R}$ is differentiable, then $\nabla f(x_0)$ is the unique vector in $M_n(\mathbb{C})_{\mathrm{sa}}^d$ satisfying
$$f(x) - f(x_0) = \langle x - x_0, \nabla f(x_0) \rangle_2 + o(\|x - x_0\|_2^2).$$

(2) In accordance with standard convention for complex-valued functions, if $f : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{C}$, then $\nabla f(x_0)$ is the unique vector in $M_n(\mathbb{C})^d$ satisfying the above relation. (Here we took our inner product to be $\mathbb{C}$-linear in the right argument, hence $\nabla f(X_0)$ is on the right.)

(3) If $f : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{R}$ is differentiable, then we denote by $Jf(X_0)$ the unique $\mathbb{R}$-linear

transformation $M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})$ such that

$$f(x) - f(x_0) = Jf(x_0)[x - x_0] + o(\|x - x_0\|_2^2).$$

(4) If $f : M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})_{\mathrm{sa}}^d$, then the divergence $\nabla^\dagger f$ is obtained by summing the directional derivative of $\langle b, f \rangle_2$ in direction b over b in some orthonormal basis $\mathcal{B}$. We make the same definition for $f : M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})^d$ in accordance with standard conventions for complex-valued functions. It is well known that this is independent of the choice of orthonormal basis.

(5) Finally, the Laplacian Δf of a function $f : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{C}$ or a function $M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})$ is defined as the sum of the second directional derivatives in direction b over b in some orthonormal basis $\mathcal{B}$, and this is independent of the choice of orthonormal basis.

Similarly, we will use the notation ∇_{x_j}, J_{x_j}, $\nabla_{x_j}^\dagger$, Δ_{x_j} to denote the analogous differentiation operators with respect to the jth matrix in the tuple $x = (x_1, \dots, x_d)$. In other words, they are the differential operators obtained by freezing x_i for $i \neq j$ and viewing f as a function of $x_j \in M_n(\mathbb{C})_{\mathrm{sa}}$.

As the algebra TrP_d was defined in several stages, the computation of derivatives will naturally proceed in stages as well, starting with non-commutative polynomials, then examining the trace of a non-commutative polynomial, and finally multiplying these terms together.

The differentiation of non-commutative polynomials is described using the *free difference quotient operators* of Voiculescu [Voi98a]. We will use the letter $\mathcal{D}$ rather than the original symbol ∂ for these difference quotients as we prefer to reserve the symbol ∂ for its classical meaning.

Definition 14.1.1. The *free difference quotient* $\mathcal{D}_{X_j} : \mathbb{C}\langle X_1, \dots, X_d \rangle \to \mathbb{C}\langle X_1, \dots, X_d \rangle \otimes \mathbb{C}\langle X_1, \dots, X_d \rangle$ is the linear operator whose action on monomials is given by

$$\mathcal{D}_{X_j}[X_{i(1)} \dots X_{i(\ell)}] = \sum_{k : i(k) = j} X_{i(1)} \dots X_{i(k-1)} \otimes X_{i(k+1)} \dots X_{i(\ell)}.$$

Definition 14.1.2. If $\mathcal{A}$ is an algebra, we then define the *hash operation* $(\mathcal{A} \otimes \mathcal{A}) \times \mathcal{A} \to \mathcal{A}$ as the map given on simple tensors by $(a \otimes b, c) \mapsto acb$. The value of the hash operation on a pair $(w, c) \in (\mathcal{A} \otimes \mathcal{A}) \times \mathcal{A}$ is denoted $w \# c$, so that for instance $(a \otimes b) \# c = acb$.

Lemma 14.1.3. *Let $f \in \mathbb{C}\langle X_1, \dots, X_d \rangle$ and let $f^{(n)}$ be the restriction of f to a function $M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})$. For $x, y \in M_n(\mathbb{C})_{\mathrm{sa}}^d$, we have*

$$Jf^{(n)}(x)[y] = \sum_{j=1}^d [\mathcal{D}_{X_j} f](x) \# y_j,$$

$$J_{x_j} f^{(n)}(x)[y_j] = [\mathcal{D}_{X_j} f](x) \# y_j \text{ for each } j.$$

Proof. Since both sides are linear in f, it suffices to check the claim when f is a monomial. Suppose $f(X) = X_{i(1)} \dots X_{i(\ell)}$. Then for $t \in \mathbb{R}$, we have

$$\begin{aligned} f(x+ty) &= x_{i(1)} \dots x_{i(\ell)} + t \sum_{k=1}^{\ell} x_{i(1)} \dots x_{i(k-1)} y_{i(k)} x_{i(k+1)} \dots x_{i(\ell)} + O(t^2) \\ &= f(x) + t \sum_{k=1}^{\ell} (x_{i(1)} \dots x_{i(k-1)} \otimes x_{i(k+1)} \dots x_{i(\ell)}) \# y_{i(k)}. \end{aligned}$$

We sort the terms in the sum based on the value of $i(k)$ and thus obtain

$$\begin{aligned} Jf^{(n)}(x)[y] &= \sum_{k=1}^{\ell} (x_{i(1)} \dots x_{i(k-1)} \otimes x_{i(k+1)} \dots x_{i(\ell)}) \# y_{i(k)} \\ &= \sum_{j=1}^{d} \left(\sum_{k: i(k) = j} x_{i(1)} \dots x_{i(k-1)} \otimes x_{i(k+1)} \dots x_{i(\ell)} \right) \# y_j \\ &= \sum_{j=1}^{d} \mathcal{D}_{X_j} f(x) \# y_j. \end{aligned}$$

This proves the first formula, and the second one is equivalent to it. □

Next, the gradient of the function $\tau_n(f^{(n)}) : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{C}$ when $f \in \mathbb{C}\langle X_1, \dots, X_d \rangle$ is described in terms of Voiculescu's cyclic derivatives.

Definition 14.1.4. If $\mathcal{A}$ is an algebra, we define the *flipped multiplication map* $\mathfrak{m} : \mathcal{A} \otimes \mathcal{A} \to \mathcal{A}$ by $\mathfrak{m}(a \otimes b) = ba$.

Definition 14.1.5. The *jth cyclic derivative* is the map $\mathcal{D}_{X_j}^{\circ} := \mathfrak{m} \circ \mathcal{D}_{X_j} : \mathbb{C}\langle X_1, \dots, X_d \rangle \to \mathbb{C}\langle X_1, \dots, X_d \rangle$.

Observation 14.1.6. *If $\mathcal{A}$ is an algebra and $\tau : \mathcal{A}$ is a linear functional with $\tau(ab) = \tau(ba)$, then for $w \in \mathcal{A} \otimes \mathcal{A}$ and $c \in \mathcal{A}$, we have $\tau(w \# c) = \tau(\mathfrak{m}(w)c) = \tau(c\mathfrak{m}(w))$.*

Proof. Check it when w is a simple tensor $a \otimes b$:

$$\tau[(a \otimes b) \# c] = \tau(acb) = \tau(bac) = \tau[\mathfrak{m}(a \otimes b)a].$$

□

Lemma 14.1.7. *Let $f \in \mathbb{C}\langle X_1, \dots, X_d \rangle$ and let $f^{(n)}$ be the restriction of f to a function $M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{C}$. Then*

$$\begin{aligned} \nabla[\tau_n(f^{(n)})](x) &= [\mathcal{D}^{\circ} f](x), \\ \nabla_{x_j}[\tau_n(f^{(n)})](x) &= [\mathcal{D}_{X_j}^{\circ} f](x) \text{ for each } j. \end{aligned}$$

Proof. Let $X, Y \in M_n(\mathbb{C})_{\mathrm{sa}}^d$. Then

$$\begin{aligned}
\left.\frac{d}{dt}\right|_{t=0} \tau_n(f(x+ty)) &= \tau_n\left(\left.\frac{d}{dt}\right|_{t=0} f(x+ty)\right) \\
&= \sum_{j=1}^d \tau_n\left(\mathcal{D}_{X_j} f(x) \# y_j\right) \\
&= \sum_{j=1}^d \tau_n\left(y_j \cdot \mathfrak{m} \circ \mathcal{D}_{X_j} f(x)\right) \\
&= \langle Y, \mathcal{D}^\circ f(x)\rangle_2.
\end{aligned}$$

This proves the first formula, and the second one is equivalent to it. □

Next, we turn our attention to the divergence of $f^{(n)}$ when $f \in \mathbb{C}\langle X_1, \dots, X_d\rangle^d$. To compute this, we pick a convenient orthonormal basis $\mathcal{B}$ for $M_n(\mathbb{C})_{\mathrm{sa}}$, given by

$$\begin{aligned}
\mathcal{B} = \{\sqrt{n} e_{k,k} : 1 \le k \le n\} \cup \{\sqrt{n/2}(e_{k,\ell} + e_{\ell,k}) : 1 \le k < \ell \le n\} \\
\cup \{\sqrt{n/2}(ie_{k,\ell} - ie_{\ell,k} : 1 \le k < \ell \le n\}. \quad (14.1)
\end{aligned}$$

Of course, we obtain an orthonormal basis for $M_n(\mathbb{C})_{\mathrm{sa}}^d$ by using d copies of $\mathcal{B}$. The convenience of this basis lies in the following "magic lemma."

Lemma 14.1.8. *For $x \in M_n(\mathbb{C})$, we have*

$$\frac{1}{n^2} \sum_{b \in \mathcal{B}} bxb = \tau_n(x) I.$$

Proof. Let $x_{i,j}$ be the (i,j) entry of x. If we sum $(1/n^2)(bxb)$ over the first portion of the basis, we obtain

$$\frac{1}{n} \sum_k e_{k,k} x e_{k,k} = \frac{1}{n} \sum_k x_{k,k} e_{k,k}.$$

The second portion of the basis contributes

$$\frac{1}{2n} \sum_{k<\ell} (e_{k,\ell} + e_{\ell,k}) x (e_{k,\ell} + e_{\ell,k}) = \frac{1}{2n} \sum_{k<\ell} (x_{\ell,k} e_{k,\ell} + x_{\ell,\ell} e_{k,k} + x_{k,\ell} e_{\ell,k} + x_{k,k} e_{\ell,\ell}),$$

and the third portion contributes

$$\frac{1}{2n} \sum_{k<\ell} (ie_{k,\ell} - ie_{\ell,k}) X (ie_{k,\ell} - ie_{\ell,k}) = \frac{1}{2n} \sum_{k<\ell} (-x_{\ell,k} e_{k,\ell} + x_{\ell,\ell} e_{k,k} - x_{k,\ell} e_{\ell,k} + x_{k,k} e_{\ell,\ell}).$$

When we add the second and third parts, the terms involving off-diagonal entries of x cancel, and thus we are left with

$$\frac{1}{n}\sum_{k<\ell}(x_{\ell,\ell}e_{k,k}+x_{k,k}e_{\ell,\ell})=\sum_{k\neq\ell}x_{\ell,\ell}e_{k,k}.$$

Adding this to the contribution of the first part produces

$$\frac{1}{n}\sum_{k,\ell}x_{\ell,\ell}e_{k,k}=\left(\frac{1}{n}\sum_{\ell}x_{\ell,\ell}\right)\left(\sum_{k}e_{k,k}\right)=\tau_n(X)I. \qquad \square$$

Lemma 14.1.9.

(1) Let $f \in \mathbb{C}\langle X_1,\dots,X_d\rangle$ and let $f^{(n)}$ be its restriction to $M_n(\mathbb{C})_{\mathrm{sa}}^d$. Then

$$\frac{1}{n^2}\nabla_{x_j}^{\dagger}f^{(n)}(x)=\tau_n\otimes\tau_n[\mathcal{D}_{X_j}f(x)].$$

(2) Let $f \in \mathbb{C}\langle X_1,\dots,X_d\rangle^d$ and let $f^{(n)}$ be its restriction to $M_n(\mathbb{C})_{\mathrm{sa}}^d$. Then

$$\frac{1}{n^2}\nabla^{\dagger}f^{(n)}(x)=\sum_{j=1}^{d}\tau_n\otimes\tau_n[\mathcal{D}_{X_j}f(x)].$$

Proof. (2) clearly follows from (1). To prove (1),

$$\begin{aligned}\frac{1}{n^2}\nabla_{x_j}^{\dagger}f^{(n)}(x)&=\frac{1}{n^2}\sum_{b\in\mathcal{B}}\frac{d}{dt}\Big|_{t=0}\langle b,f^{(n)}(x_1,\dots,x_{j-1},x_j+tb,x_{j+1},\dots,x_d)\rangle_2\\&=\frac{1}{n^2}\sum_{b\in\mathcal{B}}\tau_n[b(\mathcal{D}_{X_j}f(x)\#b)].\end{aligned}$$

The proof will be complete once we show that for $w\in M_n(\mathbb{C})\otimes M_n(\mathbb{C})$,

$$\frac{1}{n^2}\sum_{b\in\mathcal{B}}\tau_n[b(w\#b)]=\tau_n\otimes\tau_n(w).$$

It suffices to check this when w is a simple tensor $a_1\otimes a_2$, and in that case, we have by Lemma 14.1.8 that

$$\frac{1}{n^2}\sum_{b\in\mathcal{B}}\tau_n[b(w\#b)]=\frac{1}{n^2}\sum_{b\in\mathcal{B}}\tau_n[ba_1b\cdot a_2]=\tau_n[\tau_n(a_1)I\cdot a_2]=\tau_n(a_1)\tau_n(a_2). \qquad \square$$

Corollary 14.1.10. *Let $f\in\mathbb{C}\langle X_1,\dots,X_d\rangle$ and let $f^{(n)}$ be its restriction to $M_n(\mathbb{C})_{\mathrm{sa}}^d$. Then*

$$\frac{1}{n^2}\Delta_{x_j}\tau_n(f^{(n)}(x))=\tau_n\otimes\tau_n[\mathcal{D}_{X_j}\mathcal{D}_{X_j}^{\circ}f(x)] \text{ for each } j,$$

$$\frac{1}{n^2}\Delta\tau_n(f^{(n)}(x)=\sum_{j=1}^{d}\tau_n\otimes\tau_n[\mathcal{D}_{X_j}\mathcal{D}_{X_j}^{\circ}f(x)].$$

Proof. Recall that $\Delta = \nabla^\dagger \nabla$ and then combine Lemmas 14.1.7 and 14.1.9. □

Next, we will compute the Laplacian of a non-commutative polynomial p as a vector-valued function. This will be done in terms of the map

$$\mathcal{D}^2_{X_j} : \mathbb{C}\langle X_1, \dots, X_d \rangle \to \mathbb{C}\langle X_1, \dots, X_d \rangle^3$$

given by

$$\mathcal{D}^2_{X_j} := (\mathcal{D}_{X_j} \otimes \mathrm{id}) \circ \mathcal{D}_{X_j} + (\mathrm{id} \otimes \mathcal{D}_{X_j}) \circ \mathcal{D}_{X_j}.$$

This makes sense because $\mathcal{D}_{X_j} \otimes \mathrm{id}$ and $\mathrm{id} \otimes \mathcal{D}_{X_j}$ are maps $\mathbb{C}\langle X_1, \dots, X_d \rangle^{\otimes 2} \to \mathbb{C}\langle X_1, \dots, X_d \rangle^{\otimes 3}$. Actually,

$$(\mathcal{D}_{X_j} \otimes \mathrm{id}) \circ \mathcal{D}_{X_j} = (\mathrm{id} \otimes \mathcal{D}_{X_j}) \circ \mathcal{D}_{X_j}.$$

To see this, consider the action of both operators on a monomial $X_{i(1)} \dots X_{i(\ell)}$. The operator $\mathcal{D}_{X_j}$ will pick a value of k with $i(k) = j$ and replace $X_{i(k)}$ with a tensor sign (and sum up the results over all such k). Then if we apply $\mathcal{D}_{X_j} \otimes \mathrm{id}$, that means we pick an index k' in the left tensorand and replace $X_{i(k')}$ with a tensor sign. Thus, the overall result is to consider two indices (k, k') with $i(k) = i(k') = j$ and $k' < k$ and replace the kth and k'th terms with tensor signs, and then sum the result over all k and k'. On the other hand, if we apply $\mathcal{D}_{X_j}$ and then $\mathrm{id} \otimes \mathcal{D}_{X_j}$, then the same thing occurs except that the second index k' is to the right of k instead of to the left. But if we switch the names of k and k', then that is the same as $(\mathcal{D}_{X_j} \otimes \mathrm{id}) \circ \mathcal{D}_{X_j}$. Thus,

$$\mathcal{D}^2_{X_j} = 2(\mathcal{D}_{X_j} \otimes \mathrm{id}) \circ \mathcal{D}_{X_j} = 2(\mathrm{id} \otimes \mathcal{D}_{X_j}) \circ \mathcal{D}_{X_j}.$$

We need one more piece of notation.

Definition 14.1.11. Let $\mathcal{A}$ be an algebra and τ a linear functional with $\tau(ab) = \tau(ba)$. Then we define the map $\eta : \mathcal{A}^{\otimes 3} \to \mathcal{A}$ by $\eta(a \otimes b \otimes c) = \tau(b)ac$.

Lemma 14.1.12. *Let $f \in \mathbb{C}\langle X_1, \dots, X_d \rangle$ and let $f^{(n)}$ be its restriction to $M_n(\mathbb{C})^d_{\mathrm{sa}}$. Then*

$$\frac{1}{n^2} \Delta_{x_j} f^{(n)}(x) = \eta[\mathcal{D}^2_{X_j} f(x)] \text{ for each } j,$$

$$\frac{1}{n^2} \Delta f^{(n)} = \sum_{j=1}^d \eta[\mathcal{D}^2_{X_j} f(x)].$$

Proof. By linearity, we may assume that $f(X)$ is a monomial $X_{i(1)} \dots X_{i(\ell)}$. Let $x \in M_n(\mathbb{C})^d_{\mathrm{sa}}$ and $b \in \mathcal{B}$. Then

$$\begin{aligned} f(x_1, \dots, x_{j-1}, x_j + tb, x_{j+1}, \dots, x_d) &= f(x) \\ &+ t \sum_{k : i(k) = j} x_1 \dots x_{i(k-1)} b x_{i(k+1)} \dots x_{i(\ell)} \\ &+ t^2 \sum_{\substack{k, k' : k < k' \\ i(k) = i(k') = j}} x_1 \dots x_{i(k-1)} b x_{i(k+1)} \dots x_{i(k'-1)} b x_{i(k'+1)} \dots x_{i(\ell)} \\ &+ O(t^3). \end{aligned}$$

The coefficient of t^2 is this Taylor expansion is half the second derivative with respect to x_j in the direction of b. Therefore,

$$\begin{aligned}
\frac{1}{n^2}\Delta_{x_j} f^{(n)}(x) &= \frac{2}{n^2}\sum_{b\in\mathcal{B}} \sum_{\substack{k,k':k<k' \\ i(k)=i(k')=j}} x_1 \dots x_{i(k-1)} b x_{i(k+1)} \dots x_{i(k'-1)} b x_{i(k'+1)} \dots x_{i(\ell)} \\
&= 2 \sum_{\substack{k,k':k<k' \\ i(k)=i(k')=j}} x_1 \dots x_{i(k-1)} \tau_n(x_{i(k+1)} \dots x_{i(k'-1)}) x_{i(k'+1)} \dots x_{i(\ell)} \\
&= \eta[\mathcal{D}^2_{X_j} f(x)],
\end{aligned}$$

where the second equality follows from Lemma 14.1.8 and the last equality follows from our above discussion of how $\mathcal{D}^2_{X_j}$ is computed. This proves the first asserted formula and the second follows from it. □

Now we have computed the Jacobian / gradient and the Laplacian for both a non-commutative polynomial f and the trace of a non-commutative polynomial. Using these formulas and the product rule for differentiation, we may compute these same first and second derivatives for arbitrary trace polynomials. For simplicity, consider the scalar-valued case first. Suppose

$$f = \operatorname{tr}(p_1)\dots\operatorname{tr}(p_\ell)$$

where $p_1, \dots, p_\ell \in \mathbb{C}\langle X_1, \dots, X_d\rangle$. Then by the product rule, for $x \in M_n(\mathbb{C})^d_{\mathrm{sa}}$, we have

$$\begin{aligned}
\nabla_{x_j} f^{(n)}(x) &= \sum_{k=1}^{\ell} \nabla_{x_j}[\tau_n(p_k(x))] \prod_{k'\neq k} \tau_n[p_{k'}(x)] \\
&= \sum_{k=1}^{\ell} [\mathcal{D}^\circ_{X_j} p_k](x) \prod_{k'\neq k} \tau_n[p_{k'}(x)].
\end{aligned}$$

Moreover,

$$\begin{aligned}
\frac{1}{n^2}\Delta_{X_j} f^{(n)}(x) &= \frac{1}{n^2}\sum_k \Delta_{x_j}[\tau_n(p_k(x))] \prod_{k'\neq k} \tau_n(p_{k'}(x)) \\
&\quad + \frac{1}{n^2} \sum_{k_1,k_2:k_1\neq k_2} \tau_n\Big(\nabla_{x_j}[\tau_n(p_{k_1}(x))]\nabla_{x_j}[\tau_n(p_{k_2}(x))]\Big) \prod_{k'\neq k_1,k_2} \tau_n(p_{k'}(x)) \\
&= \sum_k \tau_n \otimes \tau_n[\mathcal{D}^2_{X_j} p_k(x)] \prod_{k'\neq k} \tau_n(p_{k'}(x)) \\
&\quad + \frac{1}{n^2} \sum_{k_1,k_2:k_1\neq k_2} \tau_n\Big(\mathcal{D}^\circ_{X_j} p_{k_1}(x) \mathcal{D}^\circ_{X_j} p_{k_2}(x)\Big) \prod_{k'\neq k_1,k_2} \tau_n(p_{k'}(x))
\end{aligned}$$

In the above computation, since $\tau_n(p_k)$ is complex-valued, when we compute the "cross terms" that pair two gradients together, we must make sure to use the complex-bilinear extension of the inner product on $M_n(\mathbb{C})_{\mathrm{sa}}$, which is $(x,y) \mapsto \tau_n(xy)$, rather than the complex inner product $\langle \cdot, \cdot \rangle_2$.

Remark 14.1.13. Unlike everything else we have computed so far, the normalized Laplacian $(1/n^2)\Delta$ of a trace polynomial does not have the same formula for every n. In the large-n limit the "cross terms" from the product rule vanish and thus the Laplacian becomes a derivation (meaning it satisfies the Leibniz product rule).

Next, suppose that $g = \mathrm{tr}(p_0) \dots \mathrm{tr}(p_\ell) p$ where $p_0, \dots, p_\ell, p \in \mathbb{C}\langle X_1, \dots, X_d \rangle$. Letting $f = \mathrm{tr}(p_0) \dots \mathrm{tr}(p_\ell)$ as above, the product rule tells us that

$$J_{x_j} g^{(n)}(x)[y] = \tau_n(\nabla_{x_j} f^{(n)}(x) y) p^{(n)}(X) + g^{(n)}(x) J_{x_j} p^{(n)}(x)[y].$$

We have computed $\nabla_{x_j} f^{(n)}(x)$ above, and $J_{x_j} p^{(n)}(x)[y] = \mathcal{D}_{X_j} p(x) \# y$ by Lemma 14.1.3. Similarly, by the product rule,

$$\frac{1}{n^2} \Delta_{x_j} g^{(n)}(x) = \frac{1}{n^2} \Delta f^{(n)}(x) p^{(n)}(x) + f^{(n)}(x) \frac{1}{n^2} \Delta p^{(n)}(x) + \frac{2}{n^2} J_{x_j} p^{(n)}(x)[\nabla_{x_j} f^{(n)}(x)]$$

Like the scalar-valued case above, the $(1/n^2)\Delta$ terms are independent of n, but the cross terms are $O(1/n^2)$.

These formulas shows that the Laplacian of a trace polynomial is a trace polynomial, and in fact, the Laplacian can be computed purely formally as an element of TrP_d. Let us make this claim explicit.

Lemma 14.1.14. *There exists a unique linear operator $D^0_{X_j} : \mathrm{TrP}^0_d \to \mathrm{TrP}_d$ satisfying*

$$D^0_{X_j}[\mathrm{tr}(p_1) \dots \mathrm{tr}(p_\ell)] = \sum_{k=1}^{\ell} \mathcal{D}^\circ_{X_j} p_k \prod_{k' \neq k} \mathrm{tr}(p_{k'}). \tag{14.2}$$

Moreover, there exist unique linear operators $L^0_{X_j} : \mathrm{TrP}^0_d \to \mathrm{TrP}^0_d$ and $L^{0,(n)}_{X_j} : \mathrm{TrP}^0_d \to \mathrm{TrP}^0_d$ for $n \in \mathbb{N}$ such that

$$L^0_{X_j}[\mathrm{tr}(p_1) \dots \mathrm{tr}(p_\ell)] = \sum_{k=1}^{\ell} \mathrm{tr} \otimes \mathrm{tr}(\mathcal{D}_{X_j} \mathcal{D}^\circ_{X_j} p_k) \prod_{k' \neq k} \mathrm{tr}(p_{k'}). \tag{14.3}$$

and

$$L^{0,(n)}_{X_j}[\mathrm{tr}(p_1) \dots \mathrm{tr}(p_\ell)] = L^0_{X_j}[\mathrm{tr}(p_1) \dots \mathrm{tr}(p_\ell)] + \frac{1}{n^2} \sum_{k_1, k_2 : k_1 \neq k_2} \mathrm{tr}(\mathcal{D}^\circ_{X_j} p_k \cdot \mathcal{D}^\circ_{X_j} p_{k'}) \prod_{k' \neq k_1, k_2} \mathrm{tr}(p_{k'}). \tag{14.4}$$

Moreover, $\mathcal{D}^0_{X_j}$ decreases the grading by 1, and $L^0_{X_j}$ and $L^{0,(n)}_{X_j}$ decrease the grading by 2.

Proof. First, let us show that if p and q are monomials that are cyclically equivalent, then $\mathcal{D}^\circ_{X_j} p = \mathcal{D}^\circ_{X_j} q$. Let $\sigma : \{1, \dots, \ell\} \to \{1, \dots, \ell\}$ be the permutation with $\sigma(i) = i + 1 \bmod \ell$. Let

$$\begin{aligned} p(X) &= X_{i(1)} \dots X_{i(\ell)} \\ q(X) &= X_{i(\sigma^m(1))} \dots X_{i(\sigma^m(\ell))}. \end{aligned}$$

Then

$$\mathcal{D}^\circ_{X_j} p(X) = \sum_{k : i(k) = j} X_{i(k+1)} \dots X_{i(\ell)} X_{i(1)} \dots X_{i(k-1)} = \sum_{k : i(k) = j} X_{i(\sigma^k(1))} \dots X_{i(\sigma^k(\ell-1))}.$$

and similarly,

$$\mathcal{D}^\circ_{X_j} q(X) = \sum_{k' : i(\sigma^m(k')) = j} X_{i(\sigma^{k'+m}(1))} \dots X_{i(\sigma^{k'+m}(\ell-1))}.$$

The two sums are equivalent by the substitution $k = \sigma^m(k')$. Thus, $\mathcal{D}^\circ_{X_j} p$ is uniquely determined by $\operatorname{tr}(p)$ when p is a monomial. Recall that a basis for TrP_d^0 is given by terms of the form $\operatorname{tr}(p_1) \dots \operatorname{tr}(p_\ell)$ where each p_j is a monomial modulo cyclic equivalence. Hence, there are unique linear operators $D^0_{X_j}$, L_{X_j}, and $L^{(n)}_{X_j}$ that satisfy (14.2), (14.3), and (14.4) respectively when $p_1, \dots, p_\ell$ are monomials. But then the relations (14.2), (14.3), (14.4) hold for general non-commutative polynomials p_j by linearity. Finally, the claim about the grading is clear by inspection of the definitions of $\mathcal{D}^\circ_{X_j}$ and $\mathcal{D}_{X_j}$. □

Lemma 14.1.15. *There exists unique linear operators $L_{X_j} : \mathrm{TrP}_d \to \mathrm{TrP}_d$ and $L^{(n)}_{X_j} : \mathrm{TrP}_d \to \mathrm{TrP}_d$ for $n \in \mathbb{N}$ such that for $f \in \mathrm{TrP}_d^0$ and $p \in \mathbb{C}\langle X_1, \dots, X_d \rangle$,*

$$L_{X_j}(fp) = L^0_{X_j}(f) \cdot p + f \cdot \eta \circ \mathcal{D}^2_{X_j} p, \tag{14.5}$$

where η denotes the map $\mathbb{C}\langle X_1, \dots, X_d \rangle^{\otimes 3} \to \mathrm{TrP}_d$ given by

$$\eta(p_1 \otimes p_2 \otimes p_3) = \operatorname{tr}(p_2) p_1 p_3.$$

Similarly, for each $n \in \mathbb{N}$, there exists a unique linear operator $L^{(n)}_{X_j} : \mathrm{TrP}_d \to \mathrm{TrP}_d$ such that

$$L^{(n)}_{X_j}(fp) = L_{X_j}(fp) + \frac{2}{n^2} \mathcal{D}_{X_j} p \# D^0_{X_j}(f). \tag{14.6}$$

where $\#$ denotes the map $\mathbb{C}\langle X_1, \dots, X_2 \rangle^{\otimes 2} \times \mathrm{TrP}_d \to \mathrm{TrP}_d$ given by

$$(p_1 \otimes p_2) \# f = p_1 f p_2.$$

Moreover, the operators L_{X_j} and $L^{(n)}_{X_j}$ decrease the grading by 2. Moreover, we have

$$L_{X_j}|_{\mathrm{TrP}_d^0} = L^0_{X_j} \qquad L^{(n)}_{X_j}|_{\mathrm{TrP}_d^0} = L^{0,(n)}_{X_j}. \tag{14.7}$$

and

$$\operatorname{tr} \circ L_{X_j} = L_{X_j} \circ \operatorname{tr} \tag{14.8}$$

Proof. Recall that TrP_d was defined as the tensor product of TrP_d^0 and $\mathbb{C}\langle X_1, \dots, X_d \rangle$. The right-hand sides of (14.5) and (14.6) are bilinear in (f, p) and thus pass to well-defined maps on the tensor product. The fact about the grading follows by inspection of all the definitions leading up to this point. Next, (14.7) follows from substituting $p = 1$ in (14.5) and (14.6).

Finally, to show (14.8), first consider the case where p is a non-commutative monomial $p(X) = X_{i(1)} \dots X_{i(\ell)}$. Then $L_{X_j}[\mathrm{tr}(p)] = \mathrm{tr} \otimes \mathrm{tr}(\mathcal{D}_{X_j} \circ \mathcal{D}_{X_j}^\circ p)$. Now

$$\mathcal{D}_{X_j}^\circ p = \sum_{k : i(k) = j} X_{i(k+1)} \dots X_{i(\ell)} X_{i(1)} \dots X_{i(k-1)},$$

and hence

$$\begin{aligned}\mathcal{D}_{X_j} \circ \mathcal{D}_{X_j}^\circ p = \sum_{k : i(k) = j} \Bigg(&\sum_{\substack{k' > k \\ i(k') = j}} X_{i(k+1)} \dots X_{i(k'-1)} \otimes x_{i(k'+1)} \dots X_{i(\ell)} X_{i(1)} \dots X_{i(k-1)} \\ &+ \sum_{\substack{k' < k \\ i(k') = j}} X_{i(k+1)} \dots X_{i(\ell)} X_{i(1)} \dots X_{i(k'-1)} \otimes X_{i(k'+1)} \dots X_{i(k-1)} \Bigg).\end{aligned}$$

Thus, $\mathrm{tr} \otimes \mathrm{tr}(\mathcal{D}_{X_j} \circ \mathcal{D}_{X_j}^\circ p)$ can be expressed as

$$\begin{aligned}&\sum_{\substack{k,k' : k' > k \\ i(k) = i(k') = j}} \mathrm{tr}(X_{i(k+1)} \dots X_{i(k'-1)}) \, \mathrm{tr}(X_{i(k'+1)} \dots X_{i(\ell)} X_{i(1)} \dots X_{i(k-1)}) \\ &= \sum_{\substack{k,k' : k' > k \\ i(k) = i(k') = j}} \mathrm{tr} \left[X_{i(1)} \dots X_{i(k-1)} \, \mathrm{tr}(X_{i(k+1)} \dots X_{i(k'-1)}) X_{i(k'+1)} \dots X_{i(\ell)} \right] \\ &= \mathrm{tr} \circ \eta \circ (\mathrm{id} \circ \mathcal{D}_{X_j}) \circ \mathcal{D}_{X_j} p = \frac{1}{2} \mathrm{tr} \circ \eta \circ \mathcal{D}_{X_j}^2 p.\end{aligned}$$

In a similar way, the second term can be expressed as $(1/2) \, \mathrm{tr} \circ \eta \circ \mathcal{D}_{X_j}^2 p$. Therefore, $\mathrm{tr} \otimes \mathrm{tr}(\mathcal{D}_{X_j} \circ \mathcal{D}_{X_j}^\circ p) = \mathrm{tr}(L_{X_j} p)$, and this identity extends to all $p \in \mathbb{C}\langle X_1, \dots, X_d \rangle$ by linearity.

For the general case, let $f \in \mathrm{TrP}_d^0$ and let $p \in \mathbb{C}\langle X_1, \dots, X_d \rangle$. Then

$$L_{X_j}[f \, \mathrm{tr}(p)] = L_{X_j} f \cdot \mathrm{tr}(p) + f \cdot L_{X_j}[\mathrm{tr}(p)].$$

(This follows directly from (14.3) in the case where f is a product of traces of non-commutative polynomials.) Next, using the case of $\mathrm{tr}(p)$ we just proved,

$$L_{X_j}[f \, \mathrm{tr}(p)] = L_{X_j} f \cdot \mathrm{tr}(p) + f \cdot \mathrm{tr}(L_{X_j} p) = \mathrm{tr}[L_{X_j} f \cdot p + f \cdot L_{X_j} p] = \mathrm{tr}(L_{X_j}[fp]),$$

where the last equality follows from (14.5). □

The next proposition follows immediately from our earlier discussion of the Laplacian of a trace polynomial acting on $M_n(\mathbb{C})_{\text{sa}}^d$. This proposition is one rationale for studying trace polynomials rather than only non-commutative polynomials in asymptotic random matrix theory. The Laplacian of a non-commutative polynomial (given in Lemma 14.1.12) is not in general a non-commutative polynomial, but rather a trace polynomial (and this is true even in the large-n limit).

Proposition 14.1.16. *Let $f \in \operatorname{TrP}_d$ and let $f^{(n)}$ be the restriction of f to $M_n(\mathbb{C})_{\text{sa}}^d$. Then*

$$\Delta_{x_j} f^{(n)}(x) = [L_{X_j}^{(n)} f](x) \text{ for } x \in M_n(\mathbb{C})_{\text{sa}}^d.$$

Moreover, $L_{X_j}^{(n)} f = L_{X_j} f + O(1/n^2)$ coefficient-wise.

14.2 The heat semigroup

Proposition 14.1.16 suggests that the heat semigroup $\Phi_t^{(n)} = e^{t\Delta/2n^2}$ acting on trace polynomials has a well-defined large-n limit. More generally, for trace polynomials in $d_1 + d_2$ variables $X_1, \ldots, X_{d_1}, Y_1, \ldots, Y_{d_1}$, we will study the heat semigroup with respect to X, viewing Y as an auxiliary parameter. This is useful for our results on conditional expectation in §15, since Y represents the variables that we are conditioning on.

Let $f \in \operatorname{TrP}_{d_1+d_2}$ written as a function of $X = (X_1, \ldots, X_{d_1})$ and $Y = (Y_1, \ldots, Y_{d_2})$, let $L_X f = \sum_{j=1}^{d_1} L_{X_j} f$ and $L_X^{(n)} f = \sum_{j=1}^{d_1} L_{X_j}^{(n)} f$. Fix $t \in \mathbb{R}$. Let $\operatorname{TrP}_{d_1+d_2,k}$ be the space of trace polynomials of degree $\leq k$. Now L_X and $L_X^{(n)}$ are linear transformations of $\operatorname{TrP}_{d_1+d_2}$ that map $\operatorname{TrP}_{d_1+d_2,k}$ into $\operatorname{TrP}_{d_1+d_2,k-2}$. In particular, since $\operatorname{TrP}_{d_1+d_2,k}$ is finite-dimensional, the operator $T_k = \exp((t/2) L_X|_{\operatorname{TrP}_{d_1+d_2,k}})$ is defined by analytic functional calculus. But clearly $T_{k'}|_{\operatorname{TrP}_{d_1+d_2,k}} = T_k$ for $k' > k$, hence the union of the graphs of the T_k's defines a linear operator $e^{tL_X/2}$ on $\operatorname{TrP}_{d_1+d_2}$. We define $e^{tL_X^{(n)}/2} : \operatorname{TrP}_{d_1+d_2} \to \operatorname{TrP}_{d_1+d_2}$ similarly. (Here we speak of linear operators in the purely algebraic sense.)

Lemma 14.2.1. *For each n, the operators $(e^{tL_X^{(n)}/2})_{t \in \mathbb{R}}$ on TrP_d form a one-parameter group. Moreover, for $(x,y) \in M_n(\mathbb{C})_{\text{sa}}^{d_1+d_2}$ and $f \in \operatorname{TrP}_{d_1+d_2}$ and $t \geq 0$, we have*

$$[e^{tL_X^{(n)}/2} f](x,y) = \int_{M_n(\mathbb{C})_{\text{sa}}^{d_1}} f(x+z, y)\, d\sigma_t^{(n)}(z),$$

where $\sigma_t^{(n)}$ is the normalized Gaussian measure

$$d\sigma_t^{(n)}(z) = (2\pi t/n^2)^{-dn^2/2} e^{-n^2 \|z\|_2^2/2t}\, dz.$$

Proof. The operators form a one parameter group on $\operatorname{TrP}_{d_1+d_2}$ because this is true on each of the subspaces $\operatorname{TrP}_{d_1+d_2,k}$.

For the second claim, first observe by Proposition 14.1.16 that the function $f_t^{(n)}$ on $M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2}$ given by

$$f_t^{(n)}(x,y) := [e^{tL_X^{(n)}/2} f](x,y)$$

solves the heat equation

$$\partial_t f^{(n)}(x,y) = \frac{1}{2}[L_X^{(n)} f](x,y) = \frac{1}{2n^2}\Delta_x f^{(n)}(x,y).$$

The function

$$g_t^{(n)}(x,y) := \int_{M_n(\mathbb{C})_{\mathrm{sa}}^{d_1}} f(x+z,y)\, d\sigma_t^{(n)}(z)$$

also solves the heat equation for $t \geq 0$ by well-known classical results. Thus, $f_t^{(n)}$ must equal $g_t^{(n)}$ because the solution to the heat equation with initial data of polynomial growth is unique.

Alternatively, since we are dealing with polynomials, we can show $f_t^{(n)} = g_t^{(n)}$ using purely finite-dimensional linear algebra. Note that $f^{(n)} = f_0^{(n)}$ is a classical polynomial of degree $\leq k$ with respect to the coordinates given by any orthonormal basis. It follows that $g_t^{(n)}$ is also a classical polynomial of degree $\leq k$; to see this, expand $f^{(n)}(x+z,y)$ into a sum of monomials in x, y, z and then integrate with respect to $d\sigma_t^{(n)}(z)$. Therefore, $f_t^{(n)}$ and $g_t^{(n)}$ are two solutions to the heat equation in the space of classical polynomials of degree $\leq k$, and thus they must both be given by the exponential of the restriction of $\Delta/2n^2$ to the space of classical polynomials of degree $\leq k$. □

The probabilistic interpretation of the previous lemma is that if $X^{(n)}$ and $Y^{(n)}$ are random matrix tuples, and $Z^{(n)}$ is an independent Gaussian d_1-tuple of matrices, then $E[f(X^{(n)} + Z^{(n)}, Y^{(n)})|X^{(n)}, Y^{(n)}] = [e^{tL_X^{(n)}/2} f](X^{(n)}, Y^{(n)})$. In order to state the analogous result for free random variables in the large-n limit, let us first describe the non-commutative law that plays the role of the Gaussian distribution.

Definition 14.2.2. We denote by σ_t the non-commutative law of a d-tuple $(Z_1, \dots, Z_d)$ in a tracial W*-algebra such that $Z_1, \dots, Z_d$ are freely independent and each Z_j has the spectral distribution $(1/2\pi t)\sqrt{4t - x^2}1_{|x|<2t^{1/2}}\, dx$ with respect to τ.

This is well-defined by Lemmas 5.2.8 and 10.2.7. Note that this implicitly depends on d, but we have suppressed d in the notation.

Recall from §9.2.1 that the one-variable semicircle laws ν_t given by $(1/2\pi t)\sqrt{4t - x^2}1_{|x|<2t^{1/2}}\, dx$ form a semigroup under free convolution since the Voiculescu transform $\Phi_{\nu_t}(t) = t/z$. It follows that if $Z = (Z_1, \dots, Z_d) \sim \sigma_s$ and $Z' = (Z_1', \dots, Z_d') \sim \sigma_t$ are freely independent, then $Z + Z' \sim \sigma_{s+t}$. Indeed, since Z_j and Z_j' are freely independent, we have $Z_j + Z_j' \sim \nu_s \boxplus \nu_t = \nu_{s+t}$. Moreover, using the associativity properties of free independence, $Z_1 + Z_1', \dots, Z_d + Z_d'$ are freely independent, hence $Z + Z' \sim \sigma_{s+t}$. It is also easy to verify that if $Z \sim \sigma_t$, then $sZ \sim \sigma_{|s|t}$ for any $s \in \mathbb{R}$.

The following Proposition is due to [Ceb13, Theorem 2.4] for trace polynomials, although the special case of the trace of a non-commutative polynomial is more or less contained in prior work.

Proposition 14.2.3. *Let $f \in \mathrm{TrP}_{d_1+d_2}$, and let (X,Y) be a self-adjoint (d_1+d_2)-tuple from a tracial* W*-*algebra* $\mathcal{M}$. *Let Z be a freely independent d_1-tuple with the law σ_t. Then*

$$E_{\mathrm{W}^*(X,Y)}[f(X+t^{1/2}Z,Y)] = [e^{tL_X/2}f](X,Y).$$

Proof. We claim that for any $g \in \mathrm{TrP}_{d_2}$, we have

$$\frac{d}{dt}\Big|_{t=0} \tau[f(X+t^{1/2}Z,Y)g(Y)] = \frac{1}{2}\tau[L_X f(X,Y)\cdot g(Y)]. \tag{14.9}$$

First consider the case where $f(x)$ is a non-commutative polynomial. We will compute using the non-commutative Taylor expansion of f. Similar to the hash notation that we used previously, if $\mathcal{A}$ is an algebra and $w \in \mathcal{A}^{\otimes 3}$ and $b_1, b_2 \in \mathcal{A}$, we will write $w\#(b_1,b_2)$ to mean the linear extension of the map $(a_1 \otimes a_2 \otimes a_3)\#(b_1,b_2) = a_1b_1a_2b_2a_3$. Observe that

$$\begin{aligned} f(X+t^{1/2}Z,Y) = f(X,Y) + t^{1/2}\sum_{j=1}^d \mathcal{D}_{X_j}f(X,Y)\#Z_j \\ + t\sum_{j,j'=1}^d (\mathcal{D}_{X_j}\otimes \mathrm{id})\circ \mathcal{D}_{X'_j}f(X,Y)\#(Z_j,Z_{j'}) + O(t^{3/2}). \end{aligned}$$

The proof is just to compute what happens for monomials, and the method should be clear enough from our proofs in §14.1. Regarding the $O(t^{1/2})$ terms, we claim that

$$\tau\left([\mathcal{D}_{X_j}f(X,Y)\#Z_j]g(Y)\right) = 0.$$

It suffices to show that for any functions $h_1(X,Y)$ and $h_2(X,Y)$, we have

$$\tau(h_1(X,Y)Z_jh_2(X,Y)) = 0.$$

But this follows from freeness since

$$\begin{aligned} \tau(h_1(X,Y)Z_jh_2(X,Y)) &= \tau\Big([h_1(X,Y)-\tau(h_1(X,Y))]Z_j[h_2(X,Y)-\tau(h_2(X,Y))]\Big) \\ &\quad + \tau(h_1(X,Y))\tau\Big(Z_j[h_2(X,Y)-\tau(h_2(X,Y))]\Big) \\ &\quad + \tau\Big([h_1(X,Y)-\tau(h_1(X,Y))]Z_j\Big)\tau(h_2(X,Y)) \\ &= \tau(h_1(X,Y))\tau(Z_j)\tau(h_2(X,Y)), \end{aligned}$$

and $\tau(Z_j) = 0$. Then for the $O(t)$ terms, we claim that

$$\tau\Big((\mathcal{D}_{X_j} \otimes \mathrm{id}) \circ \mathcal{D}_{X'_j} f(X,Y)\#(Z_j, Z_{j'}) \cdot g(Y)\Big) = \tau\Big(\eta \circ (\mathcal{D}_{X_j} \otimes \mathrm{id}) \circ \mathcal{D}_{X'_j} f(X,Y) \cdot g(Y)\Big)\tau(Z_j Z_{j'}).$$

This follows from the fact that for any h_1, h_2, and h_3,

$$\tau\Big(h_1(X,Y)Z_j h_2(X,Y)Z_{j'} h_3(X,Y)\Big) = \tau(h_2(X,Y))\tau(h_1(X,Y)h_3(X,Y))\tau(Z_j Z_{j'}).$$

This is another direct computation using free independence and the fact that $\tau(Z_j) = \tau(Z_{j'}) = 0$. Next, because $\tau(Z_j Z_{j'})$ is 1 if $j = j'$ and 0 otherwise, we have

$$\begin{aligned}\tau(f(X + t^{1/2}Z, Y)) - \tau(f(X,Y)) &= t \sum_{j=1}^{d} \tau\Big(\eta \circ (\mathcal{D}_{X_j} \otimes \mathrm{id}) \circ \mathcal{D}_{X_j} f(X,Y) g(Y)\Big) + O(t^{3/2}) \\ &= \frac{1}{2} t L_X f(X,Y).\end{aligned}$$

Thus, (14.9) holds in the case where f is a non-commutative polynomial.

Next, consider the case of $\mathrm{tr}(p)$ for a non-commutative polynomial p. By applying our previous claim with $g = 1$, we obtain

$$\frac{d}{dt}\Big|_{t=0} \tau(p(X + t^{1/2}, Y)) = \frac{1}{2}\tau([L_X p](X,Y)) = \frac{1}{2}[L_X(\mathrm{tr}(p))](X,Y).$$

Next, using the product rule for d/dt and (14.3)

$$\frac{d}{dt}\Big|_{t=0} [\tau(p_1(X + t^{1/2}Z) \dots \tau(p_\ell(X + t^{1/2}Z)] = \frac{1}{2} L_X[\mathrm{tr}(p_1) \dots \mathrm{tr}(p_\ell)](X,Y),$$

and thus by linearity for any $h \in \mathrm{TrP}_{d_1+d_2}$, we have

$$\frac{d}{dt}\Big|_{t=0} h(X + t^{1/2}Z, Y) = \frac{1}{2}[L_X h](X,Y).$$

Finally, suppose that $f = hp$ where $h \in \mathrm{TrP}_d^0$ and $p \in \mathbb{C}\langle x_1, \dots, x_{d_1}, y_1, \dots, y_{d_2}\rangle$. Then by the product rule and (14.5),

$$\begin{aligned}&\frac{d}{dt}\Big|_{t=0} h(X + t^{1/2}Z, Y)\tau(f(X + t^{1/2}Z, Y)g(Y)) \\ &\quad= \frac{1}{2}[L_X h](X,Y)\tau(f(X,Y)g(Y)) + \frac{1}{2}h(X,Y)\tau\Big([L_X p](X,Y)g(Y)\Big) \\ &\quad= \frac{1}{2}\tau\Big([L_X f](X,Y)g(Y)\Big).\end{aligned}$$

Therefore, (14.9) holds in every case.

Next, we claim that

$$\tau(f(X + t^{1/2}Z, Y)g(Y)) = \tau([e^{tL_X/2}f](X, Y)g(Y)). \tag{14.10}$$

Since $e^{tL_X/2}$ is an invertible linear transformation on TrP_d, it suffices to prove the claim with $e^{-tL_X/2}f$ substituted for f, that is,

$$\tau([e^{-tL_X/2}f](X + t^{1/2}Z, Y)g(Y)) = \tau(f(X, Y)g(Y)).$$

It suffices to show that right-hand side is a constant function of $t \geq 0$. Consider

$$\phi(t_1, t_2) = \tau([e^{-t_1L_X/2}f](X + t_2^{1/2}Z, Y)g(Y)).$$

The coefficients of $e^{-t_1L_X/2}f$ are polynomials of t_1. Moreover, any joint moment of $(X, t_2^{1/2}Z, Y)$ is a polynomial function of $t_2^{1/2}$. Since $-t_2^{1/2}Z \sim t_2^{1/2}Z$, this polynomial function remains the same when we substitute $-t_2^{1/2}$ for $t_2^{1/2}$, and hence it contains only even powers of $t_2^{1/2}$, which makes it a polynomial in t_2. So overall $\phi(t_1, t_2)$ is a polynomial in (t_1, t_2). Hence, we can apply the chain rule and

$$\partial_t[\phi(t, t)] = \partial_{t_1}\phi(t, t) + \partial_{t_2}\phi(t, t).$$

Clearly,

$$\partial_{t_1}\phi(t_1, t_2) = -\frac{1}{2}\tau([L_X e^{-t_1L_X/2}f](X + t_2^{1/2}Z, Y).$$

For ∂_{t_2}, note that if $\tilde{Z}$ is a copy of Z freely independent from (X, Y, Z), then $t_2^{1/2}Z + \epsilon^{1/2}\tilde{Z} \sim (t_2 + \epsilon)^{1/2}Z$ and it is freely independent from (X, Y). Hence,

$$\begin{aligned}
\partial_{t_2}\phi(t_1, t_2) &= \frac{d}{d\epsilon}\Big|_{\epsilon=0} \tau([e^{-t_1L_X/2}f](X + (t_2 + \epsilon)^{1/2}Z, Y) \\
&= \frac{d}{d\epsilon}\Big|_{\epsilon=0} \tau([e^{-t_1L_X/2}f](X + t_2^{1/2}Z + \epsilon^{1/2}\tilde{Z}, Y) \\
&= \frac{1}{2}\tau([L_X e^{-t_1L_X/2}f](X + Z_{t_2}, Y),
\end{aligned}$$

where the last equality follows from applying (14.9) to the function $e^{-t_1L_X/2}f$ and the variables $(X + t_2^{1/2}Z, Y)$ and the free semicircular $\tilde{Z}$. Thus, $\partial_{t_1}\phi(t, t) + \partial_{t_2}\phi(t, t) = 0$, which proves (14.10).

By the same token, for a $(d_1 + d_2 + d_1)$-tuple (X, Y, X') and $f' \in \mathrm{Tr}_{d_1+d_2+d_1}$, we have

$$\tau(f'(X + Z_t, Y, X')g(Y, X')) = \tau([e^{tL_X/2}f'](X, Y, X')g(Y, X')). \tag{14.11}$$

There is a natural inclusion from the space of trace polynomials in a $(d_1 + d_2)$-tuple of variables (X, Y) to the space of a trace polynomials in a $(d_1 + d_2 + d_1)$-tuple (X, Y, X'), and the restriction of the operator "L_X" on trace polynomials of (X, Y, X') to the subspace of trace polynomials in (X, Y) is exactly the original operator L_X on trace polynomials of (X, Y) (this is direct verification from the definitions). Hence, $e^{tL_X/2}$ on $\mathrm{TrP}_{d_1+d_2+d_1}$ restricts to $e^{tL_X/2}$ on $\mathrm{TrP}_{d_1+d_2}$. Thus, if we substitute $f'(X, Y, X') = f(X, Y)$ in (14.11), we get

$$\tau(f(X + Z_t, Y)g(Y, X')) = \tau([e^{tL_X/2}f](X, Y)g(Y, X')] \text{ for } g \in \mathrm{TrP}_{d_2+d_1}.$$

However, there is nothing to prevent X' from being equal to X and thus

$$\tau(f(X + Z_t, Y)g(X, Y)) = \tau([e^{tL_X/2}f](X, Y)g(X, Y)] \text{ for } g \in \mathrm{TrP}_{d_1+d_2}.$$

Therefore, $[e^{tL_X/2}f](X, Y)$ is an element of $\mathrm{W}^*(X, Y)$ whose inner product with every non-commutative polynomial $g(X, Y)$ agrees with the inner product of $f(X+Z_t, Y)$ with $g(X, Y)$, and by continuity this equality extends from polynomials $g(X, Y)$ to elements of $L^2(\mathrm{W}^*(X, Y))$. This means by definition that $E_{\mathrm{W}^*(X,Y)}[f(X + Z_t, Y)] = [e^{tL_X/2}f](X, Y)$ as desired. □

The heat semigroup provides one proof of a well-known and fundamental fact in free probability, which is closely related to Voiculescu's asymptotic freeness theorem [Voi98b].

Lemma 14.2.4. *Suppose that $X^{(n)}$ is a random variable in $M_n(\mathbb{C})_{\mathrm{sa}}^{d_1}$ such that*

$$\limsup_{n\to\infty} \|X^{(n)}\|_\infty \leq R \text{ in probability}$$

and $\lambda_{X^{(n)}}$ converges in probability to λ_X in $\Sigma_{d_1,R}$ for some self-adjoint d_1-tuple from a tracial W^*-algebra. Let $Z^{(n)}$ be an independent random variable on $M_n(\mathbb{C})_{\mathrm{sa}}^{d_2}$ with the Gaussian distribution $\sigma_t^{(n)}$. Then $\lambda_{(X^{(n)},Z^{(n)})}$ converges in probability to $\lambda_{(X,Z)}$, where Z is d_2-tuple with the free semicircular distribution $\lambda_Z = \sigma_t$, freely independent of X.

Proof. Let $f \in \mathbb{C}\langle X_1, \dots, X_{d_1}, Z_1, \dots, Z_{d_2}\rangle$. By Lemma 14.2.1

$$E[\tau_n(f(X^{(n)}, Z^{(n)}))|X^{(n)}] = \tau_n\big([e^{tL_Z^{(n)}/2}f](X^{(n)}, 0)\big).$$

Because $e^{tL_Z^{(n)}/2}f$ converges coefficient-wise to $e^{tL_Z/2}f$ and $\lambda_{X^{(n)}} \to \lambda_X$ in probability, we have that in probability

$$E[\tau_n(f(X^{(n)}, Z^{(n)}))|X^{(n)}] \to \tau\big([e^{tL_Z/2}f](X, Z)\big) = \tau\big(E_{\mathrm{W}^*(Z)}[f(X, Z)]\big) = \tau(f(X, Z)).$$

Thus, to show convergence of $\lambda_{(X^{(n)},Z^{(n)})}$ to $\lambda_{(X,Z)}$ in probability, it suffices to show that

$$\tau_n(f(X^{(n)}, Z^{(n)}) - E[\tau_n(f(X^{(n)}, Z^{(n)}))|X^{(n)}] \to 0 \text{ in probability.} \tag{14.12}$$

This is a technical argument whose main idea is to swap out f for a Lipschitz function and then apply concentration of measure. Recall that $\sigma_t^{(n)}$ is given by the potential $\|z\|_2^2/2t$ and hence by Corollary 11.5.1, it satisfies the log-Sobolev inequality with constant $n^2/2t$, hence by Lemma 11.5.2, we have

$$P\left(\|Z_j^{(n)} - 0\|_\infty \geq t^{1/2}(\Theta + \delta)\right) \leq e^{-n\delta^2/2}.$$

Thus, letting $R_1 > \max(R, t^{1/2}\Theta)$, we have

$$\limsup_{n\to\infty} \|(X^{(n)}, Z^{(n)})\|_\infty < R \text{ in probability.}$$

Let $\phi \in C_c^\infty(\mathbb{R})$ such that $\phi(t) = t$ for $|t| \leq R_1$. Then we have in probability that

$$\tau_n(f(X^{(n)}, Z^{(n)}) - \tau_n(f(\phi(X^{(n)}), \phi(Z^{(n)}))) \to 0, \tag{14.13}$$

since the two expressions are equal when $\|(X^{(n)}, Z^{(n)})\|_\infty \leq R_1$. By Lemma 13.3.6, $(x, z) \mapsto \operatorname{tr}(f(\phi(x), \phi(z))$ is globally Lipschitz in $\|\cdot\|_2$ with some Lipschitz constant K. Therefore, by Herbst's concentration inequality,

$$P\left(\left|\tau(f(\phi(X^{(n)}), \phi(Z^{(n)}))) - E[\tau(f(\phi(X^{(n)}), \phi(Z^{(n)})))|X^{(n)}]\right| \geq \delta \middle| X^{(n)}\right) \leq 4e^{-\delta^2 n^2/4tK^2},$$

and hence in probability

$$\tau_n(f(\phi(X^{(n)}), \phi(Z^{(n)}))) - E[\tau(f(\phi(X^{(n)}), \phi(Z^{(n)})))|X^{(n)}] \to 0. \tag{14.14}$$

Since f is a non-commutative polynomial and $f(0,0)$ is a constant, we have

$$|\tau(f(X^{(n)}, Z^{(n)})| \leq K_1 e^{K_2\|X^{(n)}\|_\infty + K_2\|Z^{(n)}\|_\infty}$$

for some constants K_1 and K_2, and by the same token $f(\phi(X^{(n)}), \phi(Z^{(n)}))$ is bounded by some K_1' since ϕ is bounded. Thus, by Corollary 11.5.3 applied to the Gaussian distribution,

$$E[1_{\|Z^{(n)}\|_\infty > R_1}\tau(f(\phi(X^{(n)}), Z^{(n)}))|X^{(n)}] - E[1_{\|Z^{(n)}\|_\infty > R_1}\tau(f(\phi(X^{(n)}), \phi(Z^{(n)})))|X^{(n)}] \to 0$$

in probability. Since $\phi(Z^{(n)}) = Z^{(n)}$ when $\|Z^{(n)}\|_\infty \leq R_1$, this means that

$$E[\tau(f(\phi(X^{(n)}), Z^{(n)}))|X^{(n)}] - E[\tau(f(\phi(X^{(n)}), \phi(Z^{(n)})))|X^{(n)}] \to 0 \text{ in probability.} \tag{14.15}$$

Finally, because $\|X^{(n)}\|_\infty$ is eventually $\leq R_1$ in probability,

$$E[\tau(f(\phi(X^{(n)}), Z^{(n)}))|X^{(n)}] - E[\tau(f(X^{(n)}, Z^{(n)}))|X^{(n)}] \to 0 \text{ in probability.} \tag{14.16}$$

Combining (14.13), (14.14), (14.15), and (14.16) proves (14.12) and hence completes the proof. □

Our next goal is to show that $(e^{tL_X/2})_{t\geq 0}$ extends to a semigroup $(\Phi_t)_{t\geq 0}$ on $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$, and that it describes the large-n limit of the heat semigroup $(\Phi_t^{(n)})$ on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ given by Gaussian convolution.

Lemma 14.2.5. *For each $t \geq 0$ there exists a continuous linear transformation Φ_t of $C_{\mathrm{tr,app}}(\mathbb{R}^{*(d_1+d_2)}, \|\cdot\|_2)$ such that $\Phi_t f = e^{tL_X/2} f$ for $f \in \mathrm{TrP}_d$. Moreover, $(\Phi_t)_{t\geq 0}$ is a semigroup, and we have*

$$\|\Phi_t f\|_{2,R} \leq \|f\|_{2,R+2t^{1/2}}, \qquad \|\Phi_t f\|_{\infty,R} \leq \|f\|_{\infty,R+2t^{1/2}}.$$

Proof. Suppose that $f \in \mathrm{TrP}_{d_1+d_2}$, and let (X, Y) be a self-adjoint $(d_1 + d_2)$-tuple from $(\mathcal{M}, \tau)$ with $\lambda_{(X,Y)} \in \Sigma^{\mathrm{app}}_{d_1+d_2,R}$. Let Z be a freely independent d_1-tuple with law σ_t. Note that $\|Z\|_\infty = 2t^{1/2}$.

We claim that $\lambda_{(X,Y,Z)} \in \Sigma_{d_1+d_2+d_1,R+2t^{1/2}}$. Indeed choose a sequence of deterministic tuples of $n(k) \times n(k)$ matrices $(X^{(k)}, Y^{(k)})$ converging to (X, Y) in non-commutative law. Then proceeding as in the proof of Lemma 13.5.5, we may modify them to get a sequence of $(X^{(n)}, Y^{(n)}) \in M_n(\mathbb{C})^{d_1+d_2}_{\mathrm{sa}}$ converging in non-commutative law to (X, Y). Let $Z^{(n)}$ be random Gaussian d_1-tuple. Then by Lemma 14.2.4, the non-commutative law of $(X^{(n)}, Y^{(n)}, Z^{(n)})$ converges in probability to that of (X, Y, Z), so that $\lambda_{(X,Y,Z)} \in \Sigma^{\mathrm{app}}_{d,R'}$ for some R'. By Proposition 14.2.3, we have

$$\|[e^{tL_X/2} f](X, Y)\|_2 = \|E_{\mathrm{W}^*(X,Y)}[f(X + Z, Y)]\|_2 \leq \|f(X + Z, Y)\|_2.$$

Since $\lambda_{(X+Z,Y)} \in \Sigma^{\mathrm{app}}_{d,2R'}$, and since $\|X + Z\|_\infty \leq R + 2t^{1/2}$, we have by Lemma 13.3.6 that $\lambda_{(X+Z,Y)} \in \Sigma^{\mathrm{app}}_{d,R+2t^{1/2}}$. Thus, by the same lemma,

$$\|f(X + Z, Y)\|_2 \leq \|f\|_{2,R+2t^{1/2}}.$$

Therefore, since $\lambda_{(X,Y)}$ was arbitrary in $\Sigma^{\mathrm{app}}_{d_1+d_2,R}$,

$$\|e^{tL_X/2} f\|_{2,R} \leq \|f\|_{2,R+2t^{1/2}}.$$

This bound implies that $e^{tL_X/2}$ extends uniquely to a continuous linear operator Φ_t on $C_{\mathrm{tr,app}}(\mathbb{R}^{*(d_1+d_2)}, \|\cdot\|_2)$, which still satisfies the same bounds. Finally, to show $(\Phi_t)_{t\geq 0}$ is a semigroup, we observe that $\Phi_s\Phi_t$ is a continuous linear operator satisfying $\Phi_s\Phi_t f = e^{(s+t)L_X/2} f$ when f is a trace polynomial, and hence $\Phi_s\Phi_t = \Phi_{s+t}$. □

Finally, we can show that the operation of convolution with Gaussian measure preserves asymptotic approximability by trace polynomials [Jek18, Lemma 3.28]. This is one of the main technical tools underlying our results in the rest of the thesis.

Lemma 14.2.6. *For $f^{(n)} : M_n(\mathbb{C})^{d_1+d_2}_{\mathrm{sa}} \to M_n(\mathbb{C})$, let us denote*

$$\Phi_t^{(n)} f^{(n)}(x, y) := \int f^{(n)}(x + z, y)\, d\sigma_t^{(n)}(z)$$

whenever the integral is defined. Suppose there exist constants K_1 and K_2 such that

$$\|f^{(n)}(x, y)\|_2 \leq K_1 e^{K_2(\|x\|_\infty+\|y\|_\infty)} \text{ for } (x, y) \in M_n(\mathbb{C})^{d_1+d_2}_{\mathrm{sa}} \text{ for all } n.$$

If $f^{(n)} \rightsquigarrow f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*(d_1+d_2)}, \|\cdot\|_2)$, then $\Phi_t^{(n)} f^{(n)} \rightsquigarrow \Phi_t f$.

Proof. Fix $t > 0$, $R > 0$, $\epsilon > 0$. Let Θ be the constant from Lemma 11.5.2. By Lemma 13.5.3, there is a trace polynomial g such that

$$\lim_{n\to\infty} \|f^{(n)} - g\|^{(n)}_{2,R} = \|f - g\|_{2,R+t^{1/2}\Theta} \leq \epsilon.$$

Clearly, $\|g(x,y)\|_2 \leq K_1' e^{K_2(\|x\|_\infty + \|y\|_\infty)}$ for some constant K_1'. By Corollary 11.5.3,

$$\begin{aligned}\int_{\|z\|_\infty > t^{1/2}\Theta} &\|f^{(n)}(x+z,y) - g(x+z,y)\|_2 \, d\sigma_t^{(n)}(z) \\ &\leq \int_{\|z\|_\infty > t^{1/2}\Theta} (K_1 + K_1') e^{K_2(\|x\|_\infty + \|y\|_\infty)} e^{K_2 \|z\|_\infty} \, d\sigma_t^{(n)}(z) \\ &\leq (K_1 + K_1') e^{K_2(\|x\|_\infty + \|y\|_\infty)} d \sqrt{\frac{2\pi}{nc}} e^{K_2^2/nc} e^{t^{1/2}\Theta}.\end{aligned}$$

But $\|z\|_\infty \leq t^{1/2}\Theta$ and $\|x\|_\infty, \|y\|_\infty \leq R$, we have

$$\|f^{(n)}(x+z,y) - g(x+z,y)\|_2 \leq \|f^{(n)} - g\|_{2,R+t^{1/2}\Theta}^{(n)},$$

and hence

$$\|\Phi_t^{(n)} f^{(n)}(x,y) - \Phi_t^{(n)} g(x,y)\|_2 \leq \|f^{(n)} - g\|_{2,R+t^{1/2}\Theta}^{(n)} + (K_1 + K_1') e^{2K_2 R} d \sqrt{\frac{2\pi}{nc}} e^{K_2^2/nc} e^{t^{1/2}\Theta}.$$

Therefore,

$$\limsup_{n\to\infty} \|\Phi_t^{(n)} f^{(n)} - \Phi_t^{(n)} g\|_{2,R}^{(n)} \leq \limsup_{n\to\infty} \|f^{(n)} - g\|_{2,R+t^{1/2}\Theta}^{(n)} \leq \epsilon.$$

Clearly, $\Phi_t^{(n)} g \rightsquigarrow \Phi_t g$ since $e^{tL_X^{(n)}/2} g \to e^{tL_X/2} g$ coefficient-wise. Since Θ can be taken ≥ 2,

$$\|\Phi_t f - \Phi_t g\|_{2,R} \leq \|f - g\|_{2,R+2t^{1/2}} \leq \|f - g\|_{2,R+t^{1/2}\Theta} \leq \epsilon.$$

Because

$$\|\Phi_t^{(n)} f^{(n)} - \Phi_t f\|_{2,R}^{(n)} \leq \|\Phi_t^{(n)} f^{(n)} - \Phi_t^{(n)} g\|_{2,R}^{(n)} + \|\Phi_t^{(n)} g - \Phi_t g\|_{2,R}^{(n)} + \|\Phi_t g - \Phi_t f\|_{2,R},$$

we have

$$\limsup_{n\to\infty} \|\Phi_t^{(n)} f^{(n)} - \Phi_t f\|_{2,R}^{(n)} \leq 2\epsilon.$$

Since R and ϵ were arbitrary, $\Phi_t^{(n)} f^{(n)} \rightsquigarrow \Phi_t f$. □

14.3 ODE for non-commutative variables

Several times in our study of partial differential equations, we will use flows along vector fields given by functions in $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$, and by asymptotically approximable sequences of functions on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ (see the proofs of Lemma 15.5.1 and Proposition 17.5.1).

Roughly speaking, given $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\mathrm{sa}}^d$ and given $H_t \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\mathrm{sa}}^d$ for $t \in [0,T]$, we want to construct $F_t \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\mathrm{sa}}^d$ satisfying

$$F_0 = f, \qquad \partial_t F_t = H_t \circ F_t. \tag{14.17}$$

Similarly, if we are given $H_t^{(n)}, f^{(n)} : M_n(\mathbb{C})_{\text{sa}}^d \to M_n(\mathbb{C})_{\text{sa}}^d$ with $H_t^{(n)} \rightsquigarrow H_t$ and $f^{(n)} \rightsquigarrow f$, then we want to show that the corresponding solutions $F_t^{(n)}$ are asymptotic to F_t. It will be convenient for the sake of brevity to list the assumptions that we make about the vector field H_t.

Assumption 14.3.1. We are given a function $H_t \in C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\text{sa}}^m$ for each $t \in [0,T]$ such that

(A) The mapping $t \mapsto H_t$ is a continuous function $[0,T] \to C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\text{sa}}^m$ with respect to the Fréchet topology on $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\text{sa}}^m$.

(B) There is a constant K such that $\|H_t\|_{\text{Lip}} \leq K$ for all t.

Remark 14.3.2. Condition (A) means more explicitly that $t \mapsto H_t$ is continuous with respect to the seminorm $\|\cdot\|_{2,R}$ for each $R > 0$. Since $[0,T]$ is compact, this upgrades automatically to uniform continuity. Hence, (A) is equivalent to the statement that for every $\epsilon > 0$ and $R > 0$, there exists a $\delta > 0$ such that for all $s, t \in [0,T]$,

$$|s - t| < \delta \implies \|H_s - H_t\|_{2,R} < \epsilon.$$

It should be no surprise that we will solve (14.17) by Picard iteration. We first verify that Assumption 14.3.1 is preserved under the composition and integration operations used to define Picard iterates.

Lemma 14.3.3. *Let H_t satisfy Assumption 14.3.1 for some Lipschitz constant K and let $g \in C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\text{sa}}^m$ be globally Lipschitz. Then*

$$G_t := g + \int_0^t H_s \, ds$$

is well-defined by Riemann integration and $(G_t)_{t \in [0,T]}$ also satisfies Assumption 14.3.1, except with the Lipschitz constant $KT + \|g\|_{\text{Lip}}$. Moreover, we have in the Fréchet topology on $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ that

$$\frac{d}{dt} G_t = H_t.$$

Proof. First, consider the case $g = 0$. The argument is completely standard, but we will sketch it since one does not often need to integrate functions with values in abstract Fréchet spaces. First, let $\omega_R : [0,\infty) \to [0,\infty)$ be a modulus of continuity for $t \mapsto H_t$ in $\|\cdot\|_{2,R}$, so that

$$\|H_s - H_t\|_{2,R} \leq \omega_R(t).$$

For a partition $\mathcal{P} = \{t_0, \dots, t_k\}$ of $[0,t]$, let $\operatorname{mesh}(\mathcal{P}) = \max_j |t_j - t_{j-1}|$. Define the right Riemann sum

$$G_{\mathcal{P}} = \sum_{j=1}^k (t_j - t_{j-1}) G_{t_j}.$$

If $\mathcal{Q}$ is a refinement of $\mathcal{P}$, then

$$|G_{\mathcal{P}} - G_{\mathcal{Q}}| \leq t\omega_R(\mathrm{mesh}(\mathcal{P}));$$

the proof is to replace each G_t for $t \in \mathcal{Q}$ by $G_{t'}$ where t' is the right endpoint of the interval in $\mathcal{P}$ containing t, note $|t - t'| \leq \mathrm{mesh}(\mathcal{P})$, and then estimate with the definition of ω_R.

We can define G_t to be the limit over all partitions $\mathcal{P}$ of $[0,t]$ as $\mathrm{mesh}(\mathcal{P}) \to 0$, since this net is Cauchy with respect to $\|\cdot\|_{2,R}$ for each R, since $\lim_{t\to 0^+} \omega_R(t) = 0$. Moreover,

$$\|G_t - G_{\mathcal{P}}\|_{2,R} \leq t\omega_R(\mathrm{mesh}(\mathcal{P})).$$

If $s \leq t$, then the same holds for $G_t - G_s$ using partitions of $[s,t]$. In particular, if we use the partition $\{s,t\}$, we obtain

$$\|G_t - G_s - (t-s)H_t\|_{2,R} \leq (t-s)\omega_R(t-s).$$

This implies also that

$$\|G_t - G_s - (t-s)H_s\|_{2,R} \leq 2(t-s)\omega_R(t-s),$$

since $|H_s - H_t| \leq \omega_R(t-s)$. Overall, for any s, t,

$$\left\| \frac{G_t - G_s}{t-s} - H_t \right\|_{2,R} \leq 2\omega_R(t-s),$$

which goes to zero as $t - s \to 0$, and hence proves differentiability.

Now we check that G_t satisfies Assumption 14.3.1.

(A) The above argument shows that

$$\|G_t - G_s\|_{2,R} \leq (t-s)\|H_s\|_{2,R} + 2\omega_R(t-s),$$

which goes to zero as $t \to s$.

(B) Since the Riemann sums $G_{\mathcal{P}}$ for partitions of $[0,t]$ converge to G_t in $\|\cdot\|_{2,R}$, the evaluation of these Riemann sums on matrix tuples also converges to the evaluation of G_t on the same matrix tuples. Hence, for $X, Y \in M_n(\mathbb{C})^d_{\mathrm{sa}}$,

$$G_t(X) - G_t(Y) = \int_0^t [H_s(X) - H_s(Y)]\, ds,$$

so that $\|G_t\|_{\mathrm{Lip}} \leq Kt \leq KT$, where K is the given bound on $\sup_t \|H_t\|_{\mathrm{Lip}}$.

Now if $g \neq 0$, we just add g to the functions described above. The modulus of continuity of $t \mapsto G_t$ is unchanged by adding the constant g, and (B) is still true with Lipschitz constant $KT + \|g\|_{\mathrm{Lip}}$ because we assumed g is Lipschitz. □

Lemma 14.3.4. *If $(H_t)_{t\in[0,T]}$ and $(G_t)_{t\in[0,T]}$ satisfy Assumption 14.3.1 with $d = m$ and with Lipschitz constants K_1 and K_2 respectively, then $(H_t \circ G_t)_{t\in[0,T]}$ satisfies it with Lipschitz constant K_1K_2.*

Proof. Because we assumed that H_t and G_t are Lipschitz, $H_t \circ G_t$ is well-defined by Proposition 13.4.3. And clearly, $\|H_t \circ G_t\|_{\mathrm{Lip}} \leq \|H_t\|_{\mathrm{Lip}}\|G_t\|_{\mathrm{Lip}}$, so (B) holds for $H_t \circ G_t$ with Lipschitz constant K_1K_2.

To check (A), note that

$$\|H_s \circ G_s - H_s \circ G_t\|_{2,R} \leq K_1\|G_s - G_t\|_{2,R}.$$

By Lemma 13.3.4 (3), we have

$$\|G_t\|_{\infty,R} \leq \|G_t\|_{2,R} + 2\sqrt{2}d^{1/2}\Theta R\|G_t\|_{\mathrm{Lip}}.$$

Since $t \mapsto G_t$ is continuous and $[0,T]$ is compact, $\|G_t\|_{2,R}$ is bounded by some constant M_R. So letting $R_1 = M_R + 2\sqrt{2}d^{1/2}\Theta RK_2$, we have

$$\|H_s \circ G_t - H_t \circ G_t\|_{2,R} \leq \|H_s - H_t\|_{2,R_1}.$$

Therefore,

$$\|H_s \circ G_s - H_t \circ G_t\|_{2,R} \leq K_1\|G_s - G_t\|_{2,R} + \|H_s - H_t\|_{2,R_1},$$

which demonstrates (B). □

Proposition 14.3.5. *Let $(H_t)_{t\in[0,T]}$ satisfy Assumption 14.3.1, and let $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ be Lipschitz. Then there exists a unique $(F_t)_{t\in[0,T]}$ such that*

$$\frac{d}{dt}F_t = H_t \circ F_t \qquad\qquad F_0 = f.$$

Moreover, $(F_t)_{t\in[0,T]}$ satisfies Assumption 14.3.1.

Proof. We define the Picard iterates $F_{t,\ell}$ inductively by

$$F_0 = f,$$
$$F_{\ell+1} = f + \int_0^t H_s \circ F_{s,\ell}\, ds.$$

The previous two lemmas imply that $(F_{t,\ell})_{t\in[0,T]}$ is well-defined and satisfies Assumption 14.3.1.

By the same token, $(H_t \circ g)_{t\in[0,T]}$ satisfies Assumption 14.3.1, and hence $M_R := \sup_{t\in[0,T]}\|H_t \circ g\|_{2,R}$ is finite. Note that

$$\|F_{t,1} - F_{t,0}\|_{2,R} \leq M_R t.$$

We have assumed $\|H_t\|_{\text{Lip}}$ is bounded by some constant K and hence

$$\begin{aligned}\|F_{t,\ell+1} - F_{t,\ell}\|_{2,R} &\leq \int_0^t \|H_s \circ F_{s,\ell} - H_s \circ F_{s,\ell-1}\|_{2,R}\, ds \\ &\leq K \int_0^t \|F_{s,\ell} - F_{s,\ell-1}\|_{2,R}.\end{aligned}$$

From here a straightforward induction on ℓ shows that for $\ell \geq 1$,

$$\|F_{t,\ell} - F_{t,\ell-1}\|_{2,R} \leq \frac{M_R K^{\ell-1} t^\ell}{\ell!}$$

because $K \int_0^t K^{\ell-1} s^\ell/\ell!\, ds = K^\ell s^{\ell+1}/(\ell+1)!$. Now because $\sum_{\ell=1}^\infty K^{\ell-1} s^\ell/\ell!$ converges, we know that $(F_{t,\ell})_{\ell\in\mathbb{N}}$ converges to some $F_t \in C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)^d_{\text{sa}}$, and in fact

$$\|F_{t,\ell} - F_t\|_{2,R} \leq M_R \sum_{j=\ell+1}^\infty \frac{K^{j-1}T^j}{j!}.$$

Since the convergence is uniform, it follows that $t \mapsto F_t$ is continuous, and hence $(F_t)_{t\in[0,T]}$ satisfies (A) of Assumption 14.3.1.

To check (B), we claim that $\|F_{t,\ell}\| \leq e^{Kt}\|f\|_{\text{Lip}}$ for all t and ℓ. We proceed by induction. The base case $F_{t,0} = f$ is trivial. For the induction step, for matrix tuples X and Y,

$$\begin{aligned}\|F_{t,\ell+1}(X) - F_{t,\ell+1}(Y)\|_2 &\leq \|f(X) - f(Y)\|_2 + \int_0^t \|H_s \circ F_{s,\ell}(X) - H_s \circ F_{s,\ell}(Y)\|_2\, ds \\ &\leq \|f(X) - f(Y)\|_2 - \int_0^t K \cdot e^{Kt}\|f\|_{\text{Lip}}\, ds \\ &\leq \|f\|_{\text{Lip}} + (e^{Kt} - 1)\|f\|_{\text{Lip}},\end{aligned}$$

so that $\|F_{t,\ell+1}\|_{\text{Lip}} \leq e^{Kt}\|f\|_{\text{Lip}}$. Then it is easy to see that

$$\|F_t\|_{\text{Lip}} \leq \liminf_{n\to\infty} \|F_{t,\ell}\|_{\text{Lip}} \leq e^{Kt} \leq e^{KT},$$

hence $(F_t)_{t\in[0,T]}$ satisfies Assumption 14.3.1 with Lipschitz constant e^{Kt}.

Because $(F_t)_{t\in[0,T]}$ satisfies Assumption 14.3.1, the integral $\int_0^t H_s \circ F_s\, ds$ is well-defined. Given our estimates for $F_{t,\ell} - F_t$, it is straighforward to check that

$$f + \int_0^t H_s \circ F_s\, ds = f + \lim_{\ell\to\infty} \int_0^t H_s \circ F_{s,\ell}\, ds = \lim_{\ell\to\infty} F_{t,\ell+1} = F_t.$$

Because $F_t = f + \int_0^t H_s \circ F_s\, ds$, it follows from Lemma 14.3.3 that $(F_t)_{t\in[0,T]}$ solves the initial value problem.

Finally, $(F_t)_{t\in[0,T]}$ is unique, even if we do not assume it satisfies Assumption 14.3.1, because for every n the restriction of F_t to $M_n(\mathbb{C})^d_{\text{sa}}$ is uniquely determined by the classical existence and uniquenesss theory for ODE. □

Now that we have developed the Picard-Lindelöf theorem for $C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$, let us explain the "asymptotic approximation" version of these results. First, the version of Assumption 14.3.1 for a sequence of functions on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ is as follows.

Assumption 14.3.6. We are given $T > 0$ and for each $n \in \mathbb{N}$ a function $H_t^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})_{\mathrm{sa}}^m$ such that

(A) For each t, the sequence $(H_t^{(n)})_{n \in \mathbb{N}}$ is asymptotically approximable by trace polynomials.

(B) There is some constant K such that $\|H_t^{(n)}\|_{\mathrm{Lip}} \leq K$ for all n and t.

(C) For every $R > 0$ and for every $\epsilon > 0$, there exists $\delta > 0$, such that

$$|s - t| < \delta \implies \sup_{n \in \mathbb{N}} \left\| H_s^{(n)} - H_t^{(n)} \right\|_{2,R}^{(n)} < \epsilon.$$

Proposition 14.3.7. *Let $(H_t^{(n)})_{t \in [0,T], n \in \mathbb{N}}$ satisfy Assumption 14.3.6 with $m = d$, and let $(f^{(n)})_{n \in \mathbb{N}}$ be uniformly Lipschitz sequence of unitarily equivariant functions that is asymptotically approximable by trace polynomials.*

(1) Let $H_t^{(n)} \rightsquigarrow H_t \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\mathrm{sa}}^d$ and $f^{(n)} \rightsquigarrow f$. Then $(H_t)_{t \in [0,T]}$ satisfies Assumption 14.3.1.

(2) For each n, there is a unique family of functions $F_t^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})_{\mathrm{sa}}^d$ for $t \in [0,T]$ satisfying

$$\frac{d}{dt} F_t^{(n)} = H_t^{(n)} \circ F_t^{(n)}, \qquad F_0^{(n)} = f.$$

(3) If $(F_t)_{t \in [0,T]}$ is the solution for the family $(H_t)_{t \in [0,T]}$ and f given by Proposition 14.3.5, then $F_t^{(n)} \rightsquigarrow F_t$.

Proof. (1) The existence of H_t follows from Lemma 13.5.3, and by the same lemma,

$$\|H_s - H_t\|_{2,R} = \lim_{n \to \infty} \|H_s^{(n)} - H_t^{(n)}\|_{2,R},$$

which shows that $(H_t)_{t \in [0,T]}$ satisfies Assumption 14.3.1 (A). Moreover, (B) follows since $\|H_t\|_{\mathrm{Lip}} \leq K$ by Lemma 13.5.8.

(2) The existence and uniqueness of the solution for each n follows the classical Picard-Lindelöf theorem. In fact, the explicit Picard iterates and estimates are exactly analogous to the proof of Proposition 14.3.5 above.

(3). Fix $R > 0$. Using the integral equation for both $F_t^{(n)}$ and F_t,

$$\left\| F_t^{(n)} - F_t \right\|_{2,R}^{(n)} \leq \|f^{(n)} - f\|_{2,R} + \int_0^t \left\| H_s^{(n)} \circ F_s^{(n)} - H_s \circ F_s \right\|_{2,R}^{(n)} ds.$$

Note that

$$\left\|H_s^{(n)} \circ F_s^{(n)} - H_s^{(n)} \circ F_s\right\|_{2,R}^{(n)} \leq K \|F_s^{(n)} - F_s\|_{2,R}^{(n)}.$$

Now because F_t is $e^{tK}\|f\|_{\mathrm{Lip}}$-Lipschitz, we have

$$\|F_t\|_{\infty,R} \leq \|F_t\|_{2,R} + 2\sqrt{2}d^{1/2}\Theta e^{TK}\|f\|_{\mathrm{Lip}}R =: R_1.$$

Hence,

$$\left\|H_s^{(n)} \circ F_s - H_s \circ F_s\right\|_{2,R}^{(n)} \leq \|H_s^{(n)} - H_s\|_{2,R_1}^{(n)}.$$

Let

$$M^{(n)} = \sup_{s \in [0,T]} \|H_s^{(n)} - H_s\|_{2,R_1}^{(n)}.$$

Using Assumption 14.3.1 (A) and Assumption 14.3.6 (B), the functions $t \mapsto \|H_t^{(n)} - H_t\|_{2,R}^{(n)}$ are equicontinuous for $n \in \mathbb{N}$. These functions converge to zero pointwise since $H_t^{(n)} \rightsquigarrow H_t$. Therefore, they converge to zero uniformly and hence

$$\lim_{n \to \infty} M^{(n)} = 0.$$

In the above integral formula,

$$\left\|F_t^{(n)} - F_t\right\|_{2,R}^{(n)} \leq \|f^{(n)} - f\|_{2,R} + M_{(n)} + \int_0^t K \|F_s^{(n)} - F_s\|_{2,R}^{(n)}\, ds.$$

It follows from the integral form of Grönwall's inequality that

$$\left\|F_t^{(n)} - F_t\right\|_{2,R}^{(n)} \leq \left(\|f^{(n)} - f\|_{2,R} + M_{(n)}\right) e^{Kt}.$$

Since $\|f^{(n)} - f\|_{2,R} \to 0$ and $M^{(n)} \to 0$ as $n \to \infty$, we are done. □

Remark 14.3.8. In Proposition 14.3.7, one can show with a little more work that the family $(F_t^{(n)})_{t \in [0,T], n \in \mathbb{N}}$ satisfies Assumption 14.3.6. One way to do this is to simply repeat the arguments used to prove Lemma 14.3.3, Lemma 14.3.4, and Proposition 14.3.5.

In §15 and §17, we will want to study initial value problems of the form

$$\partial_t F_t^{(n)}(x,y) = H_t(F_t^{(n)}(x,y), y) \qquad F_0^{(n)}(x,y) = x. \tag{14.18}$$

Thus, we present the previous result in the following form.

Corollary 14.3.9. *Let $H_t^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d+2} \to M_n(\mathbb{C})_{\mathrm{sa}}^{d_1}$ satisfy Assumption 14.3.6. Let $F_t^{(n)}$ be the solution to* (14.18)*. Then $(F_t^{(n)})_{n \in \mathbb{N}}$ is asymptotically approximable by trace polynomials for every n.*

Proof. Consider the vector fields in $d_1 + d_2$ variables given by

$$\tilde{H}_t^{(n)}(x, y) = (H_t(x, y), 0)$$
$$\tilde{F}_t^{(n)}(x, y) = (F_t(x, y), y).$$

Then $\tilde{F}_t^{(n)}$ solves the initial value problem

$$\frac{d}{dt}\tilde{F}_t^{(n)} = \tilde{H}_t \circ \tilde{F}_t, \qquad \tilde{F}_0^{(n)} = \mathrm{id}\,.$$

Moreover, $\tilde{H}_t^{(n)}$ clearly satisfies Assumption 14.3.6. Therefore, by the previous proposition $(\tilde{F}_t^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials for each t, hence the same holds for $(F_t^{(n)})_{n\in\mathbb{N}}$. □

CHAPTER 15

Results: Conditional expectation for free Gibbs states

15.1 Main result and consequences

Our work concerns the large n behavior of random d-tuples of self-adjoint matrices. Recall $M_n(\mathbb{C})_{\mathrm{sa}}^d$ denotes the space of d-tuples of $n \times n$ self-adjoint matrices with the inner product $\langle x, y \rangle = \sum_{j=1}^d \tau_n(x_j y_j)$ where τ_n is the normalized trace, which can be isometrically identified with $\mathbb{R}^{dn^2}$.

We consider a probability measure of the form

$$d\mu^{(n)}(x) = e^{-n^2 V^{(n)}(x)}\, dx,$$

where $V^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{R}$. Past work such as [GM06, GS09, GS14, DGS16] has focused on the case where $V^{(n)}(x) = \tau_n(p(x))$ for some non-commutative polynomial p or more generally a trace polynomial (or power series defined by trace polynomials). For multiple matrices, the most progress has been made when $V^{(n)}$ is uniformly convex.

In the present work, our assumptions will be more flexible in some ways but more restrictive in others. Rather than assume that $V^{(n)}$ is a trace polynomial, we only assume $\nabla V^{(n)}$ is asymptotically approximable by trace polynomials, or $\nabla V^{(n)} \rightsquigarrow F$. However, we will assume not only uniform convexity, but also semi-concavity, as this is technically convenient to get easy estimates. While the class of trace polynomials that satisfy global convexity and semi-concavity is very restrictive, there are many functions whose gradient is *approximated* by trace polynomials that still satisfy our assumptions.

Our standard assumptions are as follows:

Assumption 15.1.1. We are given $0 < c \leq C$ and for each n a function $V^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{R}$ such that

(1) $HV^{(n)} \geq cI$, that is, $V^{(n)}(x) - \frac{1}{2}c\|x\|_2^2$ is convex.

(2) $HV^{(n)} \leq CI$, that is, $V^{(n)}(x) - \frac{1}{2}C\|x\|_2^2$ is concave.

(3) $\{\nabla V^{(n)}\}_{n \in \mathbb{N}}$ is asymptotically approximable by trace polynomials.

(4) $V^{(n)}$ is invariant under unitary conjugation, that is, $V^{(n)}(uxu^*) = V^{(n)}(x)$ for any unitary u.

We denote by $\mu^{(n)}$ the probability measure on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ associated to the potential $n^2 V^{(n)}$.

The main technical result of the section is as follows:

Theorem 15.1.2. *Let $V^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to \mathbb{R}$ be a sequence of functions satisfying Assumption 15.1.1. Let $(f^{(n)})_{n \in \mathbb{N}}$ be a sequence of functions $M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to M_n(\mathbb{C})$ that is asymptotically approximable by trace polynomials and is uniformly Lipschitz, that is, $\sup_n \|f^{(n)}\|_{\mathrm{Lip}} < +\infty$. Let*

$$g^{(n)}(y) = \frac{\int_{M_n(\mathbb{C})_{\mathrm{sa}}^{d_1}} f^{(n)}(x,y) e^{-n^2 V^{(n)}(x,y)}\, dx}{\int_{M_n(\mathbb{C})_{\mathrm{sa}}^{d_1}} e^{-n^2 V^{(n)}(x,y)}\, dx}.$$

Then $(g^{(n)})_{n \in \mathbb{N}}$ is also asymptotically approximable by trace polynomials in y and uniformly Lipschitz with

$$\left\|g^{(n)}\right\|_{\mathrm{Lip}} \leq (1 + C/c) \left\|f^{(n)}\right\|_{\mathrm{Lip}}$$

Remark 15.1.3. The function $g^{(n)}$ is obtained from $f^{(n)}$ by taking a conditional expectation. Indeed, let $(X^{(n)}, Y^{(n)})$ be a random variable distributed according to the probability measure $\mu^{(n)}$ given by the potential $n^2 V^{(n)}$, where $X^{(n)}$ takes values in $M_n(\mathbb{C})_{\mathrm{sa}}^{d_1}$ and $Y^{(n)}$ takes values in $M_n(\mathbb{C})_{\mathrm{sa}}^{d_2}$. Then $g^{(n)}(Y^{(n)}) = E[f^{(n)}(X^{(n)}, Y^{(n)}) | Y^{(n)}]$.

Remark 15.1.4. We can take $d_2 = 0$ in the theorem, and then $g^{(n)}$ will be a constant (a function of 0 variables), and will also be scalar-valued (that is, a scalar multiple of the identity matrix) because we assumed that $\mu^{(n)}$ is unitarily invariant. Then we interpret the statement that $g^{(n)}$ is asymptotically approximable by trace polynomials to mean that the limit as $n \to \infty$ of $g^{(n)}$ exists.

While the proof of Theorem 15.1.2 will occupy most of the chapter, let us first explain special cases and consequences of the theorem. The first result is that these random matrix models have a large-n limit in a suitable sense. The following is from [Jek18, Theorem 4.1], and is comparable to earlier results such as [GS09, Theorem 4.4], [DGS16, Proposition 50 and Theorem 51], [Dab17, Theorem 4.4].

Theorem 15.1.5. *Let $V^{(n)}$ and $\mu^{(n)}$ be as in Assumption 15.1.1. Let $X^{(n)}$ be a random variable distributed according to the law $\mu^{(n)}$. Then $\lambda_{X^{(N)}}$ converges almost surely to some non-commutative law λ, which can be represented as the non-commutative law of some self-adjoint d-tuple X from a tracial von Neumann algebra $(\mathcal{M}, \tau)$. The convergence in the large-n limit satisfies the following properties.*

(1) Fix $R > \|X\|_\infty$ and let $\mathcal{U}$ be a neighborhood of $\lambda_X = \lambda$ in $\Sigma_{d,R}$. Then

$$\limsup_{n \to \infty} \frac{1}{n^2} \log P(\|X^{(n)}\|_\infty \leq R \text{ and } \lambda_{X^{(n)}} \notin \mathcal{U}) < 0.$$

(2) We have

$$P\left(\|X^{(n)}\|_\infty \geq R_0 + \delta\right) \leq d e^{-cn\delta^2/2},$$

where

$$R_0 = \left(\frac{2d}{c} + \|\nabla V(0)\|_2^2\right)^{1/2} + c^{-1/2}\Theta,$$

where Θ is the universal constant from Lemma 11.5.2.

(3) Let $f^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{C}$ be a sequence of functions such that $f^{(n)} \rightsquigarrow f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ and

$$|f^{(n)}(x)| \leq K_1 e^{K_2 \|x\|_\infty}$$

for some constants K_1 and K_2 independent of n. Then

$$\lim_{n\to\infty} E[f^{(n)}(X^{(n)})] = f(X).$$

Remark 15.1.6. We call the limiting law λ in this theorem the *free Gibbs law associated to* $(V^{(n)})_{n\in\mathbb{N}}$. We remark that λ only depends on the function $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\mathrm{sa}}^d$ such that $\nabla V^{(n)} \rightsquigarrow f$. This is automatic from the theorem by the following standard trick. Suppose that $V^{(n)}$ and $W^{(n)}$ are two sequences of potentials satisfying Assumption 15.1.1 such that $\nabla V^{(n)}$ and $\nabla W^{(n)}$ are both asymptotic to f. Then the sequence $\nabla V^{(1)}$, $\nabla W^{(2)}$, $\nabla V^{(3)}$, $\nabla W^{(4)}$, ...is also asymptotic to f, and hence by the theorem, the random matrix models associated to this sequence converge in non-commutative law to some ν almost surely. The limiting non-commutative laws for the even terms and the odd terms of the sequence must be the same, and hence the limiting non-commutative laws for $V^{(n)}$ and $W^{(n)}$ are the same.

Proof of Theorem 15.1.5. The idea is to combine Theorem 15.1.2 in the case $d_2 = 0$ with Herbst's concentration inequality (Corollary 11.5.1) and the operator norm bounds in Corollary 11.5.3. However, it requires a series of steps to get to the exact statements claimed in the theorem.

First, let us prove the operator norm bound (2). By Corollary 11.2.6,

$$\|E(X^{(n)})\|_2 \leq \left(\frac{2d}{c} + \|\nabla V^{(n)}(0)\|_2^2\right)^{1/2}.$$

However, because $V^{(n)}$ is invariant under unitary conjugation, $E(X)$ is a d-tuple of scalar multiples of the identity matrix. Therefore, $\|E(X_j^{(n)})\|_\infty = \|E(X_j^{(n)})\|_2$. Hence,

$$\|E(X_j^{(n)})\|_\infty + c^{-1/2}\Theta \leq R_0.$$

Now $x \mapsto x_j$ is 1-Lipschitz, and therefore by Lemma 11.5.2,

$$P\left(\left\|X_j^{(n)}\right\|_\infty \geq R_0 + c^{-1/2}\delta\right) \leq e^{-n\delta^2/2}.$$

To obtain (2), we substitute $c^{1/2}\delta$ for δ, and then use a union bound to get an estimate for $\|X\|_\infty$ rather than $\|X_j^{(n)}\|_\infty$.

Next, we turn our attention to the existence of the limit in (3). We claim that if $f^{(n)} : M_n(\mathbb{C})_{\text{sa}}^d \to \mathbb{C}$ is K-Lipschitz for all n and $f^{(n)} \rightsquigarrow f$, then $\lim_{n\to\infty} E[f^{(n)}(X^{(n)})]$ exists. This is actually a special case of Theorem 15.1.2. Indeed, $f^{(n)}$ can be viewed as a map into $M_n(\mathbb{C})$ by identifying $1 \in \mathbb{C}$ with $1 \in M_n(\mathbb{C})$, and as such it is asymptotically approximable by trace polynomials. We apply the theorem with $d_1 = d$ and $d_2 = 0$, and as we remarked earlier, the meaning is that $\lim_{n\to\infty} E[f^{(n)}(X^{(n)})]$ exists.

We claim that if $f^{(n)}$ is as in (3), then $\lim_{n\to\infty} E[f^{(n)}(X^{(n)})]$ exists. It suffices to show that for every $\epsilon > 0$, there is a constant L such that

$$\limsup_{n\to\infty} \left| E[f^{(n)}(X^{(n)})] - L \right| \leq \epsilon.$$

Let $f \in C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ such that $f^{(n)} \rightsquigarrow f$. Let $\tilde{f}$ be a globally $\|\cdot\|_2$-Lipschitz element of $C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ such that $\|f - \tilde{f}\|_{2,R_0} < \epsilon$. For instance, as in the previous chapters, we can choose $\tilde{f}$ to be a trace polynomial composed with a smooth cut-off function (see Lemma 13.3.5 and the proof of Proposition 13.6.6). We already know that $L := \lim_{n\to\infty} E[\tilde{f}(X^{(n)})]$ exists. Thus, it suffices to show that

$$\limsup_{n\to\infty} E|f^{(n)}(X^{(n)}) - \tilde{f}(X^{(n)})| \leq \epsilon.$$

Clearly,

$$\int_{\|x\|_\infty \leq R_0} |f^{(n)} - \tilde{f}| \, d\mu^{(n)} \leq \int_{\|x\|_\infty \leq R_0} |f^{(n)} - f| \, d\mu^{(n)} + \int_{\|x\|_\infty \leq R_0} \left| f - \tilde{f} \right| \, d\mu^{(n)},$$

and the first term goes to zero, while the second term is bounded by ϵ. Finally, observe that $|f^{(n)}(x) - \tilde{f}(x)| \leq K_1' e^{K_2 \|x\|_\infty}$ for some constant K_1'. Thus, using Corollary 11.5.3,

$$\int_{\|x\|_\infty > R_0} |f^{(n)} - \tilde{f}| \, d\mu^{(n)} \leq K_1' d \sqrt{\frac{2\pi}{nc}} e^{K_2^2/2nc} e^{R_0} \to 0.$$

Hence, $\limsup_{n\to\infty} |E[f^{(n)}(X^{(n)})] - L| \leq \epsilon$ as desired, which shows the existence of the limit.

Thus, we have shown that the limit in (3) always exists. In particular, for any non-commutative polynomial p, $\lim_{n\to\infty} E[\tau_n(p(X^{(n)}))]$ exists. We define $\lambda : \mathbb{C}\langle t_1, \dots, t_d \rangle \to \mathbb{R}$ by

$$\lambda(p) := \lim_{n\to\infty} E[\tau_n(p(X^{(n)}))].$$

Then $\lambda(p^* p) \geq 0$ since $\tau_n(p(X^{(n)})^* p(X^{(n)})) \geq 0$. Similarly, we have $\lambda(pq) = \lambda(qp)$ because $\tau_n(p(X^{(n)}) q(X^{(n)})) = \tau_n(q(X^{(n)}) p(X^{(n)})$. Finally, because

$$\lim_{n\to\infty} \int_{\|x\|_\infty > R_0} \tau_n(p) \, d\mu^{(n)} = 0,$$

we have $|\lambda(p)| \leq \|\tau(p)\|_{R_0}$. In particular, $|\lambda(t_{i_1} \dots t_{i_k})| \leq R_0^k$. Therefore, by Proposition 10.2.2, λ can be realized as the law of a self-adjoint d-tuple X from a tracial W*-algebra $(\mathcal{M}, \tau)$.

Now let us prove (1). Suppose $R > \|X\|_\infty$ and $\mathcal{U}$ is a neighborhood of λ in $\Sigma_{d,R}$. Because the topology on $\Sigma_{d,R}$ is generated by the functions $\lambda \mapsto \lambda(p)$ for a non-commutative polynomial p, there are non-commutative polynomials $p_1, \dots, p_k$ and some $\epsilon_1, \dots, \epsilon_k > 0$ such that

$$\lambda \in \bigcap_{j=1}^k \{\mu : |\mu(p_j) - \lambda(p_j)| < \epsilon_j\} \subseteq \mathcal{U}.$$

Hence,

$$P\left(\|X^{(n)}\|_\infty \leq R \text{ and } \lambda_{X^{(n)}} \notin \mathcal{U}\right) \leq \sum_{j=1}^k P\left(\|X^{(n)}\|_\infty \leq R \text{ and } |\lambda_{X^{(n)}}(p_j) - \lambda(p_j)| \geq \epsilon_j\right).$$

Our goal is to show that the left-hand side is bounded by $e^{-n^2\delta}$ for some $\delta > 0$, and to prove this, it suffices to establish such a bound for each term on the right hand side. That is, it suffices to show that for every $R > \|X\|_\infty$ and $\epsilon > 0$ and non-commutative polynomial p, we have

$$\limsup_{n \to \infty} \frac{1}{n^2} \log P\left(\|X^{(n)}\|_\infty \leq R \text{ and } |\tau_n(p(X^{(n)})) - \lambda(p)| \geq \epsilon\right) < 0.$$

Since the probability will only increase if R is made larger, we may assume without loss of generality that $R \geq R_0$. Let $f \in C_{\text{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)$ be a globally $\|\cdot\|_2$-Lipschitz function such that $f(x) = \tau_n(p(x))$ for $\|x\|_\infty \leq R$. Because $R \geq R_0$ and $|f - \tau(p)| \leq K_1 e^{K_2\|x\|_\infty}$ for some constants K_1 and K_2, our earlier arguments show that

$$\lim_{n \to \infty} E[f^{(n)}(X^{(n)})] = \lim_{n \to \infty} E[\tau_n(p(X^{(n)}))] = \lambda(p).$$

Thus, for sufficiently large n, we have

$$\left|E[f(X^{(n)})] - \lambda(p)\right| < \frac{\epsilon}{2}.$$

Meanwhile, by Herbst's concentration inequality (see Definition 11.4.3 and Corollary 11.5.1), we have

$$P\left(|f(X^{(n)}) - E[f(X^{(n)})]| \geq \frac{\epsilon}{2}\right) \leq 4e^{-cn^2\epsilon^2/16\|f\|_{\text{Lip}}^2}.$$

Therefore, for sufficiently large n,

$$\begin{aligned} \frac{1}{n^2} \log P\left(\|X^{(n)}\|_\infty \leq R \text{ and } |\tau_n(p(X^{(n)})) - \lambda(p)| \geq \epsilon\right) &\leq \frac{1}{n^2} \log 4e^{-cn^2\epsilon^2/16\|f\|_{\text{Lip}}^2} \\ &\to -\frac{cn^2\epsilon^2}{16\|f\|_{\text{Lip}}^2} < 0. \end{aligned}$$

This completes the proof of (1).

Next, let us prove almost sure convergence. Fix $R > R_0$. Let p be a non-commutative polynomial and $\epsilon > 0$. Note that

$$P(|\tau_n(p(X^{(n)}) - \lambda(p)| \geq \epsilon) \leq P(\|X^{(n)}\|_\infty > R) + P(\|X^{(n)}\|_\infty \leq R \text{ and } |\tau_n(p(X^{(n)}) - \lambda(p)| \geq \epsilon).$$

Using (2), the first term is bounded by $e^{-n\delta_1}$ for some $\delta_1 > 0$ since $R > R_0$. Using (1), the second term is bounded by $e^{-n^2\delta_2}$ for some $\delta_2 > 0$. This implies that

$$\sum_{n=1}^{\infty} P(|\tau_n(p(X^{(n)})) - \lambda(p)| \geq \epsilon) < +\infty,$$

and hence $\tau_n(p(X^{(n)}) \to \lambda(p)$ almost surely by the Borel-Cantelli lemma. This holds for every p and thus $\lambda_{X^{(n)}} \to \lambda$ almost surely.

Finally, let us finish the proof of (3). Let $f^{(n)} \rightsquigarrow f$ as in (3). Fix $R > R_0$. From the foregoing argument, we have $\lambda_{X^{(n)}} \to \lambda = \lambda_X$ almost surely, and also $\limsup_{n\to\infty} \|X^{(n)}\|_\infty \leq R_0$ almost surely. In this case, because f defines a continuous function on $\Sigma_{d,R}$, we have $f(X^{(n)}) \to f(X)$. Also, because $\sup_{\|x\|_\infty \leq R} |f^{(n)} - f| \to 0$, we have $f^{(n)}(X^{(n)}) \to f(X)$ almost surely. Thus, $1_{\|X^{(n)}\|_\infty \leq R} f^{(n)}(X^{(n)}) \to f(X)$ almost surely and this function is uniformly bounded (since $f^{(n)}$ is uniformly bounded for $\|x\|_\infty \leq R$ because $f^{(n)} \rightsquigarrow f$). It follows from dominated convergence that

$$\lim_{n\to\infty} E\left[1_{\|X^{(n)}\|_\infty \leq R} f^{(n)}(X^{(n)})\right] = f(X).$$

However, using Corollary 11.5.3 as above,

$$\lim_{n\to\infty} E\left[1_{\|X^{(n)}\|_\infty > R} f^{(n)}(X^{(n)})\right] = 0.$$

Therefore, (3) is proved. □

The next result is an extension of Theorem 15.1.2 by approximation arguments. The idea of the proposition is that the classical conditional expectation behaves in the large-n limit like a von Neumann algebraic conditional expectation.

Theorem 15.1.7. *Let $V^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to \mathbb{R}$ be a sequence of functions satisfying Assumption 15.1.1. Let $(X^{(n)}, Y^{(n)})$ be a random variable distributed according to the corresponding measure $\mu^{(n)}$, and let (X, Y) be a tuple from a tracial* W^**-algebra $(\mathcal{M}, \tau)$ which realizes the limiting free Gibbs law.*

Let $(f^{(n)})_{n\in\mathbb{N}}$ be a sequence of functions $M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to M_n(\mathbb{C})$ such that $f^{(n)} \rightsquigarrow f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*(d_1+d_2)}, \|\cdot\|_2)$ satisfies

$$\|f^{(n)}(x, y)\|_2 \leq K_1 e^{K_2 \|(x,y)\|_\infty}.$$

Define $g^{(n)}$ by

$$g^{(n)}(Y^{(n)}) = E[f^{(n)}(X^{(n)}, Y^{(n)})|Y^{(n)}].$$

Then $g^{(n)}$ is asymptotic to some $g \in C_{\text{tr,app}}(\mathbb{R}^{*d_2}, \|\cdot\|_2)$, and we have

$$g(Y) = E_{\mathrm{W}^*(Y)}[f(X, Y)],$$

where $E_{\mathrm{W}^*(Y)} : L^2(\mathrm{W}^*(X, Y)) \to L^2(\mathrm{W}^*(Y))$ denotes the orthogonal projection (which restricts to the conditional expectation $\mathrm{W}^*(X, Y) \to \mathrm{W}^*(Y)$). Hence, in particular,

$$\|E[f^{(n)}(X^{(n)}, Y^{(n)})|Y^{(n)}]\|_{L^2} \to \|E_{\mathrm{W}^*(Y)}[f(X, Y)]\|_2.$$

Proof. Fix $R > 0$ and $\epsilon > 0$, and we will approximate $g^{(n)}$ on the operator norm ball $\|y\|_\infty \leq R$ by some function from $C_{\text{tr,app}}(\mathbb{R}^{*d_2}, \|\cdot\|_2)$.

Since the argument of course relies on truncation to operator norm balls, we begin by deriving some operator norm bounds on $h_j^{(n)}(Y^{(n)}) := E[X_j^{(n)}|Y^{(n)}]$. By Theorem 15.1.2, $h_j^{(n)}(y)$ is a $(1 + C/c)$-Lipschitz function of y. Moreover, because $V^{(n)}$ is invariant under unitary conjugation, we also know that $h_j^{(n)}$ is equivariant under unitary conjugation. Therefore, by Lemma 11.5.4, we have

$$\|h_j^{(n)}(y) - \tau_n(h_j^{(n)}(y))\|_\infty \leq 2\sqrt{2}d^{1/2}\Theta(1 + C/c) \max_j \|y_j - \tau_n(y_j)\|_\infty,$$

where Θ is a universal constant. Next, because $\tau_n(h_j^{(n)}(y))$ is asymptotically approximable by trace polynomials, there is some constant $R_1 > 0$ such that

$$\sup_n \sup_{\|y\|_\infty \leq R} |\tau_n(h_j^{(n)}(y))| \leq R_1.$$

Therefore, for $\|y\|_\infty \leq R$,

$$\|h_j^{(n)}(y)\|_\infty \leq R_1 + 4\sqrt{2}d^{1/2}\Theta(1 + C/c)R =: R_2.$$

Let us denote $h^{(n)} = (h_1^{(n)}, \dots, h_{d_1}^{(n)})$, and then $\|h^{(n)}(y)\|_\infty \leq R_2$ for $\|y\|_\infty \leq R$.

Now let $R_3 = R_2 + c^{-1/2}\Theta$. By applying Corollary 11.5.3 to the conditional distribution of X given Y, we see that for any constant K_2,

$$\int_{\|x\|_\infty > R_3} e^{K_2\|x\|_\infty}\, d\mu^{(n)}(x|y) \leq d\sqrt{\frac{2\pi}{nc}} e^{K_2^2/2nc + R_3} \to 0,$$

since $\|E[X^{(n)}|Y^{(n)}]\|_\infty = \|h^{(n)}(Y^{(n)})\|_\infty \leq R_2$.

Recall that $f^{(n)} \rightsquigarrow f$. We can construct some $\tilde{f}$ such that $\|f - \tilde{f}\|_{2,\max(R,R_3)} < \epsilon$, and such that $\tilde{f}$ is globally Lipschitz with respect to $\|\cdot\|_2$. The way to obtain this $\tilde{f}$ is to approximate f by a trace polynomial, and then compose it with a compactly supported cut-off function as in Lemma 13.3.5.

Let $\tilde{g}^{(n)}(y) = \int \tilde{f}(x,y)\, d\mu^{(n)}(x|y)$. By Theorem 15.1.2, $\tilde{g}^{(n)}$ is asymptotic to some $\tilde{g} \in C_{\text{tr,app}}(\mathbb{R}^{*d_2}, \|\cdot\|_2)$. Now $\|f^{(n)}(x,y) - \tilde{f}(x,y)\|_2 \leq K_1' e^{K_2 \|x\|_\infty}$ for some constant K_1'. Therefore, for $\|y\|_\infty \leq R$,

$$\int_{\|x\|_\infty > R_3} \|f^{(n)}(x,y) - \tilde{f}(x,y)\|_2 \, d\mu^{(n)}(x|y) \leq K_1' d \sqrt{\frac{2\pi}{nc}} e^{K_2^2/2nc + R_3} \to 0.$$

Meanwhile, our choice of $\tilde{f}$ implies that for sufficiently large n,

$$\int_{\|x\|_\infty \leq R_3} \|f^{(n)}(x,y) - \tilde{f}(x,y)\|_2 \, d\mu^{(n)}(x|y) < \epsilon.$$

Thus, overall for $\|y\|_\infty \leq R$,

$$\|g^{(n)}(y) - \tilde{g}^{(n)}(y)\|_2 \leq \int \|f^{(n)}(x,y) - \tilde{f}(x,y)\|_2 \, d\mu^{(n)}(x|y) \leq K_1' d \sqrt{\frac{2\pi}{nc}} e^{K_2^2/2nc + R_3} + \epsilon.$$

Hence,

$$\limsup_{n \to \infty} \sup_{\|y\|_\infty \leq R} \left\| g^{(n)}(y) - \tilde{g}(y) \right\|_2 \leq \epsilon,$$

since $\tilde{g}^{(n)} \rightsquigarrow \tilde{g}$. Since R and ϵ were arbitrary, this shows that $g^{(n)}$ is asymptotically approximable by trace polynomials as desired.

Now let $g^{(n)} \rightsquigarrow g$. It remains to show that $g(Y) = E_{\mathrm{W}^*(Y)}[f(X,Y)]$. If $Z \in L^2(\mathrm{W}^*(Y))$, then we can pick a function $\phi \in C_{\text{tr,app}}(\mathbb{R}^{*d_2}, \|\cdot\|_2)$ such that $Z = \phi(Y)$, and we can arrange that ϕ is globally $\|\cdot\|_2$-bounded. Then the function $y \mapsto \tau_n(g^{(n)}(y)\phi(y))$ on $M_n(\mathbb{C})_{\text{sa}}^{d_2}$ is asymptotic to $\operatorname{tr}(g\phi)$ in $C_{\text{tr,app}}(\mathbb{R}^{*d_2}, \|\cdot\|_2)$ as $n \to \infty$, and similarly, the function $y \mapsto \tau_n(f^{(n)}(x,y)\phi(y))$ on $M_n(\mathbb{C})_{\text{sa}}^{d_1+d_2}$ is asymptotic to $\operatorname{tr}(f\phi(\pi_2))$ in $C_{\text{tr,app}}(\mathbb{R}^{*(d_1+d_2)}, \|\cdot\|_2)$, where $\pi_2(x,y) = y$. Therefore, by the previous lemma,

$$\begin{aligned} \tau(f(X,Y)\phi(Y)) &= \lim_{n \to \infty} E[\tau_n(f^{(n)}(X^{(n)}, Y^{(n)})\phi(Y^{(n)})] \\ &= \lim_{n \to \infty} E[\tau_n(g^{(n)}(Y^{(n)})\phi(Y^{(n)})] \\ &= \tau(g(Y)\phi(Y)). \end{aligned}$$

Thus, $\tau(f(X,Y)Z) = \tau(g(Y)Z)$ for all $Z \in L^2(\mathrm{W}^*(Y))$, hence $g(Y) = E_{\mathrm{W}^*(Y)}[f(X,Y)]$. □

A final consequence of Theorem 15.1.2 is that Assumption 15.1.1 is preserved under taking marginals of the measure $\mu^{(n)}$.

Proposition 15.1.8. *Let $V^{(n)} : M_n(\mathbb{C})_{\text{sa}}^{d_1+d_2} \to \mathbb{R}$ satisfy Assumption 15.1.1 for some $0 < c \leq C$. Define $\widehat{V}^{(n)} : M_n(\mathbb{C})_{\text{sa}}^{d_2} \to \mathbb{R}$ by*

$$\widehat{V}^{(n)}(y) = -\frac{1}{n^2} \log \int e^{-n^2 V^{(n)}(x,y)} \, dx.$$

Then $\widehat{V}^{(n)}$ satisfies Assumption 15.1.1 with the same constants c and C.

Proof. The inequality $cI \leq HV^{(n)} \leq CI$ can be written in block matrix form as

$$\begin{pmatrix} cI_{n^2d_1} & 0 \\ 0 & cI_{n^2d_2} \end{pmatrix} \leq HV^{(n)} \leq \begin{pmatrix} CI_{n^2d_1} & 0 \\ 0 & CI_{n^2d_2} \end{pmatrix}.$$

Thus, by Theorem 11.3.3 (taking $x_1 = y$ and $x_2 = x$), we have

$$(cI_{n^2d_2} - (0)(cI_{n^2d_1})^{-1}(0)) \leq H\widehat{V}^{(n)} \leq (CI_{n^2d_2} - (0)(CI_{n^2d_1})^{-1}(0)).$$

Now 11.3.4, $\nabla\widehat{V}^{(n)}(y) = \int \nabla_y V^{(n)}(x,y)\, d\mu^{(n)}(x|y)$. So by Theorem 15.1.2, $(\nabla\widehat{V}^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials. Finally, $\widehat{V}^{(n)}$ is invariant under unitary conjugation because $V^{(n)}$ is invariant under unitary conjugation. □

15.2 Strategy

This section explains the heuristics behind the proof of Theorem 15.1.2. We cannot hope to analyze the integral formula for $g^{(n)}$ directly because this integral formula is heavily dimension-dependent — for instance, all the integrals are in different dimensions and they all have n^2 in the exponent. Instead, we will express the conditional expectation through the action of a certain semigroup. Let us first consider the case where $d_2 = 0$, that is, where we evaluate an expectation rather than a conditional expectation.

Let $\mu^{(n)}$ be a measure given by the potential $n^2V^{(n)}$, where $cI \leq HV^{(n)} \leq CI$. Given a function $f^{(n)} : M_n(\mathbb{C})^d_{\text{sa}} \to M_n(\mathbb{C})$, let f evolve according to the normalized heat equation with respect to $\mu^{(n)}$. If $f^{(n)}$ were a scalar function, then the heat equation would be given by

$$\partial_t f_t^{(n)} = \frac{1}{2n^2}\Delta f_t^{(n)} - \frac{1}{2}\langle \nabla f_t^{(n)}, \nabla V^{(n)}\rangle. \tag{15.1}$$

But of course, $f^{(n)}$ is vector-valued, so we will express the equation as

$$\partial_t f_t^{(n)} = \frac{1}{2n^2}\Delta f_t^{(n)} - \frac{1}{2} J f_t^{(n)} \cdot \nabla V^{(n)}, \tag{15.2}$$

where $Jf_t^{(n)}(x)$ is the Jacobian linear transformation of $f_t^{(n)}$ at the point x. Then we denote $T_t^{(n)} f^{(n)} := f_t^{(n)}$, that is, $T_t^{(n)}$ is the heat semigroup associated to $V^{(n)}$.

One can check using integration by parts (Lemma 11.2.5) that $\int f_t \, d\mu^{(n)}$ is independent of t. Moreover, due to the uniform convexity of $V^{(n)}$, it turns out that $\|f_t^{(n)}\|_{\text{Lip}} \to 0$ as $t \to \infty$, and thus f_t converges to its expectation $\int f \, d\mu^{(n)}$.

Remark 15.2.1. There is a dual viewpoint on this semigroup in terms of SDE rather than PDE theory. Let $S_t^{(n)}$ be a Brownian motion on $M_n(\mathbb{C})^d_{\text{sa}}$ normalized so that $E\|S_t^{(n)}\|_2^2 = d \cdot t$. Let $X_t^{(n)}$ be a solution of the SDE

$$dX_t^{(n)} = dS_t^{(n)} - \frac{1}{2} DV^{(n)}(X_t^{(n)})\, dt \tag{15.3}$$

with any given initial distribution. Then the law of $X_t^{(n)}$ will evolve toward the law $\mu^{(n)}$ as $t \to +\infty$. More precisely, the probability density of $X_t^{(n)}$ evolves according to the heat equation with respect to $\mu^{(n)}$. Moreover, the solution to (15.1) is given by $f_t^{(n)}(x) = E[f^{(n)}(X_t)]$ where $X_0 = x$.

Similar to §12.4, we can study the conditional setting by applying the non-conditional results fiberwise. Indeed, if $V^{(n)}$ is now a function of (x, y) and $f^{(n)} : M_n(\mathbb{C})_{\text{sa}}^{d_1+d_2} \to M_n(\mathbb{C})$, then we can study the conditional diffusion equation

$$\partial_t f_t^{(n)} = \frac{1}{2n^2} \Delta_x f_t^{(n)} - \frac{1}{2} J_x f_t^{(n)} \cdot \nabla_x V^{(n)}, \tag{15.4}$$

where the differentiation occurs with respect to x, and y only plays the role of a parameter. The conditional expectation of $f_t^{(n)}(X^{(n)}, Y^{(n)})$ can be evaluated as the limit as $t \to \infty$ of $f_t^{(n)}(Y^{(n)})$.

The point of using the heat equation associated to $V^{(n)}$ is that it is dimension-independent. The formula $e^{-n^2 V^{(n)}}$ has an n in the exponent, which is difficult to deal with directly in the large-n limit. However, the equation (15.4) is dimension-independent; the only occurrence of n is in the expression $(1/n^2)\Delta$, but we saw in §14.1 that this is the dimension-independent normalization of the Laplacian on trace polynomials.

It is precisely for this reason that much of the previous work on free Gibbs laws relied on this heat semigroup, though this was more often expressed from the SDE viewpoint rather than the PDE viewpoint. We should also point out that this semigroup is the basis for the standard proof of the log-Sobolev and Herbst concentration inequalities (see [AGZ09, §4.4.2]).

Our main technical contribution is to show that, roughly speaking, the conditional heat semigroup associated to $V^{(n)}$ preserves asymptotic approximation by polynomials. More precisely, let $T_t^{(n)}$ denotes the conditional heat semigroup with respect to $V^{(n)}(x, y)$, let $(V^{(n)})_{n \in \mathbb{N}}$ satisfy Assumption 15.1.1, and let $f^{(n)} : M_n(\mathbb{C})_{\text{sa}}^{d_1+d_2} \to M_n(\mathbb{C})$ be uniformly Lipschitz and asymptotically approximable by trace polynomials. Then $T_t^{(n)} f^{(n)}$ is asymptotically approximable by trace polynomials.

In order to do this, we express $T_t^{(n)}$ in terms of two simpler semigroups. Let $\Phi_t^{(n)}$ be the flat heat semigroup given by

$$\partial_t(\Phi_t^{(n)} f) = \frac{1}{2n^2} \Delta_x (\Phi_t^{(n)} f),$$

and let $\Psi_t^{(n)}$ be the semigroup given by

$$\partial_t(\Psi_t^{(n)} f) = -\frac{1}{2} \langle \nabla_x (\Psi_t^{(n)} f), \nabla_x V^{(n)} \rangle.$$

We already saw in Lemma 14.2.6 that $\Phi_t^{(n)}$ preserves asymptotic approximation by trace polynomials. Meanwhile, $\Psi_t^{(n)}$ can be expressed in terms of solving ODE, so it preserves asymptotic approximation as well. But the semigroup $T_t^{(n)}$ can be obtained as

$$T_t^{(n)} = \lim_{k \to \infty} (\Phi_{t/k}^{(n)} \Psi_{t/k}^{(n)})^k$$

with dimension-independent rates of convergence as $k \to \infty$. Thus, $T_t^{(n)}$ also preserves asymptotic approximation by trace polynomials.

This method of "blending" the two semigroups $\Phi_t^{(n)}$ and $\Psi_t^{(n)}$ to obtain $T_t^{(n)}$ is motivated by Trotter's product formula which asserts that $e^{t(A+B)} = \lim_{k \to \infty} (e^{tA/k} e^{tB/k})^k$ for nice enough self-adjoint operators A and B (see [Tro59], [Kat78], [Sim79, pp. 4 - 6]). Actually, rather than cutting $[0, t]$ into k subintervals of equal length, we will consider an arbitrary partition into k subintervals. Thus, we will proceed very much in the spirit of Riemann integration (like for instance, the results in Part I). Our approach, if we were to translate it into the SDE viewpoint, is comparable to using a discrete-time approximation to solve the conditional version of the SDE (15.3).

15.3 Diffusion semigroup via Trotter's formula

In this section, we begin the technical details of the proof of Theorem 15.1.2. Our first goal is to carry out the "Trotter's formula" construction of the semigroup $T_t^{(n)}$ by blending $\Phi_t^{(n)}$ and $\Psi_t^{(n)}$. In particular, we will first prove that the approximations converge to some semigroup $T_t^{(n)}$, that this semigroup actually gives a solution to (15.1), and that the limit as $t \to \infty$ actually exists and evaluates expectations with respect to the measure $\mu^{(n)}$. Furthermore, all our estimates should be dimension-independent.

In order to simplify notation, we will freeze the variable y and only study functions of $x \in M_n(\mathbb{C})_{\text{sa}}^d$. Moreover, for the rest of the section, we will fix n, fix $V : M_n(\mathbb{C})_{\text{sa}}^d \to \mathbb{R}$ with $cI \leq HV \leq CI$ for some $0 < c \leq C$, and let μ be the associated measure. Hence, there will not be any more superscripts (n) until we return to study asymptotic approximation in §15.5.

We will study the actions of the semigroups Φ_t and Ψ_t on Lipschitz functions $f : M_n(\mathbb{C})_{\text{sa}}^d \to M_n(\mathbb{C})$. Here $\|f\|_{\text{Lip}}$ denotes the Lipschitz seminorm with respect to $\|\cdot\|_2$ on $M_n(\mathbb{C})_{\text{sa}}^d$ and $\|f\|_{L^\infty}$ denotes

$$\|f\|_{L^\infty} = \operatorname{ess\,sup}\{\|f(x)\|_2 : x \in M_n(\mathbb{C})_{\text{sa}}^d\}.$$

On the other hand, $\|f\|_2$ denotes $\tau_n(f^* f)^{1/2}$, which is still a function of x; if we need to refer to the L^2 norm with respect to a measure, we will explicitly write L^2 in the subscript. Finally, since a Lipschitz function f is differentiable almost everywhere by Rademacher's theorem, we will denote by Jf the Jacobian linear transformation, by $\|Jf\|$ its operator norm, and by $\|Jf\|_{L^\infty}$ the essential supremum of $\|Jf\|$, which is the same as the Lipschitz seminorm of f.

We begin with the definition and basic properties of semigroups Φ_t and Ψ_t.

Lemma 15.3.1. *Let* $\Phi_t f(x) = \int f(x+z)\, d\sigma_t^{(n)}(z)$ *where* $\sigma_t^{(n)}(z)$ *is the Gaussian measure on* $M_n(\mathbb{C})_{\mathrm{sa}}^d$ *as in §14.2. Then*

(1) $\|\Phi_t f\|_{\mathrm{Lip}} \leq \|f\|_{\mathrm{Lip}}$.

(2) $\|\Phi_t f\|_{L^\infty} \leq \|f\|_{L^\infty}$.

(3) $\|\Phi_t f - f\|_{L^\infty} \leq d^{1/2} t^{1/2} \|f\|_{\mathrm{Lip}}$.

Proof. (1) and (2) follow from the fact that $\Phi_t u$ is u convolved with a probability measure. To prove (3), suppose $\|f\|_{\mathrm{Lip}} < +\infty$. Then

$$\begin{aligned}\|\Phi_t f(x) - f(x)\|_2 &= \left\| \int (f(x+z) - f(x))\, d\sigma_t^{(n)}(z) \right\|_2 \\ &\leq \int \|f(x+z) - f(x)\|_2 \, d\sigma_t^{(n)}(z) \\ &\leq \|f\|_{\mathrm{Lip}} \int \|z\|_2 \, d\sigma_t^{(n)}(z).\end{aligned}$$

Meanwhile,

$$\int \|y\|_2 \, d\sigma_t^{(n)}(z) \leq \left(\int 1 \, d\sigma_t^{(n)}(z) \right) \left(\int \|z\|_2^2 \, d\sigma_t^{(n)}(z) \right)^{1/2} = (d \cdot t)^{1/2},$$

since z an d-tuple $(z_1, \dots, z_d)$ and $\int \tau_n(z_j^2)\, d\sigma_t^{(n)}(z) = t$ for each j. □

Lemma 15.3.2. *Let* $W(x,t)$ *be the solution to the ODE*

$$\partial_t W(x,t) = -\frac{1}{2} \nabla V(W(x,t)), \tag{15.5}$$

which exists for all t *using the Picard-Lindelöf theorem, and define*

$$\Psi_t f(x) := f(W(x,t)).$$

Then

(1) $\|W(x,t) - W(y,t)\|_2 \leq e^{-ct/2} \|x - y\|_2$.

(2) $\|W(x,t) - x\|_2 \leq (t/2) \|\nabla V(x)\|_2$.

(3) $\|(W(x,t) - x) - (W(y,t) - y)\|_2 \leq \frac{C}{c}(1 - e^{-ct/2}) \|x - y\|_2$.

(4) $\|\Psi_t f\|_{\mathrm{Lip}} \leq e^{-ct/2} \|f\|_{\mathrm{Lip}}$.

(5) $\|\Psi_t f\|_{L^\infty} \leq \|f\|_{L^\infty}$ *if* f *is continuous.*

Proof. The convexity and semi-concavity assumptions on V imply that ∇V is C-Lipschitz (Lemma 11.1.4) and therefore global existence of the solution follows from the Picard–Lindelöf Theorem.

(1) Let $\tilde{V}(x) = V(x) - (c/2)\|x\|_2^2$. By Lemma 11.1.4,

$$\langle \nabla V(x) - \nabla V(y), x - y\rangle_2 \geq c\|x - y\|_2^2.$$

Now observe that

$$\begin{aligned}\frac{d}{dt}\|W(x,t) - W(y,t)\|_2^2 &= -\langle \nabla V(W(x,t)) - \nabla V(W(y,t)), W(x,t) - W(y,t)\rangle_2 \\ &\leq -c\|W(x,t) - W(y,t)\|_2^2,\end{aligned}$$

and hence by Grönwall's inequality, $\|W(x,t) - W(y,t)\|_2^2 \leq e^{-ct}\|W(x,0) - W(y,0)\|_2^2 = e^{-ct}\|x - y\|_2^2$.

(2) Note that

$$\begin{aligned}\frac{d}{dt}\|W(x,t) - x\|_2^2 &= -\langle \nabla V(W(x,t)), W(x,t) - x\rangle_2 \\ &= -\langle \nabla V(W(x,t)) - \nabla V(x), W(x,t) - x\rangle_2 - \langle \nabla V(x), W(x,t) - x\rangle_2 \\ &\leq \|\nabla V(x)\|_2 \|W(x,t) - x\|_2.\end{aligned}$$

Meanwhile, $\|W(x,t) - x\|_2$ is Lipschitz in t and hence differentiable almost everywhere and we have

$$\frac{d}{dt}\|W(x,t) - x\|_2^2 = 2\|W(x,t) - x\|_2 \frac{d}{dt}\|W(x,t) - x\|_2.$$

Thus,

$$\frac{d}{dt}\|W(x,t) - x\|_2 \leq \frac{1}{2}\|\nabla V(x)\|_2,$$

which proves (2).

(3) We observe that

$$\begin{aligned}\|(W(x,t) - x) - (W(y,t) - y)\|_2 &\leq \frac{1}{2}\int_0^t \|DV(W(x,s)) - DV(W(y,s))\|_2 \, ds \\ &\leq \frac{C}{2}\int_0^t \|W(x,s) - W(y,s)\|_2 \, ds \\ &\leq \frac{C}{2}\int_0^t e^{-cs/2}\|x - y\|_2 \, ds \\ &= \frac{C}{c}(1 - e^{-ct/2})\|x - y\|_2.\end{aligned}$$

(4) follows from (1).

(5) is immediate because $\Psi_t f$ is f precomposed with another function. □

Let $\mathcal{P} = \{t_0, \dots, t_k\}$ be some partition of $[s,t]$ with $s = t_0$ and $t = t_k$. Then we define

$$T_{\mathcal{P}} f = \Phi_{t_k - t_{k-1}} \Psi_{t_k - t_{k-1}} \dots \Phi_{t_1 - t_0} \Psi_{t_1 - t_0} f.$$

We define $\mathrm{mesh}(\mathcal{P}) := \max\{|t_j - t_{j-1}| : j = 1, \dots, k\}$. Our goal is to estimate $\|T_{\mathcal{P}} f - T_{\mathcal{Q}} f\|_{L^\infty}$ when $\mathcal{Q}$ is a refinement of $\mathcal{P}$. First, we observe the following lemma which follows from inductive application of our earlier estimates for Φ_t and Ψ_t.

Lemma 15.3.3. *Let $\mathcal{P}$ be a partition of $[s,t]$. Then*

(1) $\|T_{\mathcal{P}} f\|_{\mathrm{Lip}} \leq e^{-c(t-s)/2} \|f\|_{\mathrm{Lip}}$.

(2) $\|T_{\mathcal{P}} f\|_{L^\infty} \leq \|f\|_{L^\infty}$.

The good behavior of $T_{\mathcal{P}}$ under refinement of $\mathcal{P}$ derives from the basic fact that Φ_s and Ψ_t "almost commute" when s and t are small.

Lemma 15.3.4.

$$\|\Psi_t \Phi_s f - \Phi_s \Psi_t f\|_{L^\infty} \leq \frac{C}{c} d^{1/2} s^{1/2} (1 - e^{-ct/2}) \|f\|_{\mathrm{Lip}}.$$

Proof. Thus, we want to estimate $\Psi_t \Phi_t - \Phi_t \Psi_t$ and then control the propagation of the errors through the applications of the other operators. Note that for a Lipschitz function f, we have using Lemma 15.3.2 (3) that

$$\begin{aligned}
\|\Psi_t \Phi_s f(x) - \Phi_s \Psi_t f(x)\|_2 &\leq \int \|f(W(x,t) + z) - f(W(x+z,t))\|_2 \, d\sigma_s^{(n)}(z) \\
&\leq \|f\|_{\mathrm{Lip}} \int \|(W(x,t) - x) - (W(x+z,t) - (x+z))\|_2 \, d\sigma_s^{(n)}(y) \\
&\leq \|f\|_{\mathrm{Lip}} \frac{C}{c} (1 - e^{-ct/2}) \int \|z\|_2 \, d\sigma_s^{(n)}(z) \\
&\leq \|f\|_{\mathrm{Lip}} \frac{C}{c} (1 - e^{-ct/2}) (ds)^{1/2},
\end{aligned}$$

where the last inequality follows by the same reasoning as Lemma 15.3.1 (3). □

Lemma 15.3.5. *Let $\mathcal{Q}$ be a refinement of the partition $\mathcal{P}$ of the interval $[s,t]$. Then*

$$\|T_{\mathcal{P}} u - T_{\mathcal{Q}} u\|_{L^\infty} \leq \frac{C d^{1/2}}{c} (1 - e^{-c(t-s)/2}) \, \mathrm{mesh}(\mathcal{P})^{1/2} \|f\|_{\mathrm{Lip}}.$$

Proof. First, we will prove the following claim: Let $\mathcal{Q} = \{t_0, \dots, t_k\}$. Then

$$\|T_{\mathcal{P}} u - T_{\mathcal{Q}} u\|_{L^\infty} \leq \frac{C d^{1/2}}{c} \sum_{j : t_j \in \mathcal{Q} \setminus \mathcal{P}} (e^{-c(t_j - s)/2} - e^{-c(t_{j+1} - s)}) \, \mathrm{mesh}(\mathcal{P})^{1/2} \|f\|_{\mathrm{Lip}}.$$

We proceed by induction on the cardinality of $\mathcal{Q} \setminus \mathcal{P}$. In the base case, $|\mathcal{Q} \setminus \mathcal{P}| = 0$, so $\mathcal{Q} = \mathcal{P}$, and the claim is trivial.

For the inductive step, fix $\mathcal{Q}$. Let t_{j^*} be the least element of $\mathcal{Q} \setminus \mathcal{P}$ and let

$$\begin{aligned} \mathcal{Q}_1 &= \{t_0, \dots, t_{j^*-1}\}, \\ \mathcal{Q}_2 &= \{t_{j^*+1}, \dots, t_k\}, \\ \mathcal{Q}' &= \mathcal{Q} \setminus \{t_{j^*}\} = \mathcal{Q}_1 \cup \mathcal{Q}_2. \end{aligned}$$

Thus,

$$\begin{aligned} T_{\mathcal{Q}} f &= T_{\mathcal{Q}_2} \Phi_{t_{j^*+1}-t_{j^*}} \Psi_{t_{j^*+1}-t_{j^*}} \Phi_{t_{j^*}-t_{j^*-1}} \Psi_{t_{j^*}-t_{j^*-1}} T_{\mathcal{Q}_1} f \\ T_{\mathcal{Q}'} f &= T_{\mathcal{Q}_2} \Phi_{t_{j^*+1}-t_{j^*-1}} \Psi_{t_{j^*+1}-t_{j^*-1}} T_{\mathcal{Q}_1} f. \end{aligned}$$

Therefore,

$$\begin{aligned} T_{\mathcal{Q}} f - T_{\mathcal{Q}'} f &= T_{\mathcal{Q}_2} (\Phi_{t_{j^*+1}-t_{j^*}} \Psi_{t_{j^*+1}-t_{j^*}} \Phi_{t_{j^*}-t_{j^*-1}} \Psi_{t_{j^*}-t_{j^*-1}} - \Phi_{t_{j^*+1}-t_{j^*-1}} \Psi_{t_{j^*+1}-t_{j^*-1}}) T_{\mathcal{Q}_1} f \\ &= T_{\mathcal{Q}_2} \Phi_{t_{j^*+1}-t_{j^*}} (\Psi_{t_{j^*+1}-t_{j^*}} \Phi_{t_{j^*}-t_{j^*-1}} - \Phi_{t_{j^*}-t_{j^*-1}} \Psi_{t_{j^*+1}-t_{j^*}}) \Psi_{t_{j^*}-t_{j^*-1}} T_{\mathcal{Q}_1} f, \end{aligned}$$

where we applied linearity and the semigroup properties of Φ and Ψ. Thus, by Lemma 15.3.3 (2) and Lemma 15.3.1 (2),

$$\|T_{\mathcal{Q}} f - T_{\mathcal{Q}'} f\|_{L^\infty} \leq \|(\Psi_{t_{j^*+1}-t_{j^*}} \Phi_{t_{j^*}-t_{j^*-1}} - \Phi_{t_{j^*}-t_{j^*-1}} \Psi_{t_{j^*+1}-t_{j^*}}) \Psi_{t_{j^*}-t_{j^*-1}} T_{\mathcal{Q}_1} f\|_{L^\infty},$$

hence by Lemma 15.3.4

$$\|T_{\mathcal{Q}} f - T_{\mathcal{Q}'} f\|_{L^\infty} \leq \frac{Cd^{1/2}}{c} (1 - e^{-c(t_{j^*+1}-t_{j^*})/2}) (t_{j^*} - t_{j^*-1})^{1/2} \|\Psi_{t_{j^*}-t_{j^*-1}} T_{\mathcal{Q}_1} f\|_{\mathrm{Lip}},$$

hence by Lemma 15.3.2 (4) and Lemma 15.3.3 (2),

$$\begin{aligned} \|T_{\mathcal{Q}} f - T_{\mathcal{Q}'} f\|_{L^\infty} &\leq \frac{Cd^{1/2}}{c} (1 - e^{-c(t_{j^*+1}-t_{j^*})/2}) (t_{j^*} - t_{j^*-1})^{1/2} e^{-c(t_{j^*}-t_{j^*-1})/2} e^{-c(t_{j^*-1}-s)/2} \|f\|_{\mathrm{Lip}} \\ &= \frac{Cd^{1/2}}{c} (e^{-c(t_{j^*}-s)/2} - e^{-c(t_{j^*+1}-s)/2}) (t_{j^*} - t_{j^*-1})^{1/2} \|f\|_{\mathrm{Lip}} \\ &\leq \frac{Cd^{1/2}}{c} (e^{-c(t_{j^*}-s)/2} - e^{-c(t_{j^*+1}-s)/2}) \operatorname{mesh}(\mathcal{P})^{1/2} \|f\|_{\mathrm{Lip}}. \end{aligned}$$

Meanwhile, by the inductive hypothesis,

$$\|T_{\mathcal{Q}'} f - T_{\mathcal{P}} f\| \leq \sum_{j : t_j \in \mathcal{Q}' \setminus \mathcal{P}} \frac{1}{c} (e^{-c(t_j - s)/2} - e^{-c(t_{j+1}-s)}) \operatorname{mesh}(\mathcal{P})^{1/2} \|f\|_{\mathrm{Lip}};$$

this works because for each t_j in $\mathcal{Q}' \setminus \mathcal{P}$, the next element in $\mathcal{Q}'$ is the same as the next element in $\mathcal{Q}$, since t_{j^*} was chosen to the be least element in $\mathcal{P} \setminus \mathcal{Q}$. Therefore, using the triangle inequality, we can bound $\|T_{\mathcal{Q}}f - T_{\mathcal{P}}f\|$ by

$$\frac{Cd^{1/2}}{c}\left((e^{-c(t_{j^*}-s)/2} - e^{-c(t_{j^*+1}-s)/2}) + \sum_{j:t_j\in\mathcal{Q}'\setminus\mathcal{P}}(e^{-c(t_j-s)/2} - e^{-c(t_{j+1}-s)})\right)\operatorname{mesh}(\mathcal{P})^{1/2}\|f\|_{\mathrm{Lip}}$$

$$=\frac{Cd^{1/2}}{c}\sum_{j:t_j\in\mathcal{Q}\setminus\mathcal{P}}(e^{-c(t_j-s)/2} - e^{-c(t_{j+1}-s)})\operatorname{mesh}(\mathcal{P})^{1/2}\|f\|_{\mathrm{Lip}}.$$

This completes the inductive proof of our initial claim. Then the lemma follows because

$$\begin{aligned}
&\frac{Cd^{1/2}}{c}\sum_{j:t_j\in\mathcal{Q}\setminus\mathcal{P}}(e^{-c(t_j-s)/2} - e^{-c(t_{j+1}-s)/2})\operatorname{mesh}(\mathcal{P})^{1/2}\|f\|_{\mathrm{Lip}} \\
\leq\ &\frac{Cd^{1/2}}{c}\sum_{j=1}^{k}(e^{-c(t_j-s)/2} - e^{-c(t_{j+1}-s)/2})\operatorname{mesh}(\mathcal{P})^{1/2}\|f\|_{\mathrm{Lip}} \\
=\ &\frac{Cd^{1/2}}{c}(1 - e^{-c(t-s)/2})\operatorname{mesh}(\mathcal{P})^{1/2}\|f\|_{\mathrm{Lip}}
\end{aligned}$$

by summation of telescoping series. □

Finally, we are ready to define the semigroup T_t.

Lemma 15.3.6. *There exists a semigroup $(T_t)_{t\geq 0}$ acting on Lipschitz functions $M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})$ such that*

(1) $\|T_t f\|_{\mathrm{Lip}} \leq e^{-ct/2}\|f\|_{\mathrm{Lip}}$.

(2) $\|T_t f\|_{L^\infty} \leq \|f\|_{L^\infty}$.

(3) If $\mathcal{P}$ is a partition of $[s,t]$, then

$$\|T_{\mathcal{P}}f - T_{t-s}f\|_{L^\infty} \leq \frac{Cd^{1/2}}{c}(1 - e^{-c(t-s)/2})\operatorname{mesh}(\mathcal{P})^{1/2}\|f\|_{\mathrm{Lip}}$$

Proof. We define

$$T_t f = \lim_{\operatorname{mesh}(\mathcal{P})\to 0} T_{\mathcal{P}}f,$$

where the limit is taken over all partitions $\mathcal{P}$ of $[0,t]$. This limit exists by Lemma 15.3.5. Moreover, since any partition can be translated along the real line by some distance s, it is also clear that for $s \leq t$, we have $T_{t-s}f = \lim_{\operatorname{mesh}(\mathcal{P})\to 0} T_{\mathcal{P}}f$ for partitions $\mathcal{P}$ of $[s,t]$, and the estimate (3) holds by letting $\operatorname{mesh}(\mathcal{Q}) \to 0$ in Lemma 15.3.5. The claims (1) and (2) follow immediately from Lemma 15.3.3.

It remains to show that T_t is a semigroup. Let $s, t \geq 0$. Let $\mathcal{P}$ and $\mathcal{Q}$ be partitions of $[0, s]$ and $[s, s+t]$ respectively. Then $\mathcal{P} \cup \mathcal{Q}$ is a partition of $[0, s+t]$ and $T_{\mathcal{P}\cup\mathcal{Q}} u = T_{\mathcal{Q}} T_{\mathcal{P}} u$. Moreover,

$$\begin{aligned}
&\|T_{\mathcal{Q}} T_{\mathcal{P}} f - T_t T_s f\|_{L^\infty} \\
&\leq \|(T_{\mathcal{Q}} - T_t) T_{\mathcal{P}} f\|_{L^\infty} + \|T_t (T_{\mathcal{P}} - T_s) f\|_{L^\infty} \\
&\leq \frac{C d^{1/2}}{c}(1 - e^{-ct/2}) \operatorname{mesh}(\mathcal{Q})^{1/2} e^{-cs/2} \|f\|_{\mathrm{Lip}} + \frac{C d^{1/2}}{c}(1 - e^{-cs/2}) \operatorname{mesh}(\mathcal{P})^{1/2} \|f\|_{\mathrm{Lip}} \\
&\leq \frac{C d^{1/2}}{c}(1 - e^{-c(s+t)/2}) \operatorname{mesh}(\mathcal{P} \cup \mathcal{Q})^{1/2} \|f\|_{\mathrm{Lip}}.
\end{aligned}$$

On the other hand,

$$\|T_{\mathcal{P}\cup\mathcal{Q}} f - T_{s+t} f\|_{L^\infty} \leq \frac{C d^{1/2}}{c}(1 - e^{-c(s+t)/2}) \operatorname{mesh}(\mathcal{P} \cup \mathcal{Q})^{1/2} \|f\|_{\mathrm{Lip}},$$

which implies

$$\|T_t T_s f - T_{s+t} f\| \leq \frac{2 C d^{1/2}}{c}(1 - e^{-c(s+t)/2}) \operatorname{mesh}(\mathcal{P} \cup \mathcal{Q})^{1/2} \|f\|_{\mathrm{Lip}},$$

and since $\mathcal{P}$ and $\mathcal{Q}$ were arbitrary, $T_{s+t} f = T_t T_s f$ as desired. □

Next, we must show that $T_t f$ solves the differential equation (15.2). As we have only assumed that V is C^1 with Lipschitz first derivative, we cannot conclude that $T_t f$ is C^2, so we must be content with showing that $T_t f$ solves the equation in a weaker sense. While the optimal results about the smoothness of $T_t f$ are irrelevant for our present goal, we know at least that $T_t f$ is Lipschitz, and therefore by Rademacher's theorem, $T_t f(x)$ is differentiable (with respect to x) almost everywhere, and of course $\|J(T_t f)\|_{L^\infty} = \|T_t u\|_{\mathrm{Lip}}$.

Lemma 15.3.7. *Let f be Lipschitz, and let $f_t = T_t f$. Then for $u \in C_c^\infty(M_N(\mathbb{C})_{\mathrm{sa}}^d)$, the map $t \mapsto \int f_t u$ is Lipschitz on $[0, +\infty)$ and for almost every t,*

$$\frac{d}{dt} \int f_t u = -\frac{1}{2} \int J f_t \cdot \left(u \nabla V + \frac{1}{n^2} \nabla u \right)$$

Proof. Observe that

$$f_{t+\epsilon} - f_t = (T_\epsilon - 1) f_t = (T_\epsilon - \Phi_\epsilon \Psi_\epsilon) f_t + (\Phi_\epsilon \Psi_\epsilon - \Psi_\epsilon \Phi_\epsilon) f_t + (\Psi_\epsilon \Phi_\epsilon - 1) f_t. \tag{15.6}$$

By Lemma 15.3.6 (3) applied to the partition $\{0, \epsilon\}$ of the interval $[0, \epsilon]$,

$$\|T_\epsilon f_t - \Phi_\epsilon \Psi_\epsilon f_t\|_{L^\infty} \leq \frac{C d^{1/2}}{c}(1 - e^{-c\epsilon/2}) \epsilon^{1/2} \|f_t\|_{\mathrm{Lip}} \leq \frac{C d^{1/2}}{2} \epsilon^{3/2} \|f_t\|_{\mathrm{Lip}},$$

since $1 - e^{-c\epsilon/2} \leq \frac{c\epsilon}{2}$. The second term of (15.6) can be estimated by Lemma 15.3.4:

$$\|(\Phi_\epsilon \Psi_\epsilon - \Psi_\epsilon \Phi_\epsilon) f_t\|_{L^\infty} \leq \frac{Cd^{1/2}}{c}(1 - e^{-c\epsilon/2})\epsilon^{1/2}\|f_t\|_{\text{Lip}} \leq \frac{Cd^{1/2}}{2}\epsilon^{3/2}\|f_t\|_{\text{Lip}}.$$

Since u is in L^1, the first two terms on the right-hand side of (15.6) will still be $O(\epsilon^{3/2})$ when we pair them with u and integrate:

$$\int (f_{t+\epsilon} - f_t) u = \int (\Psi_\epsilon \Phi_\epsilon - 1) f_t \cdot u + O(\epsilon^{3/2}).$$

Now we write

$$(\Psi_\epsilon \Phi_\epsilon - 1) f_t = (\Psi_\epsilon - 1)\Phi_\epsilon f_t + (\Phi_\epsilon - 1) f_t. \tag{15.7}$$

To understand $(\Psi_\epsilon - 1)\Phi_\epsilon f_t$, recall that Φ_ϵ is convolution with the Gaussian measure with variance $\epsilon I/n^2$. So $\Phi_\epsilon f_t$ is smooth. Moreover, $J(\Phi_\epsilon f_t) = \Phi_\epsilon(J f_t)$. The first derivatives of the Gaussian density are $O(\epsilon^{-1/2})$ in L^1, and therefore the first derivatives of $J(\Phi_\epsilon f_t) = \Phi_\epsilon(J f_t)$ are $O(\epsilon^{-1/2})$ in L^∞.

Now observe that

$$\begin{aligned}(\Psi_\epsilon - 1)\Phi_\epsilon f_t(x) &= \Phi_\epsilon f_t(W(x,\epsilon)) - \Phi_\epsilon f_t(x) \\ &= J(\Phi_\epsilon f_t)(x) \cdot (W(x,\epsilon) - x) + O(\epsilon^{-1/2}\|W(x,\epsilon) - x\|_2^2),\end{aligned}$$

where in the last line we have used Taylor expansion of $\Phi_\epsilon f$ along with the fact that the second derivatives are $O(\epsilon^{3/2})$ in L^∞. Thus, the error estimate in this equation holds uniformly for all x (though of course the estimates here depend on n). Next, by Lemma 15.3.2,

$$\|W(x,\epsilon) - x\|_2 \leq \frac{\epsilon}{2}\|\nabla V(x)\|_2$$

and

$$\begin{aligned}W(x,\epsilon) - x &= -\frac{1}{2}\int_0^\epsilon \nabla V(W(x,s))\,ds \\ &= -\frac{1}{2}\epsilon \nabla V(x) + O\left(C\int_0^\epsilon \|W(x,s) - x\|_2\,ds\right) \\ &= -\frac{1}{2}\epsilon\nabla V(x) + O\left(\epsilon^2\|\nabla V(x)\|\right).\end{aligned}$$

Thus, substituting this into our earlier equation,

$$\begin{aligned}(\Psi_\epsilon - 1)\Phi_\epsilon f_t(x) &= J(\Phi_\epsilon f_t)(x)\left[-\frac{1}{2}\epsilon\nabla V(x) + O(\epsilon^2\|\nabla V(x)\|_2)\right] + O(\epsilon^{-1/2}\epsilon^2\|\nabla V(x)\|_2) \\ &= -\frac{1}{2}\epsilon J(\Phi_\epsilon f_t)(x)\nabla V(x) + O(\epsilon^{3/2}\|\nabla V(x)\|_2),\end{aligned}$$

where again the implicit constants are independent of x. Since $\|\nabla V(x)\|_2$ is bounded on any compact set and u is compactly supported and in L^1, we get

$$\int (\Psi_\epsilon - 1)\Phi_\epsilon f_t \cdot u = -\frac{1}{2}\epsilon \int J(\Phi_\epsilon f_t) \cdot \nabla V \cdot u + O(\epsilon^{3/2}).$$

Now Φ_ϵ satisfies $\int(\Phi_\epsilon f)g = \int(\Phi_\epsilon g)f$ for any functions f and g, and hence

$$\int (\Psi_\epsilon - 1)\Phi_\epsilon f_t \cdot u = -\frac{1}{2}\epsilon \int Jf_t) \cdot \Phi_\epsilon[u\nabla V] + O(\epsilon^{3/2}).$$

By the same token

$$\int (\Phi_\epsilon - 1) f_t u = \int f_t (\Phi_\epsilon - 1) u.$$

Therefore, overall,

$$\begin{aligned}\int (T_\epsilon - 1) f_t u &= \int (\Psi_\epsilon \Phi_\epsilon - 1) f_t u + O(\epsilon^{3/2}) \\ &= -\frac{1}{2}\epsilon \int Jf_t \cdot \Phi_\epsilon[u\nabla V] + \int f_t(\Phi_\epsilon - 1)u + O(\epsilon^{3/2}).\end{aligned}$$

If we divide by ϵ and send $\epsilon \searrow 0$, then we get

$$-\frac{1}{2}\int Jf_t \cdot (u\nabla V) + \frac{1}{2n^2}\int f_t \Delta u,$$

because $(\Phi_\epsilon - 1)u/\epsilon \to (1/2n^2)\Delta u$ uniformly and $\Phi_\epsilon[u\nabla V] \to u\nabla V$ uniformly as $u\nabla V$ is Lipschitz (∇V being Lipschitz on the support of u). Finally, it follows from integration by parts that

$$\int f_t \Delta u = -\int Jf_t \cdot \nabla u,$$

since u is smooth and compactly supported (this uses a standard approximation argument since f_t is only differentiable almost everywhere with Jf_t in L^∞). Therefore,

$$\lim_{\epsilon \searrow 0} \frac{1}{\epsilon} \int (f_{t+\epsilon} - f_t) = -\frac{1}{2}\int Jf_t \cdot \left(u\nabla V + \frac{1}{n^2}\nabla u\right).$$

Moreover, our error estimates only depend on the Lipschitz norm of f_t (which is bounded by the Lipschitz norm of f) on the L^1 norm of f_t on the support of u. In other words, if we denote $h(t) = \int f_t u$, we have

$$h'_+(t) := \lim_{\epsilon \searrow 0} \frac{h(t+\epsilon) - h(t)}{\epsilon} = -\frac{1}{2}\int Jf_t \cdot \left(u\nabla V + \frac{1}{n^2}\nabla u\right),$$

and the rate of convergence is independent of t. Now it is clear that

$$|h'_+(t)| \leq \frac{1}{2}\|f_t\|_{\mathrm{Lip}} \|u\nabla V + (1/n^2)\nabla u\|_{L^1},$$

which is a bounded function of t. Hence, $h(t+\epsilon)-h(t) = O(\epsilon)$ uniformly for all t, which implies that h is Lipschitz. Hence, h is differentiable almost everywhere and $h' = h'_+$ whenever h is differentiable. □

Lemma 15.3.8. *Let μ be the measure given by the potential V. Then for Lipschitz f, we have $\int T_t f \, d\mu = \int f \, d\mu$.*

Proof. It follows from the previous lemma that for any $u \in C_c^\infty(M_n(\mathbb{C})_{\text{sa}}^d)$, we have

$$\int (T_t f - f) u = -\frac{1}{2} \int_0^t \left(\int J f_s \cdot \left(u \nabla V + \frac{1}{n^2} \nabla u \right) \right) ds.$$

Using standard approximation arguments, we can substitute $u = e^{-n^2 \Phi_\epsilon V}$ since it is smooth, and this function and its first derivatives decay rapidly at ∞ since $cI \leq H(\Phi_\epsilon V) \leq CI$. Therefore,

$$\int (T_t f - f) u = -\frac{1}{2} \int_0^t \left(\int J f_s \cdot \left(e^{-n^2 \Phi_\epsilon V} \nabla V - \nabla(\Phi_\epsilon V) e^{-n^2 \Phi_\epsilon V} \right) \right) ds.$$

But

$$\begin{aligned} \left\| \int J f_s \cdot \left(e^{-n^2 \Phi_\epsilon V} \nabla V - \nabla(\Phi_\epsilon V) e^{-n^2 \Phi_\epsilon V} \right) \right\|_2 &\leq \|J f_s\|_{L^\infty} \int \|\nabla V - \nabla(\Phi_\epsilon V)\|_2 e^{-n^2 \Phi_\epsilon V} \\ &\leq \|f\|_{\text{Lip}} C \epsilon^{1/2} \int e^{-n^2 V}. \end{aligned}$$

Here we used that $\|(\Phi_\epsilon - 1)\nabla V\|_{L^\infty} \leq C\epsilon^{1/2}$ since ∇V is C-Lipschitz. And we also observed that $\Phi_\epsilon V \geq V$ by Jensen's inequality because V is convex, and so $e^{-n^2 \Phi_\epsilon V} \leq e^{-n^2 V}$.

Therefore,

$$\left\| \int (T_t f - f) e^{-n^2 \Phi_\epsilon V} \right\|_2 \leq t \|f\|_{\text{Lip}} C \epsilon^{1/2} \int e^{-n^2 V}.$$

Since $T_t f - f$ is Lipschitz, we know $\int (T_t f - f) e^{-n^2 V}$ is finite, and hence by dominated convergence

$$\int (T_t f - f) e^{-n^2 V} = \lim_{\epsilon \searrow 0} \int (T_t f - f) e^{-n^2 \Phi_\epsilon V} = 0.$$

Thus, $\int (T_t f) e^{-n^2 V} = \int f e^{-n^2 V}$, so that $\int T_t f \, d\mu = \int f \, d\mu$ as desired. □

Lemma 15.3.9. *For a Lipschitz function f, we have $T_t f \to \int f \, d\mu$ as $t \to +\infty$ and in fact*

$$\left\| T_t f(x) - \int f \, d\mu \right\|_2 \leq e^{-ct/2} \|f\|_{\text{Lip}} \left(\|x\|_2 + \frac{1}{c} \|\nabla V(0)\|_2 + \frac{3d^{1/2}}{c^{1/2}} \right).$$

Proof. We apply Corollary 11.2.8 to the function $T_t f$ and the potential $n^2 V$ to obtain

$$\left\| T_t f(x) - \int T_t f \, d\mu \right\|_2 \leq \|T_t f\|_{\text{Lip}} \left(\|x\|_2 + \frac{1}{n^2 c} \|n^2 \nabla V(0)\| + \frac{3(n^2 d)^{1/2}}{(n^2 c)^{1/2}} \right).$$

Then all the factors of n cancel, and we substitute $\int T_t f \, d\mu = \int f \, d\mu$ and $\|T_t f\|_{\text{Lip}} \leq e^{-c/2} \|f\|_{\text{Lip}}$. □

15.4 Conditional diffusion semigroup

In the last section, we gave an explicit iterative construction of the heat semigroup T_t associated to a potential $V(x)$. In this section, we consider a potential $V(x,y)$ on $M_n(\mathbb{C})_{\text{sa}}^{d_1+d_2}$ with $cI \leq HV \leq CI$, and we the conditional heat semigroup where the differentiation occurs with respect to x, and y serves merely as a parameter. As in the last section, n and V will be fixed, and thus we will not write out any superscript (n).

Let μ denote the measure associated to V, and let

$$d\mu(x|y) = \frac{1}{\int e^{-n^2 V(x,y)}\,dx} e^{-n^2 V(x,y)}\,dx.$$

Our main goal is to show that if $f(x,y)$ is a Lipschitz function with values in $M_n(\mathbb{C})$, then $T_t f(x,y)$ is also Lipschitz in y, and then by taking $t \to +\infty$, we will obtain that $\int f(x,y)\,d\mu(x|y)$ is Lipschitz in y.

It will be convenient to consider the Lipschitz seminorms of a function with respect to x and y individually, and hence we introduce the notation

$$\|f\|_{\text{Lip},dx} = \sup\left\{\frac{\|f(x,y) - f(x',y)\|}{\|x - x'\|} : x, x' \in M_n(\mathbb{C})_{\text{sa}}^{d_1}, y \in M_n(\mathbb{C})_{\text{sa}}^{d_2}, x \neq x'\right\}$$
$$\|f\|_{\text{Lip},dy} = \sup\left\{\frac{\|f(x,y) - f(x,y')\|}{\|y - y'\|} : x \in M_n(\mathbb{C})_{\text{sa}}^{d_1}, y, y' \in M_n(\mathbb{C})_{\text{sa}}^{d_2}, y \neq y'\right\}$$

Changing notation slightly from the last section, let

$$[\Phi_t f](x,y) = \int f(x+z,y)\,d\sigma_t(z).$$

We also define $W : M_n(\mathbb{C})_{\text{sa}}^{d_1+d_2} \times [0,+\infty) \to M_n(\mathbb{C})_{\text{sa}}^{d_1}$ by

$$\partial_t W(x,y,t) = -\frac{1}{2}\nabla_x V(W(x,y,t),y), \qquad W(x,y,0) = x.$$

As we mentioned before, W is well-defined by the Picard-Lindelöf theorem since ∇V is C-Lipschitz by Lemma 11.1.4. Then we set

$$[\Psi_t f](x,y) = f(W(x,y,t),y).$$

For a partition $\mathcal{P} = \{t_0, \dots, t_k\}$ of the interval $[s,t]$, we set

$$T_{\mathcal{P}} f = \Phi_{t_k - t_{k-1}} \Psi_{t_k - t_{k-1}} \dots \Phi_{t_1 - t_0} \Psi_{t_1 - t_0} f.$$

By freezing the variable y and using the results from the previous section, it is clear that Φ_t, Ψ_t, and T_t are semigroups acting on the space of functions with $\|f\|_{\text{Lip},dx} < \infty$. Moreover, the results from the last section can be summarized as follows.

Proposition 15.4.1. *There exists a semigroup T_t acting on functions $M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \times M_n(\mathbb{C})$ with $\|f\|_{\mathrm{Lip},dx} < +\infty$ such that the following hold.*

(1) If $\mathcal{P}$ is a partition of $[s,t]$, then

$$\|T_{t-s}f - T_{\mathcal{P}}f\|_{L^\infty} \leq \frac{Cd^{1/2}}{c}(1 - e^{-c(t-s)/2}) \operatorname{mesh}(\mathcal{P})^{1/2} \|f\|_{\mathrm{Lip}}.$$

(2) $\int T_t f(x,y)\, d\mu(x|y) = \int f(x,y)\, d\mu(x|y)$ for all y.

(3) $\|T_t f(x,y) - \int f(x,y)\, d\mu(x|y)\|_2 \leq e^{-ct/2}\|f\|_{\mathrm{Lip}}(\|x\|_2 + (1/c)\|\nabla_x V(0,y)\|_2 + 3d^{1/2}/c^{1/2})$.

Proof. (1) follows from Lemma 15.3.6 (3). (2) follows from Lemma 15.3.8. (3) follows from Lemma 15.3.9. □

Now we move on to estimate the Lipschitz seminorms of Φ_t, Ψ_t, and T_t with respect to x and y.

Lemma 15.4.2. *Let $f : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to M_n(\mathbb{C})$ be Lipschitz. Then*

(1) $\|\Phi_t f\|_{\mathrm{Lip},dx} \leq \|f\|_{\mathrm{Lip},dx}$,

(2) $\|\Phi_t f\|_{\mathrm{Lip},dy} \leq \|f\|_{\mathrm{Lip},dy}$,

(3) $\|\Phi_t f\|_{L^\infty} \leq \|f\|_{L^\infty}$.

Proof. These properties are immediate since Φ_t is convolution with a probability measure. □

Lemma 15.4.3. *Let $W_t(x,y) = W(x,y,t)$. Then $\|W_t\|_{\mathrm{Lip},dx} \leq e^{-ct/2}$ and $\|W_t\|_{\mathrm{Lip},dy} \leq (C/c)(1 - e^{-ct/2})$. Therefore, if $f : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to M_n(\mathbb{C})$ is Lipschitz, we have*

(1) $\|\Psi_t f\|_{\mathrm{Lip},dx} \leq e^{-ct/2}\|f\|_{\mathrm{Lip},dx}$,

(2) $\|\Psi_t f\|_{\mathrm{Lip},dy} \leq \|f\|_{\mathrm{Lip},dy} + (C/c)(1 - e^{-ct/2})\|f\|_{\mathrm{Lip},dx}$,

(3) $\|\Psi_t f\|_{L^\infty} \leq \|f\|_{L^\infty}$.

Proof. We begin by proving the claims about W_t. Fix $x, x' \in M_n(\mathbb{C})_{\mathrm{sa}}^{d_1}$ and $y, y' \in M_n(\mathbb{C})_{\mathrm{sa}}^{d_2}$. Define

$$\phi(t) = \|W_t(x,y) - W_t(x',y')\|_2.$$

Note that ϕ is locally Lipschitz in t and hence absolutely continuous. Moreover, $\phi(t)^2$ is C^1 with

$$\begin{aligned}
\partial_t[\phi(t)^2] &= 2\langle \partial_t W_t(x,y) - \partial_t W_t(x',y'), W_t(x,y) - W_t(x',y')\rangle_2 \\
&= -\langle \nabla_x V(W_t(x,y),y) - \nabla_x V(W_t(x',y'),y'), W_t(x,y) - W_t(x',y')\rangle_2 \\
&= -\langle \nabla_x V(W_t(x,y),y) - \nabla_x V(W_t(x',y'),y), W_t(x,y) - W_t(x',y')\rangle_2 \\
&\quad - \langle \nabla_x V(W_t(x',y'),y) - \nabla_x V(W_t(x',y'),y'), W_t(x,y) - W_t(x',y')\rangle_2 \\
&\leq -c\|W_t(x,y) - W_t(x',y')\|_2^2 \\
&\quad + \|\nabla_x V(W_t(x',y'),y) - \nabla_x V(W_t(x',y'),y')\|_2 \|W_t(x,y) - W_t(x',y')\|_2 \\
&\leq -c\|W_t(x,y) - W_t(x',y')\|_2^2 + C\|y-y'\|_2 \|W_t(x,y) - W_t(x',y')\|_2.
\end{aligned}$$

Here we have employed the inequality $\langle \nabla_x V(z,w) - \nabla_x V(z',w), z-z'\rangle_2 \geq c\|z-z'\|_2^2$ coming from the uniform convexity of V as well as the Cauchy-Schwarz inequality. This implies that

$$2\phi'(t)\phi(t) = \partial_t[\phi(t)^2] \leq -c\phi(t)^2 + C\|y-y'\|\phi(t).$$

Thus, $\phi'(t) \leq -(c/2)\phi(t) + (C/2)\|y-y'\|$, so that $\partial_t[e^{ct/2}\phi(t)] \leq (C/2)e^{ct/2}\|y-y'\|_2$. This implies that

$$e^{ct/2}\phi(t) - \phi(0) \leq \frac{C}{c}(e^{ct/2} - 1)\|y-y'\|_2.$$

But $\phi(t) = \|W_t(x,y) - W_t(x',y')\|_2$ and $\phi(0) = \|x-x'\|_2$. Hence,

$$\|W_t(x,y) - W_t(x',y')\|_2 \leq e^{-ct/2}\|x-x'\|_2 + \frac{C}{c}(1-e^{-ct/2})\|y-y'\|_2.$$

This proves both of our estimates for $\|W_t\|_{\mathrm{Lip},dx}$ and $\|W_t\|_{\mathrm{Lip},dy}$ at once.

(1) is immediate from our claim about W_t, and indeed was already shown in Lemma 15.3.2.

(2) Note that

$$\begin{aligned}
\|\Psi_t f(x,y) - \Psi_t f(x,y')\|_2 &= \|f(W_t(x,y),y) - f(W_t(x,y'),y')\|_2 \\
&\leq \|f(W_t(x,y),y) - f(W_t(x,y),y')\|_2 + \|f(W_t(x,y),y') - f(W_t(x,y'),y')\|_2 \\
&\leq \|f\|_{\mathrm{Lip},dy}\|y-y'\|_2 + \|f\|_{\mathrm{Lip},dx}\|W_t(x,y) - W_t(x,y')\|_2 \\
&\leq \|f\|_{\mathrm{Lip},dy}\|y-y'\|_2 + \frac{C}{c}(1-e^{-ct/2})\|f\|_{\mathrm{Lip},dx}\|y-y'\|_2.
\end{aligned}$$

(3) is immediate, as in Lemma 15.3.2. □

Lemma 15.4.4. *If $f : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to M_n(\mathbb{C})$ is Lipschitz, we have*

(1) $\|T_t f\|_{\mathrm{Lip},dx} \leq e^{-ct/2}\|f\|_{\mathrm{Lip},dx}$,

(2) $\|T_t f\|_{\mathrm{Lip},dy} \leq \|f\|_{\mathrm{Lip},dy} + (C/c)(1-e^{-ct/2})\|f\|_{\mathrm{Lip},dx}$,

(3) $\|T_t f\|_{L^\infty} \leq \|f\|_{L^\infty}$.

The same holds for $T_{\mathcal{P}}$ *whenever* $\mathcal{P}$ *is a partition of* $[0,t]$.

Proof. First, consider the case of $T_{\mathcal{P}}$. The estimates (1) and (3) follow from iterating our estimates for Φ_t and Ψ_t, or from Lemma 15.3.6.

To prove (2), let $\mathcal{P} = \{t_0, \dots, t_k\}$, and for each j, let $\mathcal{P}_j = \{t_0, \dots, t_j\}$. We prove the claim inductively for the partition $\mathcal{P}_j$. In the base case, we adopt the convention that $T_{\mathcal{P}_0} f = f$, and hence the claim holds trivially. For the induction step, observe that $T_{\mathcal{P}_{j+1}} f = \Phi_{t_{j+1}-t_j} \Psi_{t_{j+1}-t_j} T_{\mathcal{P}} f$. Therefore, using the two previous lemmas, and the induction hypothesis,

$$\|T_{\mathcal{P}_{j+1}} f\|_{\mathrm{Lip},dy} \leq \|\Psi_{t_{j+1}-t_j} T_{\mathcal{P}_j} f\|_{\mathrm{Lip},dy} \tag{15.8}$$

$$\leq \|T_{\mathcal{P}_j} f\|_{\mathrm{Lip},dy} + \frac{C}{c}(1 - e^{-c(t_{j+1}-t_j)/2}) \|T_{\mathcal{P}_j} f\|_{\mathrm{Lip},dx} \tag{15.9}$$

$$\leq \left(\|f\|_{\mathrm{Lip},dy} + \frac{C}{c}(1 - e^{-ct_j/2}) \|f\|_{\mathrm{Lip},dx} \right) + \frac{C}{c}(1 - e^{-c(t_{j+1}-t_j)/2}) e^{-ct_j/2} \|f\|_{\mathrm{Lip},dx} \tag{15.10}$$

$$= \|f\|_{\mathrm{Lip},dy} + \frac{C}{c}(1 - e^{-ct_{j+1}/2}) \|f\|_{\mathrm{Lip},dx}. \tag{15.11}$$

This completes (2).

Finally, the case of T_t follows from the case of $T_{\mathcal{P}}$ by letting $\mathrm{mesh}(\mathcal{P}) \to 0$ and using Proposition 15.4.1 (1). □

Corollary 15.4.5. *Let* $f : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to M_n(\mathbb{C})$ *be Lipschitz. Let* $g(y) = \int f(x,y)\, d\mu(x|y)$. *Then* g *is Lipschitz with*

$$\|g\|_{\mathrm{Lip}} \leq (1 + C/c) \|f\|_{\mathrm{Lip}}.$$

Proof. By the previous lemma,

$$\|T_t f\|_{\mathrm{Lip},dy} \leq \|f\|_{\mathrm{Lip},dy} + \frac{C}{c}(1 - e^{-ct/2}) \|f\|_{\mathrm{Lip},dx}$$
$$\leq \|f\|_{\mathrm{Lip}} + \frac{C}{c} \|f\|_{\mathrm{Lip}}.$$

As $t \to \infty$, we have $T_t f(x,y) \to g(y)$ by Proposition 15.4.1 (3). Hence, $\|g\|_{\mathrm{Lip}} \leq (1 + C/c) \|f\|_{\mathrm{Lip}}$. □

15.5 Asymptotic approximation and convergence

Finally, we are ready to prove Theorem 15.1.2. To this end, fix $V^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to \mathbb{R}$ such that $cI \leq HV^{(n)} \leq CI$ for some $0 < c \leq C$ and $(\nabla V^{(n)})_{n \in \mathbb{N}}$ is asymptotically approximable by trace polynomials. Let $\mu^{(n)}$ be the measure associated to the potential $n^2 V^{(n)}$. Let $\mu^{(n)}(x|y)$ be the conditional distribution associated to $\mu^{(n)}$.

Let $P_t^{(n)}$, $S_t^{(n)}$, and $T_t^{(n)}$ be the semigroups acting on Lipschitz functions defined as in §15.4 with respect to the potential $V^{(n)}$.

Lemma 15.5.1. *With the notation above, suppose that $f^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to M_n(\mathbb{C})$, that $f^{(n)}$ is K-Lipschitz for every n, and that $f^{(n)}$ is asymptotically approximable by trace polynomials. Then*

(1) $(\Phi_t^{(n)} f^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials,

(2) $(\Psi_t^{(n)} f^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials,

(3) $(T_t^{(n)} f^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials.

Proof. (1) We proved in Lemma 14.2.6 that $\Phi_t^{(n)}$ preserves asymptotic approximability by trace polynomials.

(2) Recall that $\Psi_t^{(n)} f^{(n)}(x,y) = f^{(n)}(W_t^{(n)}(x,y), y)$, where

$$\partial_t W_t^{(n)}(x,y) = -\frac{1}{2}\nabla_x V^{(n)}(W_t^{(n)}(x,y),y). \qquad W_0^{(n)}(x,y) = x$$

Now $\nabla_x V^{(n)}(x,y)$ is C-Lipschitz in (x,y), asymptotically approximable by trace polynomials, and independent of t, and thus it satisfies Assumption 14.3.6, so by Proposition 14.3.7, $W_t^{(n)}(x,y)$ is asymptotically approximable by trace polynomials (here we rely on Lemma 13.5.3 that asymptotic approximability is equivalent to being asymptotic to some element of $\overline{\mathrm{TrP}}_{m+n}^1$). Then because $f^{(n)}$ is K-Lipschitz in (x,y), Proposition 13.5.12 implies asymptotic approximability of $f^{(n)}(W_t^{(n)}(x,y),y)$.

(3) From inductive application of (1) and (2), we see that for any partition $\mathcal{P}$ of $[0,t]$, the sequence $(T_{\mathcal{P}}^{(n)} f^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials, where of course $T_{\mathcal{P}}^{(n)}$ is the operator defined in §15.5 for the potential $V^{(n)}$. But by Proposition 15.4.1 (1), as $\operatorname{mesh}(\mathcal{P}) \to 0$, we have $T_{\mathcal{P}}^{(n)} f^{(n)} \to T_t^{(n)} f^{(n)}$ in L^∞ with a rate of convergence independent of n since $\|f^{(n)}\|_{\mathrm{Lip}} \leq K$ for all n. Therefore, $(T_t^{(n)} f^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials by Lemma 13.5.13. □

Proof of Theorem 15.1.2. Let $f^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to M_n(\mathbb{C})$ be K-Lipschitz and asymptotically approximable by trace polynomials. Let

$$g^{(n)}(y) = \int f^{(n)}(x,y)\, d\mu^{(n)}(x|y)$$

We showed in Corollary 15.4.5 that $g^{(n)}$ is Lipschitz with $\|g^{(n)}\|_{\mathrm{Lip}} \leq (1 + C/c)\|f^{(n)}\|_{\mathrm{Lip}}$.

Moreover, we know that $(T_t^{(n)} f^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials in (x,y) by the previous lemma. And by Proposition 15.4.1 (3), we have $T_t^{(n)} f^{(n)}(x,y) \to g^{(n)}(y)$ as $t \to \infty$, with the error bounded by

$$e^{-ct/2}\left(\|x\|_2 + \frac{1}{c}\|\nabla_x V(0,y)\|_2 + \frac{3d^{1/2}}{c^{1/2}}\right)\|f^{(n)}\|_{\mathrm{Lip}}.$$

Given that $(\nabla V^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials, $\|\nabla_x V\|_{2,R}^{(n)}$ is bounded as $n \to \infty$ for each $R > 0$. This implies that the rate of convergence of $T_t^{(n)} f^{(n)}(x,y) \to g^{(n)}(x,y)$ as $t \to \infty$ is uniform on $\|(x,y)\|_\infty \leq R$ and independent of n. So by Lemma 13.5.13, $g^{(n)}$ is asymptotically approximable by trace polynomials of (x,y). Yet $g^{(n)}$ is independent of x, and so we may approximate $g^{(n)}(y)$ by evaluating these trace polynomials at $(0,y)$, which reduces them to trace polynomials of y. $\square$

CHAPTER 16

Results: Free entropy and Fisher information

In this chapter, we show that free entropy theory describes the large-n limit of classical entropy for random matrix models from potentials $V^{(n)}$ satisfying Assumption 15.1.1. In particular, we show that for non-commutative random variables in the large-n limit, the microstates entropy $\chi(X)$ agrees with non-microstates entropy $\chi^*(X)$, and this is the same as the large-n limit of the classical entropy of the random matrix models. Moreover, if $(X^{(n)}, Y^{(n)})$ is a $(d_1 + d_2)$-tuple of random self-adjoint matrices arising from such potentials $V^{(n)}$, then the conditional classical entropy $h(X^{(n)}|Y^{(n)})$ converges to the conditional non-microstates free entropy $\chi * (X|Y)$.

16.1 Microstates free entropy

This section explains Voiculescu's microstates free entropy χ and $\underline{\chi}$ and states sufficient conditions for the microstates free entropies χ and $\underline{\chi}$ to be evaluated as the lim sup and lim inf of renormalized classical entropies of random matrix models.

Because there is no integral formula known for free entropy of multiple non-commuting variables as in the classical case, Voiculescu defined the free analogue of entropy [Voi93, Voi94] using Boltzmann's microstates viewpoint on entropy.

Definition 16.1.1. For $\mathcal{U} \subseteq \Sigma_d$, we define the *microstate space*

$$\Gamma^{(n)}(\mathcal{U}) := \{x \in M_N(\mathbb{C})_{\text{sa}}^m : \lambda_x \in \mathcal{U}\}$$
$$\Gamma_R^{(n)}(\mathcal{U}) := \{x \in M_N(\mathbb{C})_{\text{sa}}^m : \lambda_x \in \mathcal{U}, \|x\|_\infty \leq R\}.$$

The *microstates free entropy* of a non-commutative law λ is defined as

$$\chi_R(\lambda) := \inf_{\mathcal{U} \ni \lambda} \limsup_{n \to \infty} \left(\frac{1}{n^2} \log \operatorname{vol} \Gamma_R^{(n)}(\mathcal{U}) + d \log n \right)$$
$$\chi(\lambda) := \sup_{R > 0} \chi_R(\lambda).$$

Here vol denotes the Lebesgue measure with respect to some isometric identification of $M_n(\mathbb{C})_{\text{sa}}^d$ with $\mathbb{R}^{dn^2}$, and $\mathcal{U}$ ranges over all open neighborhoods of λ in Σ_d. Similarly, we define

$$\underline{\chi}_R(\lambda) := \inf_{\mathcal{U} \ni \lambda} \liminf_{n \to \infty} \left(\frac{1}{n^2} \log \operatorname{vol} \Gamma_R^{(n)}(\mathcal{U}) + d \log n \right)$$
$$\underline{\chi}(\lambda) := \sup_{R > 0} \chi_R(\lambda).$$

Similarly, for a self-adjoint d-tuple X from a tracial W*-algebra $(\mathcal{M}, \tau)$, we write $\chi(X) := \chi(\lambda)$.

Remark 16.1.2. In Voiculescu's original work [Voi94, Def. 2.1], the additive normalizing constant would be $(d/2) \log n$ rather than $d \log n$. However, we have chosen a different normalization of Lebesgue measure. Indeed, we identified $M_n(\mathbb{C})_{\text{sa}}^d$ isometrically with $\mathbb{R}^{dn^2}$ using the inner product $\langle x, y \rangle_2 = \sum_j \tau_n(x_j y_j)$ rather than the inner product $\langle x, y \rangle_{\text{Tr}} = \sum_j \operatorname{Tr}(x_j y_j) = n \langle x, y \rangle_2$. Heuristically, the normalizing constant can be checked by examining the case of a Gaussian. Alternatively, it is clear from Theorem 16.4.1 that the normalizing constant is "correct."

Remark 16.1.3. Note that $\mathcal{U} \subseteq \mathcal{V}$ implies that

$$\limsup_{n \to \infty} \left(\frac{1}{n^2} \log \operatorname{vol} \Gamma_R^{(n)}(\mathcal{U}) + d \log n \right) \leq \limsup_{n \to \infty} \left(\frac{1}{n^2} \log \operatorname{vol} \Gamma_R^{(n)}(\mathcal{V}) + d \log n \right).$$

Hence, to estimate the infimum over $\mathcal{U}$ (that is, $\chi_R(\lambda)$), we can always restrict our attention to neighborhoods $\mathcal{U}$ contained inside some fixed $\mathcal{V}$. The same holds for the lim inf variant of entropy.

The next proposition states sufficient conditions for the renormalized classical entropies of random matrix models to converge to the free entropy of a limiting non-commutative law. It can be regarded as a precise version of Voiculescu's heuristics for the microstates free entropy in [Voi94]. Although our main focus in the dissertation is on the potentials $V^{(n)}$ satisfying Assumption 15.1.1, we state the proposition in greater generality since it has independent interest.

Proposition 16.1.4. *Let $V^{(n)} : M_n(\mathbb{C})_{\text{sa}}^d \to \mathbb{R}$ be a potential with $\int e^{-n^2 V^{(n)}} < \infty$, let $\mu^{(n)}$ be the associated measure, and let $X^{(n)}$ be a random variable chosen according to $\mu^{(n)}$. Assume the following.*

(A) There is a non-commutative law λ such that $\lambda_{X^{(n)}} \to \lambda$ in probability.

(B) The sequence $(V^{(n)})_{n \in \mathbb{N}}$ is asymptotically approximable by scalar-valued trace polynomials.

(C) There exists $R_0 > 0$ such that

$$\lim_{n \to \infty} \int_{\|x\|_\infty \geq R_0} \left(1 + |V^{(n)}(x)|\right) \, d\mu^{(n)}(x) = 0.$$

Then λ can be realized as the law of non-commutative random variables $X = (X_1, \dots, X_m)$ in a von Neumann algebra $(\mathcal{M}, \tau)$ with $\|X_j\| \leq R_0$. Moreover, we have

$$\chi(\lambda) = \chi_{R_0}(\lambda) = \limsup_{n \to \infty} \left(\frac{1}{n^2} h(X^{(n)}) + d \log n \right) \tag{16.1}$$

$$\underline{\chi}(\lambda) = \underline{\chi}_{R_0}(\lambda) = \liminf_{n \to \infty} \left(\frac{1}{n^2} h(\mu^{(n)}) + d \log n \right). \tag{16.2}$$

Proof. It follows from assumptions (A) and (C) that for every non-commutative polynomial p,

$$\lim_{n \to \infty} \int_{\|x\|_\infty \leq R_0} \tau_n(p(x))\, d\mu^{(n)}(x) = \lambda(p).$$

In particular, this implies that $|\lambda(X_{i_1} \dots X_{i_k})| \leq R_0^k$ for every $i_1, \dots, i_k$, and hence $\lambda \in \Sigma_{d,R_0}$. The non-commutative λ can be realized by operators in a von Neumann algebra by Proposition 10.2.2.

Now let us evaluate χ_R and $\underline{\chi}_R$ for $R \geq R_0$. Recall that

$$d\mu^{(n)}(x) = \frac{1}{Z^{(n)}} e^{-n^2 V^{(n)}(x)}\, dx, \text{ where } Z^{(n)} = \int e^{-n^2 V^{(n)}},$$

and therefore

$$h(\mu^{(n)}) = n^2 \int V^{(n)}(x)\, d\mu^{(n)}(x) + \log Z^{(n)}.$$

Assumption (C) implies that

$$\lim_{n \to \infty} \int_{\|x\|_\infty \geq R} |V^{(n)}(x)|\, d\mu^{(n)}(x) = 0 \text{ and } \lim_{n \to \infty} \mu^{(n)}(x : \|x\|_\infty \geq R) = 0.$$

Therefore, if we let

$$d\mu_R^{(n)}(x) = \frac{1}{Z_R^{(n)}} \mathbf{1}_{\|x\|_\infty \leq R} e^{-n^2 V^{(n)}(x))}\, dx, \text{ where } Z_R^{(n)} = \int_{\|x\|_\infty \leq R} e^{-n^2 V^{(n)}},$$

then as $n \to \infty$, we have

$$\int V^{(n)}\, d\mu^{(n)} - \int V^{(n)}\, d\mu_R^{(n)} \to 0, \qquad \log Z^{(n)} - \log Z_R^{(n)} \to 0,$$

and hence

$$\frac{1}{n^2} h(\mu^{(n)}) - \frac{1}{n^2} h(\mu_R^{(n)}) \to 0.$$

Fix $\epsilon > 0$. By assumption (B), there is scalar-valued trace polynomial f such that $|V^{(n)}(x) - f(x)| \leq \epsilon/2$ for $\|x\|_\infty \leq R$ and for sufficiently large n. Now because the trace polynomial f is continuous with respect to convergence in non-commutative moments, the set

$\mathcal{U} = \{\lambda' : |\lambda'(f) - \lambda(f)| < \epsilon/2\}$ is open in $\Sigma_{d,R}$. Now suppose that $\mathcal{V} \subseteq \mathcal{U}$ is a neighborhood of λ. Note that

$$\lim_{n\to\infty} \mu_R^{(n)}(\Gamma_R^{(n)}(\mathcal{V})) = \lim_{n\to\infty} \frac{Z^{(n)}}{Z_R^{(n)}} \mu^{(n)}(\Gamma^{(n)}(\mathcal{V}) \cap \{x : \|x\|_\infty \leq R\}) = 1,$$

where we have used that $Z^{(n)}/Z_R^{(n)} \to 1$ as shown above, that $\mu^{(n)}(\Gamma^{(n)}(\mathcal{V})) \to 1$ by assumption (A), and that $\mu^{(n)}(\|x\|_\infty \leq R) \to 1$ by assumption (C). Moreover, by our choice of f and $\mathcal{U}$, we have

$$x \in \Gamma_R^{(n)}(\mathcal{V}) \implies \left|V^{(n)}(x) - \lambda(f)\right| \leq \epsilon.$$

Therefore,

$$\begin{aligned} Z_R^{(n)} \mu_R^{(n)} \left(\Gamma_R^{(n)}(\mathcal{V})\right) &= \int_{\Gamma_R^{(n)}(\mathcal{V})} e^{-n^2 V^{(n)}(x)}\, dx \\ &= e^{-n^2(\lambda(f) + O(\epsilon))} \operatorname{vol} \Gamma_R^{(n)}(\mathcal{V}). \end{aligned}$$

Thus,

$$\log Z_R^{(n)} + \log \mu_R^{(n)} \left(\Gamma_R^{(n)}(\mathcal{V})\right) = \log \operatorname{vol} \Gamma_R^{(n)}(\mathcal{V}) - n^2(\lambda(f) + O(\epsilon)).$$

Meanwhile, note that $|f(x)|$ is bounded by some constant K whenever $\|x\|_\infty \leq R$ (where K is independent of n). Therefore,

$$\begin{aligned} \int V^{(n)}\, d\mu_R^{(n)} &= \int_{\Gamma_R^{(n)}(\mathcal{V})} V^{(n)}\, d\mu_R^{(n)} + \int_{\Gamma_R^{(n)}(\mathcal{V}^c)} V^{(n)}\, d\mu_R^{(n)} \\ &= \int_{\Gamma_R^{(n)}(\mathcal{V})} \lambda[f]\, d\mu_R^{(n)} + \int_{\Gamma_R^{(n)}(\mathcal{V}^c)} \lambda_x[f]\, d\mu_R^{(n)} + O(\epsilon) \\ &= \lambda(f)\, \mu_R^{(n)} \left(\Gamma_R^{(n)}(\mathcal{V})\right) + O(\epsilon) + O\left(K\, \mu^{(n)} \left(\Gamma_R^{(n)}(\mathcal{V}^c)\right)\right). \end{aligned}$$

Altogether,

$$\begin{aligned} \frac{1}{n^2} h(\mu_R^{(n)}) = &\int V^{(n)}\, d\mu_R^{(n)} + \frac{1}{n^2} \log Z_{N,R} \\ =&\lambda(f) \left(\mu_R^{(n)} \left(\Gamma_R^{(n)}(\mathcal{V})\right) - 1\right) + \frac{1}{n^2} \log \operatorname{vol} \Gamma_R^{(n)}(\mathcal{V}) \\ &+ O(\epsilon) + O\left(K\, \mu^{(n)} \left(\Gamma_R^{(n)}(\mathcal{V}^c)\right)\right) - \frac{1}{n^2} \log \mu_R^{(n)}(\Gamma_R^{(n)}(\mathcal{V})). \end{aligned}$$

Now we apply the fact that $\mu_R^{(n)}(\Gamma_R^{(n)}(\mathcal{V})) \to 1$ to obtain

$$\limsup_{n\to\infty} \frac{1}{n^2} \left|h(\mu_R^{(n)}) - \log \operatorname{vol} \Gamma_R^{(n)}(\mathcal{V})\right| = O(\epsilon).$$

In light of Remark 16.1.3, because this holds for all sufficiently small neighborhoods $\mathcal{V} \subseteq \mathcal{U}$ with the error $O(\epsilon)$ only depending on $\mathcal{U}$, we have

$$\begin{aligned} \chi_R(\lambda) &= \limsup_{n\to\infty} \left(\frac{1}{n^2} h(\mu_R^{(n)}) + d\log n \right) + O(\epsilon) \\ &= \limsup_{n\to\infty} \left(\frac{1}{n^2} h(\mu^{(n)}) + d\log n \right) + O(\epsilon). \end{aligned}$$

Next, we take $\epsilon \to 0$ and obtain $\chi_R(\lambda) = \limsup_{n\to\infty}(n^{-2}\log h(\mu^{(n)}) + (d/2)\log n)$ for $R \geq R_0$. Now $\chi(\lambda) = \sup_R \chi_R(\lambda)$ and $\chi_R(\lambda)$ is an increasing function of R. Since our claim about $\chi_R(\lambda)$ holds for sufficiently large R, it also holds for $\chi(\lambda)$, so (16.1) is proved. The proof of (16.2) is identical. □

Corollary 16.1.5. *If $V^{(n)}$ satisfies Assumption 15.1.1 and $\mu^{(n)}$ is the corresponding measure, then* (16.1) *and* (16.2) *hold.*

Proof. Since $V^{(n)}$ is only determined by $\mu^{(n)}$ up to an additive constant, assume without loss of generality that $V^{(n)}(0) = 0$. Let us verify the hypotheses of the previous proposition.

(A) We showed in Theorem 15.1.5 that $\lambda_{X^{(n)}}$ converges almost surely to some non-commutative law λ.

(B) Fix $R > 0$ and $\epsilon > 0$. We have assumed that $(\nabla V^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials, so there exists a self-adjoint trace polynomial d-tuple f such that

$$\limsup_{n\to\infty} \sup_{\|x\|_\infty \leq R} \left\| \nabla V^{(n)}(x) - f(x) \right\|_2 \leq \frac{\epsilon}{R}.$$

Since $V^{(n)}(0) = 0$ by assumption, we have

$$V^{(n)}(x) = \int_0^1 \left\langle \nabla V^{(n)}(tx), x \right\rangle dt.$$

We may define a scalar-valued trace polynomial g by

$$g(x) = \int_0^1 \sum_{j=1}^d \tau(f_j(tx)x_j)\, dt.$$

By the triangle inequality for integrals,

$$|V^{(n)}(x) - g(x)| \leq \frac{\epsilon}{R}\|x\|_\infty \leq \epsilon \text{ for } \|x\|_\infty \leq R.$$

Hence, $(V^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials.

(C) Since $0 \leq HV^{(n)} \leq CI$ and $V^{(n)}(0) = 0$, we have

$$\langle \nabla V^{(n)}(x), x \rangle_2 \leq V^{(n)}(x) \leq \langle \nabla V^{(n)}(0), x \rangle + \frac{C}{2} \|x\|_2^2.$$

Because $(\nabla V^{(n)})_{n \in \mathbb{N}}$ is asymptotically approximable by trace polynomials, we have in particular that $\sup_n \|\nabla V(0)\|_2 < \infty$. Meanwhile, since $HV^{(n)} \geq cI$, we know that $\mu^{(n)}$ satisfies the log-Sobolev inequality with constant $n^2 c$, and hence by Corollary 11.5.3, there is some $R_0 > 0$ such that

$$\lim_{n \to \infty} \int_{\|x\|_\infty \geq R_0} e^{K \|x\|_\infty} \, d\mu^{(n)}(x) = 0.$$

Combined with our bound on $V^{(n)}(x)$, this easily implies (C).

□

16.2 Free Fisher information

In this section, we explain Voiculescu's free Fisher information, and we state sufficient conditions for classical Fisher information for random matrix models to converge to free Fisher information in the large-n limit.

Voiculescu's definition of free Fisher information [Voi98a] is motivated by the integration-by-parts characterization of the score function in classical probability (12.1). Suppose that $X^{(n)}$ is a random d-tuple of self-adjoint matrices with distribution

$$d\mu^{(n)}(x) = e^{-n^2 V^{(n)}(x)} \, dx.$$

If $V^{(n)}$ is C^1, then the score function is given by $\Xi^{(n)} = n^2 \nabla V^{(n)}(X)$. If we apply the integration-by-parts relation $E\langle \Xi^{(n)}, f(X) \rangle = E[\nabla^\dagger f(X)]$ to a tuple of self-adjoint non-commutative polynomials $p = (p_1, \dots, p_d)$ and use Lemma 14.1.9, then we obtain

$$E\langle \nabla V^{(n)}(X), p(X) \rangle_2 = E\langle n^{-2} \Xi, p(X) \rangle_2 = \sum_{j=1}^d E[\tau_n \otimes \tau_n(\mathcal{D}_{X_j} p_j(X^{(n)}))], \tag{16.3}$$

Of course, this is only justified under sufficient assumptions of finite moments; see for instance Lemma 11.2.5. Since we may choose all the p_j's to be zero except for one of them, we can equivalently write this relation as

$$E\langle n^{-2} \Xi_j, p(X) \rangle_2 = E[\tau_n \otimes \tau_n(\mathcal{D}_{X_j} p(X))]$$

for all j, where $\Xi = (\Xi_1, \dots, \Xi_d)$. Moreover, if this holds for self-adjoint polynomials, then it holds for all polynomials by linearity. Voiculescu therefore made the following definitions.

Definition 16.2.1 ([Voi98a, §3]). Let $X = (X_1, \dots, X_d)$ be a tuple of self-adjoint random variables in a tracial von Neumann algebra $(\mathcal{M}, \tau)$ and assume that $\mathcal{M}$ is generated by X as a von Neumann algebra. We say that $\xi = (\xi_1, \dots, \xi_d) \in L^2(\mathcal{M}, \tau)_{\text{sa}}^d$ is a *free score function* for X if

$$\tau(\xi_j p(X)) = \tau \otimes \tau(\mathcal{D}_{X_j} p(X)) \tag{16.4}$$

for every non-commutative polynomial p and every j.

Note that the free score function, if it exists, is unique. This is because a vector in $L^2(\mathcal{M}, \tau)$ is uniquely determined by its inner product with vectors in a dense subspace, such as the non-commutative polynomials in X.

Definition 16.2.2. If a free score function exists, we say that X (or equivalently the law of X) then we define the *free Fisher information* $\Phi^*(X) := \|\xi\|_2^2 = \sum_j \tau(\xi_j^2)$. If there is no free score function, then we set $\Phi^*(X) = \infty$.

We might expect free Fisher information to be the limit of renormalized classical Fisher information of random matrix models, since under sufficient assumptions $E\|\nabla V^{(n)}(X^{(n)})\|^2$ should converge to $\|f(X)\|_2^2$ if $\nabla V^{(n)} \rightsquigarrow f$. Actually, we will prove this more generally in the conditional setting.

The conditional versions of classical and free Fisher information are a straightforward generalization of the definition we have already given. Compare the classical case of conditional score functions defined in Definition 12.4.2.

Definition 16.2.3. If (X, Y) is a $(d_1 + d_2)$-tuple of non-commutative random variables from $(\mathcal{M}, \tau)$ with $\mathcal{M} = \mathrm{W}^*(X, Y)$, then $\xi \in L^2(\mathcal{M}, \tau)_{\text{sa}}^{d_1}$ is a *free score function* for X given Y if for every j and every non-commutative polynomial p in $d_1 + d_2$ variables, we have

$$\tau(\xi_j p(X, Y)) = \tau \otimes \tau[\mathcal{D}_{X_j} p(X_j, Y_j)].$$

If a free score function for X given Y exists, then we define the free Fisher information by

$$\Phi^*(X|Y) = \|\xi\|_2^2,$$

and otherwise the free Fisher information is defined to be ∞.

The following proposition gives sufficient conditions for the classical conditional Fisher information to converge to the free conditional Fisher information. This is a version of [Jek18, Proposition 5.10].

Proposition 16.2.4. *Let $V^{(n)} : M_n(\mathbb{C})_{\text{sa}}^{d_1+d_2} \to \mathbb{R}$ be a C^1 potential with $\int e^{-n^2 V^{(n)}} < \infty$, let $\mu^{(n)}$ be the associated measure, and let $(X^{(n)}, Y^{(n)})$ be a random variable with distribution $\mu^{(n)}$. Let (X, Y) be a self-adjoint $(d_1 + d_2)$-tuple in a tracial W^*-algebra $(\mathcal{M}, \tau)$. Assume the following.*

(A) $\lambda_{(X^{(n)},Y^{(n)})} \to \lambda_{(X,Y)}$ *in probability.*

(B) The sequence $(\nabla_x V^{(n)})_{n\in\mathbb{N}}$ *is asymptotically approximable by trace polynomials.*

(C) For some $R_0 > 0$*, we have*

$$\lim_{n\to\infty} \int_{\|(x,y)\|_\infty > R_0} (1 + \|\nabla_x V^{(n)}(x,y)\|_2^2)\, d\mu^{(n)}(x,y) = 0.$$

Let $\nabla_x V^{(n)} \rightsquigarrow f \in C_{\text{tr,app}}(\mathbb{R}^{*(d_1+d_2)}, \|\cdot\|_2)_{\text{sa}}^{d_1}$*. Then* $f(X,Y)$ *is the free score function for* X *given* Y*, and we have*

$$\lim_{n\to\infty} \frac{1}{n^4} \mathcal{I}(X^{(n)}|Y^{(n)}) \to \Phi^*(X|Y) < \infty.$$

Proof. The idea of the proof is that the classical integration-by-parts relation involving $\nabla_x V^{(n)}(X^{(n)}, Y^{(n)})$ leads to the free integration-by-parts relation involving $f(X,Y)$ in the large-n limit. However, as we have not assumed that $\mu^{(n)}$ has finite moments, we must justify this more carefully.

First, note that since $\nabla_x V^{(n)}$ is asymptotically approximable by trace polynomials, we have

$$\sup_n \sup_{\|(x,y)\|_\infty \le R_0} \|\nabla_x V^{(n)}\|_2 < \infty.$$

Together with (C), this implies that

$$\int (1 + \|\nabla_x V\|_2^2\, d\mu^{(n)}(x,y) < \infty$$

for sufficiently large n.

As in Proposition 16.1.4, we have $\|(X,Y)\|_\infty \le R_0$. Consider a self-adjoint tuple of non-commutative polynomials $p = (p_1, \dots, p_{d_1})$. Let $\psi \in C_c^\infty(\mathbb{R}, \mathbb{R})$ satisfy $\psi(t) = t$ for $|t| \le \mathbb{R}$, and consider the function

$$g(x,y) = p(\psi(x_1), \dots, \psi(x_{d_1}), \psi(y_1), \dots, \psi(y_{d_2})).$$

Then g is globally $\|\cdot\|_2$-Lipschitz and bounded in operator norm. Moreover, g is a C^∞ function on $M_n(\mathbb{C})_{\text{sa}}^d$ for each n. This follows from results that we will prove later in §18.1; specifically, Propositions 18.1.5 and 18.1.6 imply that ψ defines a smooth function $M_n(\mathbb{C})_{\text{sa}} \to M_n(\mathbb{C})_{\text{sa}}$. Using integration by parts, for any $R > 0$

$$\int_{\|(x,y)\|_2 \le R} \langle \nabla_x V^{(n)}, g \rangle_2\, d\mu^{(n)} = \frac{1}{n^2} \int_{\|(x,y)\|_2 \le R} \nabla_x^\dagger g\, d\mu^{(n)} + \frac{1}{n^2} \frac{1}{\int e^{-n^2 V^{(n)}}} \int_{\|x\|_2 = R} \langle g(x,y), x/R \rangle_2 e^{-n^2 V^{(n)}(x,y)}\, d\sigma_R(x,y),$$

where σ_R denotes surface measure on the sphere. Since g is bounded, we have

$$\int |\langle g(x), x/R\rangle_2| e^{-n^2 V^{(n)}(x,y)}\, dx\, dy < \infty,$$

and hence there is a sequence R_k tending to ∞ such that

$$\lim_{k\to\infty} \int_{\|x\|_2=R_k} \langle g(x,y), x/R\rangle_2 e^{-n^2 V^{(n)}(x,y)}\, d\sigma_{R_k}(x,y) = 0.$$

Since g and $\nabla^\dagger g$ are bounded functions for a fixed n and $\int (1 + \|\nabla_x V^{(n)}\|_2^2)\, d\mu^{(n)} < \infty$, we have by Cauchy-Schwarz that

$$\int |\langle \nabla_x V^{(n)}, g\rangle_2|\, d\mu^{(n)} < \infty, \qquad \int |\nabla_x^\dagger g|\, d\mu^{(n)} < \infty,$$

and hence by dominated convergence

$$\lim_{k\to\infty} \int_{\|(x,y)\|_2 \le R_k} |\langle \nabla_x V^{(n)}, g\rangle_2|\, d\mu^{(n)} = \int_{\|(x,y)\|_2 \le R_k} \langle \nabla_x V^{(n)}, g\rangle_2\, d\mu^{(n)}$$

and

$$\lim_{k\to\infty} \int_{\|(x,y)\|_2 \le R_k} \nabla^\dagger g\, d\mu^{(n)} \to \int \nabla_x^\dagger g\, d\mu^{(n)}.$$

Therefore, overall,

$$\int \langle \nabla_x V^{(n)}, g\rangle_2\, d\mu^{(n)} = \frac{1}{n^2} \int \nabla^\dagger g\, d\mu^{(n)}.$$

Using (C) and Cauchy-Schwarz,

$$\lim_{n\to\infty} \int_{\|(x,y)\|_\infty > R_0} \langle \nabla_x V^{(n)}, g\rangle_2\, d\mu^{(n)} = 0.$$

Hence,

$$\begin{aligned}\lim_{n\to\infty} \int \langle \nabla_x V^{(n)}, g\rangle_2\, d\mu^{(n)} &= \lim_{n\to\infty} \int_{\|(x,y)\|_\infty \le R_0} \langle \nabla_x V^{(n)}, g\rangle_2\, d\mu^{(n)} \\ &= \langle f(X,Y), p(X,Y)\rangle_2,\end{aligned}$$

where the last line follows because $\nabla_x V^{(n)} \rightsquigarrow f$ and $\lambda_{(X^{(n)},Y^{(n)})} \to \lambda_{(X,Y)}$ in probability (the argument is similar to those used in the proof of Theorem 15.1.7). Similarly, we claim that

$$\lim_{n\to\infty} \int_{\|(x,y)\|_\infty > R_0} \nabla_x^\dagger g\, d\mu^{(n)} = 0.$$

This follows from (C) and the fact that $(1/n^2)|\nabla^\dagger g|$ is bounded uniformly bounded. More precisely, g is K-Lipschitz for some K independent of n, hence $\|J_x g\| \le K$, where $J_x g$ is the

Jacobian linear transformation. But $(1/n^2)|\nabla^\dagger g| = (1/n^2)|\operatorname{Tr}(J_x g)| \leq (1/n^2)(d_1 n^2)\|J_x g\| \leq d_1 K$. Thus, we have

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n^2} \int \nabla_x^\dagger g \, d\mu^{(n)} &= \lim_{n\to\infty} \frac{1}{n^2} \int \nabla_x^\dagger p \, d\mu^{(n)} \\
&= \lim_{n\to\infty} \sum_{j=1}^{d_1} \int \tau_n \otimes \tau_n[\mathcal{D}_{X_j} p] \, d\mu^{(n)} \\
&= \sum_{j=1}^{d_1} \tau \otimes \tau[\mathcal{D}_{X_j} p(X,Y)],
\end{aligned}$$

where we have used that $g = p$ when $\|(x,y)\|_\infty \leq R$, and in the last line we have used the convergence in non-commutative law in probability again. Therefore, we have

$$\sum_{j=1}^{d_1} \tau(f_j(X,Y) p_j(X,Y)) = \tau \otimes \tau[\mathcal{D}_{X_j} p_j(X,Y)].$$

Therefore, $f(X,Y)$ is the free score function for X given Y.

By construction of $C_{\mathrm{tr,app}}(\mathbb{R}^{*(d_1+d_2)}, \|\cdot\|_2)$, we have $f(X,Y) \in L^2(\mathcal{M},\tau)$. Hence, $\Phi^*(X|Y) = \|f(X,Y)\|_2^2 < \infty$. Finally, since $\nabla_x V^{(n)} \rightsquigarrow f$, we have $\|\nabla_x V^{(n)}\|_2^2 \rightsquigarrow \|f\|_2^2$. Using assumption (C) and convergence of $\lambda_{(X^{(n)},Y^{(n)})}$ to $\lambda_{(X,Y)}$ in probability again,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n^4} \mathcal{I}(X^{(n)}|Y^{(n)}) &= \lim_{n\to\infty} \int \left\|\nabla_x V^{(n)}\right\|_2^2 d\mu^{(n)} \\
&= \lim_{n\to\infty} \int_{\|(x,y)\|_\infty \leq R_0} \left\|\nabla_x V^{(n)}\right\|_2^2 d\mu^{(n)} \\
&= \|f(X,Y)\|_2^2 = \Phi^*(X|Y). \qquad \square
\end{aligned}$$

Corollary 16.2.5. *The conclusions of the previous proposition hold for the random matrix models coming from any sequence of potentials $(V^{(n)})_{n\in\mathbb{N}}$ satisfying Assumption 15.1.1.*

Proof. We check the assumptions of the proposition.

(A) follows from Theorem 15.1.5.

(B) is part of Assumption 15.1.1.

(C) Note that $\sup_n \|\nabla_x V^{(n)}(0)\|_2 < \infty$ because $(\nabla V^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials. Also, $\nabla V^{(n)}$ is C-Lipschitz by Lemma 11.1.4 since $0 \leq HV^{(n)} \leq C$. Thus, we have $\|\nabla_x V^{(n)}(x,y)\|_2^2 \leq K(1 + \|x\|_2^2 + \|y\|_2^2)$ for some constant K. Hence, (C) follows from Corollary 11.5.3. $\square$

16.3 Free score functions and non-microstates free entropy

In this section, we explain further properties of the conditional free Fisher information and Voiculescu's non-microstates free entropy $\chi^*(X|Y)$. Voiculescu [Voi98a] defined the non-microstates free entropy χ^* by integrating the free Fisher information in analogy with the formula (12.4) (see Definition 16.3.4 below). As in the classical case, to show convergence of this integral formula, we need the conditional free version of Lemma 12.1.3, which is due to [Voi98a, Remark 3.2(a)].

Lemma 16.3.1. *Let Y be a d_2-tuple of self-adjoint variables in $(\mathcal{M}, \tau)$, and let X and Y be d_1-tuples of self-adjoint variables from $\mathcal{M}$. Suppose that X and Y are free with amalgamation over $\mathrm{W}^*(Y)$, or in other words, $\mathrm{W}^*(X,Y)$ and $\mathrm{W}^*(Z,Y)$ are free with respect to the conditional expectation $E_{\mathrm{W}^*(Y)} : \mathcal{M} \to \mathrm{W}^*(Y)$. If ξ is a free score function for X given Y, then $E_{\mathrm{W}^*(X+Z,Y)}$ is a free score function for $X+Z$ given Y.*

Proof. We claim that for any non-commutative polynomial $g(x,y,z)$, we have

$$\tau(\xi_j g(X,Y,Z)) = \tau \otimes \tau[\partial_{x_j} g(X,Y,Z)].$$

Taking this claim for granted for the moment, note that for a non-commutative polynomial $g(x,y)$, we have $\partial_{x_j}[g(x+z,y)] = \partial_{x_j} g(x+z,y)$ and hence

$$\tau(\xi_j g(X+Z,Y)) = \tau \otimes \tau[\partial_{x_j} g(X+Z,Y)].$$

The inner product does not change if we replace ξ_j by its expectation onto $\mathrm{W}^*(X+Z,Y)$, hence $E_{\mathrm{W}^*(X+Z,Y)}[\xi]$ is a free score function for $X+Z$ given Y.

The claim is well-known in free probability, but let us recall the proof for the sake of completeness. Let $\mathcal{B} = \mathrm{W}^*(Y)$. Note that any non-commutative polynomial in (X,Y,Z) can be expressed as a polynomial in (X,Z) with coefficients in $\mathcal{B}$. The differentiation operator with respect to x_j produces a well-defined map $\mathcal{D}_{X_j:\mathcal{B}} : \mathcal{B}\langle X,Z\rangle \to \mathcal{B}\langle X,Z\rangle \otimes \mathcal{B}\langle X,Z\rangle$. Moreover, since elements of $\mathcal{B}$ can be approximated in L^2 norm by polynomials in Y, we have

$$\tau(\xi_j G(X) = \tau \otimes \tau[\mathcal{D}_{X_j:\mathcal{B}} G(X)]$$

for any $G \in \mathcal{B}\langle X\rangle$. To accomplish our goal, the claim we want to prove is that

$$\tau(\xi_j G(X,Z)) = \tau \otimes \tau[\mathcal{D}_{X_j:\mathcal{B}} G(X,Z)]$$

for any $G \in \mathcal{B}\langle X,Z\rangle$. As in the proof of the free case of Lemma 5.2.8, $\mathcal{B}\langle X,Z\rangle$ is spanned by elements of $\mathcal{B}$ and alternating products of terms from $\mathcal{B}\langle X\rangle$ and $\mathcal{B}\langle Z\rangle$ with expectation zero onto $\mathcal{B}$, so it suffices to consider $G(X,Z)$ of this form. If $G(X,Z) \in \mathcal{B}$ or is an element in $\mathcal{B}\langle X\rangle$ with expectation zero, then $G(X,Z)$ is already in $\mathcal{B}\langle X\rangle$, so there is nothing to prove.

Thus, it suffices to consider such alternating products with at least one element from $\mathcal{B}\langle Z\rangle$ with expectation zero onto $\mathcal{B}$. Any such product can be written

$$G(X,Z) = G_0(X) H_1(Z) G_1(X) \dots H_k(Z) G_k(X), \tag{16.5}$$

where $k \geq 1$ and where $E_{\mathcal{B}}[H_j(Z)] = 0$ for all j and $E_{\mathcal{B}}[G_j(X)] = 0$ for all j except possibly the endpoints 0 and k. Note that

$$E_{\mathcal{B}}\left[[G_0(X) - E_{\mathcal{B}}[G_0(X)]]H_1(Z)G_1(X)\dots H_k(Z)[G_k(X) - E_{\mathcal{B}}[G_k(X)]]\right] = 0,$$

which after some computation implies that

$$E_{\mathcal{B}}\left[G_0(X)H_1(Z)G_1(X)\dots H_k(Z)G_k(X)\right] = 0,$$

hence

$$\tau\left(G_0(X)H_1(Z)G_1(X)\dots H_k(Z)G_k(X)\right) = 0.$$

By the same token, since ξ_j is in the L^2-closure of $\mathcal{B}\langle X\rangle$, we have

$$\tau\left(\xi_j G_0(X)H_1(Z)G_1(X)\dots H_k(Z)G_k(X)\right) = 0.$$

On the other hand, note that

$$\begin{aligned}&\partial_{x_j} G(X,Z) \\ &\quad = \sum_{i=0}^{k} \left(G_0(X)H_1(Z)\dots G_{i-1}(Z)H_i(Z)\right)\partial_{x_j}G_i(X)\left(H_{i+1}(Z)G_{i+1}(X)\dots H_k(Z)G_k(X)\right).\end{aligned} \tag{16.6}$$

We can write $\partial_{x_j}G_i(X)$ as a sum of simple tensors in $\mathcal{B}\langle X\rangle$. This turns (16.6) into a sum of simple tensors, where each side of the tensor sign has a string of the form (16.5). Either the left side or the right side of each term has to have at least one element from $\mathcal{B}\langle Z\rangle$, which implies that $\tau \otimes \tau$ will evaluate to zero on the tensor product. Thus, in this case,

$$\tau(\xi_j G(X,Z)) = 0 = \tau \otimes \tau[\partial_{x_j} G(X,Z)]. \qquad \square$$

We also record another well-known observation in free probability.

Lemma 16.3.2. *Let (X,Y) be a non-commutative (d_1+d_2)-tuple from $(\mathcal{M},\tau)$, and suppose that Z is a d_3-tuple free from (X,Y). Then Z is free from X with amalgamation over* $\mathrm{W}^*(Y)$.

Proof. Let $\mathcal{B} = \mathrm{W}^*(Y)$. To demonstrate freeness, it suffices to show that

$$E_{\mathcal{B}}[G_0(X)H_1(Z)G_1(X)\dots H_k(Z)G_k(Z)] = 0$$

whenever $k \geq 1$, $G_i(X) \in \mathcal{B}\langle X\rangle$ for $0 \leq i \leq k$ with $E_{\mathcal{B}}[G_i(X)] = 0$ for $0 < i < k$, and $H_i(Z) \in \mathcal{B}\langle Z\rangle$ with $E_{\mathcal{B}}[H_i(Z)] = 0$ for $1 \leq i \leq k$. By definition of the conditional expectation, this reduces to showing that for $b \in \mathcal{B}$, we have

$$\tau[G_0(X)H_1(Z)G_1(X)\dots H_k(Z)G_k(Z)b] = 0.$$

But b can be absorbed into $G_k(Z)$, so we can assume without loss of generality that $b = 1$.

Now $\mathcal{B}\langle Z \rangle$ is spanned by alternating strings of the form $b_0 f_1(Z_{\iota(1)}) b_1 \dots f_\ell(Z_{\iota(\ell)}) b_\ell$ where $\tau(f_j(Z_{\iota(j)})) = 0$ for $1 \leq j \leq \ell$ and $\tau(b_j) = 0$ for $0 < j < \ell$. In the case $\ell = 0$, these are elements of $\mathcal{B}$, while if $\ell > 0$, they are orthogonal to $\mathcal{B}$ since Z is free from $\mathcal{B}$. Thus, the elements of $\mathcal{B}\langle Z \rangle$ with expectation zero onto $\mathcal{B}$ are spanned by such strings with $\ell \geq 1$. Hence, we can assume that each $H_i(Z)$ has the form

$$H_i(Z) = b_{i,0} f_{i,1}(Z_{\iota(i,1)}) b_{i,1} \dots f_{i,\ell(i)}(Z_{\iota(i,\ell(i))}) b_{i,\ell(i)}$$

with the conditions described above. Plugging in this choice of $H_i(Z)$ and regrouping terms produces a string composed out of terms like

(1) $b_{i,j}$ for $0 < j < \ell(i)$ and $1 \leq i \leq k$,

(2) $f_{i,j}(Z_{i,\iota(i,j)})$ for $1 \leq j \leq \ell(i)$ and $1 \leq i \leq k$,

(3) $b_{i,\ell(i)} G_i(X) b_{i+1,0}$ for $0 < i < k$, or $G_0(X) b_{1,0}$ or $b_{k,\ell(k)} G_k(X)$ in the endpoint cases.

Note that that $\tau(b_{i,\ell(i)} G_i(X) b_{i+1,0}) = 0$ since $E_{\mathcal{B}}[G_i(X)] = 0$. Thus, all the terms in the string have trace zero except the endpoint terms, and also the terms alternate between $\mathrm{W}^*(X, Y)$ and $\mathrm{W}^*(Z)$, and hence the whole string has expectation zero. $\square$

Using the previous two lemmas, we can obtain the following.

Lemma 16.3.3. *Let (X, Y) be a self-adjoint $(d_1 + d_2)$-tuple from $(\mathcal{M}, \tau)$, and let Z_t be a freely independent d_1-tuple with non-commutative law σ_t. Let $a = \|X\|_2^2 / d_1$. Then*

$$\frac{d_1}{a + t} \leq \Phi^*(X + Z_t | Y) \leq \min\left(\frac{d_1}{t}, \mathcal{I}(X|Y) \right).$$

Proof. The proof is much the same as for the classical case (12.2). For the lower bound, note that if ξ_t is a free score function for $X + Z_t$ given Y, then

$$d_1^2 = \langle \xi, X + Z_t \rangle_2^2 \leq \|\xi_t\|_2^2 \|X + Z_t\|_2^2 = \Phi^*(X + Z_t|Y)(d_1 a + d_1 t).$$

For the upper bound, note that X and Z_t are freely independent with amalgamation over $\mathrm{W}^*(Y)$. If Ξ is a free score function for X given Y, then by the previous lemma, $\Xi_t = E[\Xi | X + Z_t, Y]$, so that $\Phi^*(X + Z_t|Y) \leq \Phi * (X|Y)$. Moreover, one can check that Z_t/t is a free score function for Z_t given Y (for instance, this follows from Proposition 16.2.4 since $Z_t^{(n)}/t$ is a score function for the Gaussian random matrix tuple with distribution $\sigma_t^{(n)}$ which converges in non-commutative law to the Gaussian by applying Lemma 14.2.4 with $d_1 = 0$). Thus, $\Xi_t = E[Z_t/t | X + Z_t, Y]$. $\square$

Definition 16.3.4 ([Voi98a, Definition 7.1])**.** Let (X, Y) be a $(d_1 + d_2)$-tuple from a tracial W^*-algebra $(\mathcal{M}, \tau)$. The *non-microstates free entropy* of X given Y is defined as

$$\chi^*(\lambda) := \frac{1}{2} \int_0^\infty \left(\frac{d}{1+t} - \Phi^*(X + Z_t | Y) \right) + \frac{d}{2} \log 2\pi e,$$

where Z_t is a free semicircular d_1-tuple with non-commutative law σ_t, freely independent of (X, Y).

Remark 16.3.5. It follows from the previous lemma that the positive part of the integrand $d/(1+t) - \Phi^*(X + Z_t|Y)$ is integrable, and hence the integral defining $\chi^*(X|Y)$ is well-defined in $[-\infty, \infty)$; the argument is the same as in Lemma 12.1.4.

Remark 16.3.6. Voiculescu's original notation in [Voi98a, §7] was $\Phi^*(X : \mathrm{W}^*(Y))$ and $\chi^*(X : \mathrm{W}^*(Y))$ rather than $\chi^*(X|Y)$, since the definition of the free score function can be rephrased so as to depend only on $\mathrm{W}^*(Y)$ rather than Y. The idea is to use $\mathrm{W}^*(Y)$-valued polynomials as we did in the proof of Lemma 16.3.1. However, we prefer to write $\chi^*(X|Y)$ instead by analogy with the classical case, using the vertical bar to denote "conditioning." This avoids potential confusion with the notation $\chi(X : Y)$ for microstates entropy of X "in the presence of Y" used in [Voi96, §1].

16.4 Convergence to conditional free entropy

Our main new result of the chapter shows that if $(X^{(n)}, Y^{(n)})$ is a random matrix $(d_1 + d_2)$-tuple coming from potentials $V^{(n)}$ as in Assumption 15.1.1, then the large-n limit of $h(X^{(n)}|Y^{(n)})$ is described by $\chi^*(X|Y)$. As a corollary, when $d_2 = 0$, we obtain that $\chi(X) = \chi^*(X)$.

Theorem 16.4.1. *Let $V^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to \mathbb{R}$ satisfy Assumption 15.1.1 for some $0 < c \leq C$, let $(X^{(n)}, Y^{(n)})$ be a random variable distributed according to the associated measure $\mu^{(n)}$, let $Z_t^{(n)}$ be an independent d_1-tuple with the Gaussian distribution $\sigma_t^{(n)}$. Let (X, Y) be a self-adjoint $(d_1 + d_2)$-tuple from a tracial W^*-algebra realizing the limiting free Gibbs law as in Theorem 15.1.5. Then for every $t \geq 0$, we have*

$$\lim_{n\to\infty} \frac{1}{n^4} \mathcal{I}(X^{(n)} + Z_t^{(n)} | Y^{(n)}) = \Phi^*(X + Z_t|Y). \tag{16.7}$$

and

$$\lim_{n\to\infty} \left(\frac{1}{n^2} h(X^{(n)} + Z_t^{(n)} | Y^{(n)}) + d_1 \log n \right) = \chi^*(X + Z_t|Y). \tag{16.8}$$

Corollary 16.4.2. *Let $V^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{R}$ satisfy Assumption 15.1.1. Let $X^{(n)}$ be the associated random matrix d-tuple and let X be a non-commutative d-tuple realizing the free Gibbs law as in Theorem 15.1.5. Then we have*

$$\chi(X) = \underline{\chi}(X) = \chi^*(X).$$

Proof. By Proposition 16.1.4 and Corollary 16.1.5 imply that $\chi(X)$ and $\underline{\chi}(X)$ can be evaluated respectively as the lim sup and lim inf of the normalized classical entropy $n^{-2}h(X^{(n)}) + d\log n$. But by Theorem 16.4.1, the normalized classical entropy converges to $\chi^*(X)$. □

In the proof of Theorem 16.4.1, the greatest difficulty is to show that the potential $V_t^{(n)}$ associated to $(X^{(n)} + Z_t^{(n)}, Y^{(n)})$ still satisfies Assumption 15.1.1, and in particular that $(\nabla V_t^{(n)})_{n\in\mathbb{N}}$ is asymptotitcally approximable by trace polynomials. However, the results of §15 were designed for just this purpose.

Lemma 16.4.3. *Let $V^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to \mathbb{R}$ be a sequence of potentials satisfying Assumption 15.1.1 with $cI \leq HV^{(n)} \leq CI$, and let $\mu^{(n)}$ be the corresponding measure. Let*

$$V_t^{(n)}(x,y) = -\frac{1}{n^2}\log \int e^{-n^2 V^{(n)}(x-z,y)}\, d\sigma_t^{(n)}(z),$$

where $\sigma_t^{(n)}$ is the GUE measure.

(1) Then $(V_t^{(n)})_{n\in\mathbb{N}}$ also satisfies Assumption 15.1.1 for constants $c/(1+ct)$ and C.

(2) More precisely, we have

$$\frac{c}{1+ct} I_{n^2 d_1} \oplus cI_{n^2 d_2} \leq HV_t^{(n)} \leq \frac{C}{1+Ct} I_{n^2 d_1} \oplus CI_{n^2 d_2}.$$

(3) Moreover,

$$\left\| \nabla V_t^{(n)}(x,y) - \nabla V^{(n)}(x,y) \right\|_2 \leq C \left(\frac{2d_1 t}{1+ct} + \frac{t^2}{(1+ct)^2} \|\nabla_x V^{(n)}(x,y)\|_2^2 \right)^{1/2}. \quad (16.9)$$

(4) Let (X,Y) be a non-commutative (d_1+d_2)-tuple realizing the limiting free Gibbs law for $(V^{(n)})_{n\in\mathbb{N}}$ as in Theorem 15.1.5, and let Z_t be a freely independent semicircular d_1-tuple with law σ_t. Then the limiting free Gibbs law for $V_t^{(n)}$ is described by $(X+Z_t, Y)$.

Proof. For $z \in M_n(\mathbb{C})_{\mathrm{sa}}^{d_1}$, let

$$U_t^{(n)}(x,y,z) = V_t^{(n)}(x,y) + \frac{1}{2t}\|z\|_2^2,$$

and

$$W_t^{(n)}(x,y,z) = U_t^{(n)}(x-z,y,z).$$

Note that

$$V_t^{(n)} = -\frac{1}{n^2}\log \int e^{-n^2 W^{(n)}(x,y,z)}\, dz + \mathrm{const}(t,n).$$

The probabilistic interpretation is as follows. Letting $(X^{(n)}, Y^{(n)}) \sim \mu^{(n)}$ and $Z_t^{(n)} \sim \sigma_t^{(n)}$ be an independent Gaussian tuple, the potential $U_t^{(n)}$ gives the law of $(X^{(n)}, Y^{(n)}, Z_t^{(n)})$, and the potential $W_t^{(n)}$ gives the law of $(X^{(n)} + Z_t^{(n)}, Y^{(n)}, Z^{(n)})$; by taking the marginal on the first $d_1 + d_2$, we obtain the law of $(X^{(n)} + Z_t^{(n)}, Y^{(n)})$, which corresponds to the potential $V_t^{(n)}$.

(1), (2) Note that

$$\begin{pmatrix} cI & 0 & 0 \\ 0 & cI & 0 \\ 0 & 0 & t^{-1}I \end{pmatrix} \leq HU_t^{(n)} \leq \begin{pmatrix} cI & 0 & 0 \\ 0 & cI & 0 \\ 0 & 0 & t^{-1}I \end{pmatrix},$$

where the block decomposition is given by dividing the rows and columns in to blocks of sizes $n^2 d_1$, $n^2 d_2$, and $n^2 d_1$. Then $W_t^{(n)} = U_t^{(n)} \circ T^{-1}$, where

$$T = \begin{pmatrix} I & 0 & I \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}, \qquad T^{-1} = \begin{pmatrix} I & 0 & -I \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}$$

with respect to the same block decomposition. It follows from Lemma 11.3.1 that

$$\begin{pmatrix} cI & 0 & -cI \\ 0 & cI & 0 \\ -cT & 0 & (t^{-1} + c)I \end{pmatrix} \leq HW_t^{(n)} \leq \begin{pmatrix} cI & 0 & -CI \\ 0 & cI & 0 \\ -CI & 0 & (t^{-1} + C)I \end{pmatrix}.$$

Hence, Theorem 11.3.3, $V_t^{(n)}$ satisfies

$$\begin{pmatrix} (c/(1+ct))I & 0 \\ 0 & cI \end{pmatrix} \leq HV_t^{(n)} \leq \begin{pmatrix} (C/(1+Ct))I & 0 \\ 0 & CI \end{pmatrix},$$

where we have used the Schur complement computation

$$\begin{aligned} \begin{pmatrix} cI & 0 \\ 0 & cI \end{pmatrix} - \begin{pmatrix} -c \\ 0 \end{pmatrix} ((t^{-1} + c)I)^{-1} \begin{pmatrix} -c & 0 \end{pmatrix} &= \begin{pmatrix} (c - c^2(t^{-1} + c)^{-1}I & 0 \\ 0 & cI \end{pmatrix} \\ &= \begin{pmatrix} (c/(1+ct))I & 0 \\ 0 & cI \end{pmatrix}, \end{aligned}$$

and the same with c replaced by C. This proves the asserted claims about the convexity/concavity of $V_t^{(n)}$.

Observe that

$$\nabla W_t^{(n)}(x, y, z) = (\nabla_x V^{(n)}(x - z, y), \nabla_y V^{(n)}(x - z, y), z - \nabla_x V^{(n)}(x - z, y)),$$

which is asymptotically approximable by trace polynomials because $\nabla V^{(n)}$ is. Hence, $W_t^{(n)}$ satisfies Assumption 15.1.1. Therefore, by Proposition 15.1.8, $\nabla V_t^{(n)}$ is asymptotically approximable by trace polynomials.

(3) To prove (16.9), let us denote

$$d\nu_t^{(n)}(z|x,y) = \frac{1}{\int e^{-n^2 W_t^{(n)}(x,y,z)}\,dz} e^{-n^2 W_t(x,y,z)}\,dz,$$

so that

$$\nabla V_t^{(n)}(x,y) = \int \nabla_{(x,y)} W_t^{(n)}(x,y,z)\,d\nu_t^{(n)}(z|x,y) = \int \nabla V^{(n)}(x-z,y)\,d\nu^{(n)}(z|x,y).$$

Therefore,

$$\begin{aligned} \left\| \nabla V_t^{(n)}(x,y) - \nabla V^{(n)}(x,y) \right\|_2 &\leq \int \left\| \nabla V^{(n)}(x-z,y) - \nabla V^{(n)}(x,y) \right\| d\nu^{(n)}(z|x,y) \\ &\leq C \int \|z\|_2\,d\nu^{(n)}(z|x,y) \end{aligned}$$

since $\nabla V^{(n)}$ is C-Lipschitz. But recall that $H_z W_t(x,y,z) \geq (t^{-1}+c)I$. Therefore, applying Corollary 11.2.6 (5) to $\nu^{(n)}(z|x,y)$, we obtain

$$\begin{aligned} \int \|z\|_2\,d\nu^{(n)}(z|x,y) &\leq \left(\int \|z\|_2^2\,d\nu^{(n)}(zx,y) \right)^{1/2} \\ &\leq \left(\frac{2d}{t^{-1}+c} + \frac{1}{(t^{-1}+c)^2} \|\nabla_z W_t(x,y,0)\|_2^2 \right)^{1/2} \\ &= \left(\frac{2dt}{1+ct} + \frac{t^2}{(1+ct)^2} \left\| \nabla_x V^{(n)}(x,y) \right\|_2^2 \right)^{1/2}, \end{aligned}$$

which combined with our previous estimate finishes the proof of (16.9).

(4) By Theorem 15.1.5, we know $\lambda_{(X^{(n)}+Z_t^{(n)},Y^{(n)})}$ converges almost surely to some non-commutative law, and we only have to show that it agrees with the non-commutative law of $(X+Z_t,Y)$. This follows from Lemma 14.2.4. □

Proof of Theorem 16.4.1. To prove (16.7), let $V_t^{(n)}$ be the potential given by Lemma 16.4.3, which corresponds to $(X^{(n)}+Z_t^{(n)},Y^{(n)})$. By the lemma, $V_t^{(n)}$ satisfies Assumption 15.1.1. Hence, by Proposition 16.2.4 and Corollary 16.2.5, we have

$$\lim_{n\to\infty} \frac{1}{n^4} \mathcal{I}(X^{(n)}+Z_t^{(n)}|Y^{(n)}) = \Phi^*(X+Z_t|Y).$$

To prove (16.8), we begin with the formula for classical entropy in Lemma 12.4.4, which states that

$$h(X^{(n)}|Y^{(n)}) = \frac{1}{2} \int \left(\frac{d_1 n^2}{1+t} - \mathcal{I}(X^{(n)}+Z_{n^2 t}^{(n)}|Y^{(n)}) \right) dt + \frac{d_1 n^2}{2} \log 2\pi e.$$

Hence,

$$\frac{1}{n^2}h(X^{(n)}|Y^{(n)}) + \frac{1}{2}\int_0^\infty \left(\frac{d_1 n^2}{1+n^2 t} - \frac{d_1}{1+t}\right) dt$$
$$= \frac{1}{2}\int_0^\infty \left(\frac{d_1 n^2}{1+n^2 t} - \frac{1}{n^2}\mathcal{I}(X^{(n)} + Z_{n^2 t}^{(n)}|Y^{(n)})\right) dt + \frac{d_1}{2}\log 2\pi e.$$

But note that

$$\frac{1}{2}\int_0^\infty \left(\frac{d_1 n^2}{1+n^2 t} - \frac{d_1}{1+t}\right) dt = \lim_{a\to\infty}\frac{d_1}{2}\left(\int_0^a \frac{1}{1/n^2+t} - \frac{1}{1+t}\right) = \frac{d_1}{2}\log n^2 = d_1 \log n.$$

Moreover, by a change of variables,

$$\frac{1}{2}\int_0^\infty \left(\frac{d_1 n^2}{1+n^2 t} - \frac{1}{n^2}\mathcal{I}(X^{(n)} + Z_{n^2 t}^{(n)}|Y^{(n)})\right) dt = \frac{1}{2}\int_0^\infty \frac{d_1}{1+t} - \frac{1}{n^4}\mathcal{I}(X^{(n)} + Z_t^{(n)}|Y^{(n)})\, dt.$$

Hence,

$$\frac{1}{n^2}h(X^{(n)}|Y^{(n)}) + d_1 \log n = \frac{1}{2}\int_0^\infty \frac{d_1}{1+t} - \frac{1}{n^4}\mathcal{I}(X^{(n)} + Z_t^{(n)}|Y^{(n)})\, dt + \frac{d_1}{2}\log 2\pi e.$$

By (16.7), $n^{-4}\mathcal{I}(X^{(n)} + Z_t^{(n)}|Y^{(n)}) \to \Phi^*(X + Z_t|Y)$. In order to exchange the limit and integration, we use the dominated convergence theorem together with the bounds of Lemma 12.4.4. Let $a^{(n)} = (1/d_1)E\|X^{(n)}\|_2^2$ which converges to $\|X\|_2^2$ and hence is uniformly bounded. By Lemma 12.4.4,

$$\frac{d_1 n^2}{a^{(n)}/n^2 + (t/n^2)} \le \mathcal{I}(X^{(n)} + Z_t^{(n)}|Y^{(n)}) \le \min\left(\frac{d_1 n^2}{t/n^2}, \mathcal{I}(X^{(n)}|Y^{(n)})\right),$$

and hence

$$\frac{d_1}{a^{(n)} + t} \le \frac{1}{n^4}\mathcal{I}(X^{(n)} + Z_t^{(n)}|Y^{(n)}) \le \min\left(\frac{d_1}{t}, \frac{1}{n^4}\mathcal{I}(X^{(n)}|Y^{(n)})\right).$$

Since $\mathcal{I}(X^{(n)}|Y^{(n)})$ is bounded as $n \to \infty$, this estimate is sufficient to apply the dominated convergence theorem and exchange the limit with integration. By the definition of $\chi^(X|Y)$, this yields

$$\lim_{n\to\infty}\left(h(X^{(n)}|Y^{(n)}) + d_1 \log n\right) = \chi^*(X|Y).$$

But since $(X^{(n)} + Z_t^{(n)}, Y^{(n)})$ satisfies the hypotheses just as well as $(X^{(n)}, Y^{(n)})$, we get (16.8). □

Remark 16.4.4. In [Jek18, §7], we did not use the conditional expectation method to prove $\nabla V_t^{(n)}$ is asymptotically approximable by trace polynomials, but rather we analyzed the evolution of $\nabla V_t^{(n)}$ directly using PDE semigroups. The proof given in [Jek19] and here for convergence of entropy is thus considerably shorter.

CHAPTER 17

Results: Conditional transport to Gaussian

17.1 Main result and consequences

Let $V^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2}$ satisfy Assumption 15.1.1 and let $(X^{(n)}, Y^{(n)})$ be the corresponding random variable. Let $Z^{(n)}$ be an independent Gaussian tuple in $M_n(\mathbb{C})_{\mathrm{sa}}^{d_1}$. In this section, we will construct Lipschitz functions $F^{(n)}$ such that $(F^{(n)}(X^{(n)}, Y^{(n)}), Y^{(n)}) \sim (Z^{(n)}, Y^{(n)})$ and $F^{(n)}$ is asymptotically approximable by trace polynomials. The inverse function will be obtained by the same construction and this will lead to an isomorphism of von Neumann algebras for the non-commutative random variables in the large-n limit.

The construction of $F^{(n)}$ follows the same strategy as used by Otto and Villani [OV00] (see Lemma 12.3.1 and Theorem 12.3.5 above), but in the conditional setting (see §12.4). Specifically, for each y, we take as our initial measure (μ in Theorem 12.3.5) the conditional distribution of $X^{(n)}$ given $Y^{(n)} = y$, and as our target measure (m in Theorem 12.3.5) the Gaussian distribution $\sigma_t^{(n)}$. Unsurprisingly in light of Theorem 12.3.5, the resulting transport functions $F^{(n)}$ produce a coupling between $(X^{(n)}, Y^{(n)})$ and $(Z^{(n)}, Y^{(n)})$ that witnesses the conditional Talagrand inequality with respect to Gaussian measure, and in the large-n limit this produces W*-algebraic transport that witnesses the conditional free Talagrand inequality with respect to the law of a free semicircular family.

The main result of the section is as follows. Here we denote by $h_g^{(n)}(X^{(n)}|Y^{(n)})$ the conditional entropy of $X^{(n)}$ given $Y^{(n)}$ relative to the Gaussian measure $\sigma_t^{(n)}$. One can compute as in §12.5, after taking account of the normalization of $\sigma_t^{(n)}$, that

$$h_g^{(n)}(X^{(n)}|Y^{(n)}) = h(X^{(n)}|Y^{(n)}) - \frac{n^2}{2} E\|X^{(n)}\|_2^2 - \frac{d_1 n^2}{2} \log \frac{2\pi}{n^2}.$$

Theorem 17.1.1. *Let $V^{(n)}(x, y)$ be a sequence of potentials on $M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2}$ satisfying Assumption 15.1.1 for some $0 < c \leq C$, and let $\mu^{(n)}$ and $(X^{(n)}, Y^{(n)})$ be the corresponding probability measures and random variables. Let $Z^{(N)}$ be an independent GUE d_1-tuple. Let $\pi_1(x, y) = x$ and $\pi_2(x, y) = y$. Then there exist functions $F^{(n)}$, $G^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2} \to M_n(\mathbb{C})_{\mathrm{sa}}^{d_1}$ such that the following hold.*

(1) $(F^{(n)}, \pi_2) \circ (G^{(n)}, \pi_2) = \mathrm{id} = (G^{(n)}, \pi_2) \circ (F^{(n)}, \pi_2)$.

(2) We have $(F^{(n)}(X^{(n)}, Y^{(n)}), Y^{(n)}) \sim (Z^{(n)}, Y^{(n)})$ *and* $(G^{(n)}(Z^{(n)}, Y^{(n)}), Y^{(n)}) \sim (X^{(n)}, Y^{(n)})$ *in law.*

(3) $(F^{(n)})_{n\in\mathbb{N}}$ *and* $(G^{(n)})_{n\in\mathbb{N}}$ *are asymptotically approximable by trace polynomials.*

(4) $F^{(n)}$ *and* $G^{(n)}$ *witness the conditional Talagrand inequality with respect to Gaussian measure, that is,*

$$\|F^{(n)}(X^{(n)},Y^{(n)})-X^{(n)}\|_{L^2}^2=\|G^{(n)}(Z^{(n)},Y^{(n)})-Z^{(n)}\|_{L^2}^2\leq\frac{2}{n^2}|h_g^{(n)}(X^{(n)}|Y^{(n)})|.$$

(5) We have

$$\max\left(\|F^{(n)}-\pi_1\|_{\mathrm{Lip}},\|G^{(n)}-\pi_1\|_{\mathrm{Lip}}\right)\leq(\max(C,1/c)^3-1)\max(C,1/c)^{1/2}.$$

Remark 17.1.2. Of course, as explained in §12.3, the existence of transport holds in much greater generality in the classical setting. The difficulty of our result is to control the asymptotic behavior as $n\to\infty$. The explicit form of the heat semigroup with respect to the Gaussian measure makes this task much easier. To get asymptotic approximability by trace polynomials (see Proposition 17.5.1), we rely on Lemma 16.4.3, which in turn depends on our earlier results about conditional expectation.

Remark 17.1.3. The transport constructed here is usually *not* optimal. Indeed, Otto and Villani's heat semigroup method rarely produces the optimal transport maps, even in the non-conditional setting. The interpolation between the two measures is given by diffusion rather than the displacement interpolation from optimal transport theory.

Similar to §15, we will postpone the proof of the main theorem until after explaining some of its consequences. First, the following theorem is obtained from Theorem 17.1.1 in the large-n limit.

Theorem 17.1.4. *Continue with the same setup and notation from the previous theorem. Let* (X,Y) *be a tuple of non-commutative random variables given by the limiting free Gibbs law* λ *of* $V^{(n)}$*, and let* Z *be a freely independent free semicircular* d_1*-tuple. Let* $F^{(n)}\rightsquigarrow F$ *and* $G^{(n)}\rightsquigarrow G$*. Then the following hold.*

(1) $(F,\pi_2)\circ(G,\pi_2)=\mathrm{id}=(G,\pi_2)\circ(F,\pi_2)$.

(2) We have $(F(X,Y),Y)\sim(Z,Y)$ *and* $(G(Z,Y),Y)\sim(X,Y)$ *in non-commutative law.*

(3) There is a unique tracial W^**-isomorphism* $\alpha:\mathrm{W}^*(X,Y)\to\mathrm{W}^*(Z,Y)=\mathrm{W}^*(Z)*\mathrm{W}^*(Y)$ *such that* $\alpha(X)=G(Z,Y)$ *and* $\alpha(Y)=Y$.

(4) F *and* G *witness the conditional free Talagrand inequality with respect to the law of a free semicircular family, that is,*

$$\|F(X,Y)-X\|_2^2=\|G(Z,Y)-Z\|_2^2\leq 2|\chi_g^*(X|Y)|,$$

where

$$\chi_g^*(X|Y):=\chi*(X|Y)-\frac{1}{2}\|X\|_2^2-\frac{d}{2}\log 2\pi.$$

(5) We have

$$\max(\|F - \pi_1\|_{\mathrm{Lip}}, \|G - \pi_1\|_{\mathrm{Lip}}) \leq (\max(C, 1/c)^3 - 1)\max(C, 1/c)^{1/2},$$

and consequently

$$\|F(X,Y) - X\|_\infty = \|G(Z,Y) - Z\|_\infty \leq \Theta(\max(C,1/c)^3 - 1)\max(C,1/c) + \max_j |\tau(X_j)|,$$

where Θ is the universal constant from Lemma 11.5.2.

Remark 17.1.5. Claim (3) implies that the subalgebra $\mathrm{W}^*(Y)$ is *freely complemented* in $\mathrm{W}^*(X,Y)$, that is, the larger algebra is the free product of $\mathrm{W}^*(Y)$ with some other subalgebra.

Remark 17.1.6. Regarding (4), we remark that the free Talagrand inequality for self-adjoint tuples was studied in greater generality in [HU06] and [Dab10, §3.3]. Although we restricted ourselves to the case where the target measure is Gaussian/semicircular, our goal in this paper was not merely to estimate the Wasserstein distance using some coupling, but rather to exhibit a coupling that arises from a transport map (and in fact a Lipschitz transport map).

Remark 17.1.7. Regarding (5), the exact constants are probably not optimal, and the precise form of the estimate is not important for our purposes. However, in formulating and proving our results, we took care to ensure that the constants approach zero as $c, C \to 1$. Thus, if $V^{(n)}(x,y)$ is a small perturbation of the quadratic potential $(1/2)(\|x\|_2^2 + \|y\|_2^2)$, then the transport function $F(x,y)$ is close to x in Lipschitz norm, and $F(X,Y)$ is close to X in operator norm. This perturbative setting was studied first in the literature, for instance in [GM06] and [GS14], and we will discuss it further in §18.

Proof. (1) Because $(F^{(n)}, \pi_2)$ and $(G^{(n)}, \pi_2)$ are $\|\cdot\|_2$-Lipschitz, Proposition 13.5.12 implies that

$$(F^{(n)}, \pi_2) \circ (G^{(n)}, \pi_2) \rightsquigarrow (F, \pi_2) \circ (G, \pi_2),$$

and hence $(F, \pi_2) \circ (G, \pi_2) = \mathrm{id}$, and the same holds in the reverse direction.

(2) First, note that because (F, π_2) and (G, π_2) are $\|\cdot\|_2$-Lipschitz, $F(X,Y)$ and $G(X,Z)$ are tuples of bounded operators by Lemma 13.3.4 (3). Let u be a $\|\cdot\|_2$-Lipschitz scalar-valued function from the space $C_{\mathrm{tr,app}}(\mathbb{R}^{d_1+d_2}, \|\cdot\|_2)$. Then $u \circ (F^{(n)}, \pi_2) \rightsquigarrow u \circ (F, \pi_2)$. Hence, by Theorem 15.1.5,

$$u(F(X,Y),Y) = \lim_{n\to\infty} Eu(F^{(n)}(X^{(n)}, Y^{(n)}), Y^{(n)}) = \lim_{n\to\infty} Eu(Z^{(n)}, Y^{(n)}) = u(Z,Y).$$

The last equality follows because the potential for $(Z^{(n)}, Y^{(n)})$ satisfies Assumption 15.1.1; indeed, the potential for $Y^{(n)}$ satisfies this assumption by Proposition 15.1.8, and the potential for $(Z^{(n)}, Y^{(n)})$ is the sum of this potential in y and the quadratic potential in z. And the non-commutative law of $(Z^{(n)}, Y^{(n)})$ converges in probability to that of (Z, Y) by the same reasoning as in Lemma 16.4.3 (4).

Therefore, $u(F(X,Y),Y) = u(Z,Y)$ for globally $\|\cdot\|_2$-Lipschitz functions u in $C_{\text{tr,app}}(\mathbb{R}^{*(d_1+d_2)}, \|\cdot\|_2)$. In particular, it holds for a trace polynomial composed with a smooth cut-off function, so that $\lambda_{(F(X,Y),Y)} = \lambda_{(Z,Y)}$. The argument for $(G(X,Z),Z)$ is symmetrical.

(3) Since $(G(Z,Y),Y) \sim (X,Y)$, it follows from Lemma 10.2.5, that there is an isomorphism $\alpha : \mathrm{W}^*(X,Y) \to \mathrm{W}^*(G(Z,Y),Y)$ sending X to $G(Z,Y)$ and Y to Y. But $Z = F(G(Z,Y),Y)$ since $(F,\pi_2) \circ (G,\pi_2) = \mathrm{id}$ (using Proposition 13.4.3), hence $Z \in \mathrm{W}^*(G(Z,Y),Y)$, so $\mathrm{W}^*(G(Z,Y),Y) = \mathrm{W}^*(Z,Y)$. Since Z and Y are freely independent, $\mathrm{W}^*(Z,Y)$ is isomorphic to the tracial W*-free product $\mathrm{W}^*(Z) * \mathrm{W}^*(Y)$. Finally, it is clear that the isomorphism is unique once the values on the generators are specified.

(4) This follows from taking the large-n limit of the relation (4) from Theorem 17.1.1. Indeed, by Theorem 15.1.5, we have

$$\lim_{n\to\infty} E\left\|F^{(n)}(X^{(n)}, Y^{(n)}) - X^{(n)}\right\|_2^2 = \|F(X,Y) - X\|_2^2,$$

and the analogous statement holds for $\|G(Z,X) - Z\|_2$. Similarly, by Theorem 16.4.1

$$\begin{aligned}\frac{1}{n^2} h_g^{(n)}(X^{(n)}|Y^{(n)}) &= h(X^{(n)}|Y^{(n)}) + d_1 \log n - \frac{1}{2} E\left\|X^{(n)}\right\|_2^2 - \frac{d_1}{2}\log 2\pi \\ &\to \chi * (X|Y) - \frac{1}{2}\|X\|_2^2 - \frac{d_1}{2}\log 2\pi \\ &= \chi_g(X|Y).\end{aligned}$$

(5) The estimate on the Lipschitz norms follows by taking the large-n limit of claim (5) from Theorem 17.1.1. For the operator norm bound, note that the first two terms of the equation are equivalent that in light of the isomorphism α. Then using Lemma 11.5.2, we have that

$$\limsup_{n\to\infty} \left\| F_j(X^{(n)}, Y^{(n)}) - X_j^{(n)} \tau_n(F_j(X^{(n)}, Y^{(n)})) + \tau_n(X_j^{(n)}) \right\|_\infty \le c^{-1/2}\Theta \|F - \pi_1\|_{\mathrm{Lip}}$$

in probability. Then because of Lemma 13.3.6 (2),

$$\|F_j(X,Y) - X_j + \tau(X_j)\|_\infty \le c^{-1/2}\Theta\|F - \pi_1\|_{\mathrm{Lip}}.$$

since $\tau(F_j(X,Y)) = \tau(Z_j) = 0$. Then we plug in our estimate for $\|F - \pi_1\|_{\mathrm{Lip}}$ and simplify using $c^{-1/2} \le \max(C, 1/c)^{1/2}$. □

The next two theorems concerning "triangular transport" are obtained by iterating Theorems 17.1.1 and 17.1.4. In the classical setting, a triangular transport between two probability measures μ and ν on $\mathbb{R}^d$ is a function f such that $f_*\mu = \nu$ and f_k only depends on x_1, $\dots x_k$ for each k, or in other words

$$f(x_1, \dots, x_d) = (f_1(x_1), f_2(x_1, x_2), \dots, f_d(x_1, \dots, x_d)),$$

similar to a lower-triangular linear transform. Triangular transport was studied in [BKM05], and in fact, the authors proved the existence of optimal triangular transport (see especially Corollary 3.10).

Triangular transport can be obtained by iterating conditional transport. Roughly speaking, given random variables $(X_1, \dots, X_d) \sim \mu$ and $(Y_1, \dots, Y_d) \sim \nu$, we first transport X_1 to Y_1, then transport X_2 to Y_2 conditioned on X_1, then transport X_3 to Y_3 conditioned on X_1 and X_2, and so on. Under some conditions, the Talagrand inequality can even be witnessed by triangular transport as also noted in [BKM05]. For instance, if the conditional distribution of Y_k given Y_1, ..., Y_{k-1} is sufficiently regular and satisfies the conditional log-Sobolev inequality for each k, then using the ideas sketched in §12.4, we can construct a conditional transport of X_k to Y_k such that $\|F_k(X) - X_k\|_{L^2}^2$ is bounded by the entropy of X_k relative to the law of Y_k conditioned on X_1, ..., X_{k-1}. Then by additivity of entropy under conditioning,

$$\|F(X) - X\|_{L^2}^2 = \sum_{j=1}^d \|F_j(X) - X_j\|_{L^2}^2 \leq 2|h(\mu\|\nu)|.$$

We will carry out these ideas in detail in the random matrix setting where $\mu^{(n)}$ is a convex Gibbs law and the target measure is $\sigma_1^{(n)}$, and thus obtain the following result.

Theorem 17.1.8. *Let $V^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^d \to \mathbb{R}$ satisfy Assumption 15.1.1 for some $0 < c \leq C$, and let $\mu^{(n)}$ be the corresponding measure and $X^{(n)}$ the corresponding random variable. Let $Z^{(n)}$ be a Gaussian d-tuple. There exist functions $F^{(n)}, G^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}}^d \to M_n(\mathbb{C})_{\mathrm{sa}}^d$ such that the following hold.*

(1) $F^{(n)}$ and $G^{(n)}$ are inverses of each other.

(2) $F^{(n)}(x_1, \dots, x_d) = (F_1^{(n)}(x_1), F_2^{(n)}(x_1, x_2), \dots, F_d^{(n)}(x_1, \dots, x_d))$ for some functions $F_1^{(n)}$, ..., $F_d^{(n)}$, and a similar expression holds for $G^{(n)}$.

(3) $F^{(n)}(X^{(n)}) \sim Z^{(n)}$ and $G^{(n)}(Z^{(n)}) \sim X^{(n)}$.

(4) $(F^{(n)})_{n \in \mathbb{N}}$ and $(G^{(n)})_{n \in \mathbb{N}}$ are asymptotically approximable by trace polynomials.

(5) $F^{(n)}$ and $G^{(n)}$ witness the Talagrand inequality relative to the Gaussian measure, that is,

$$\left\|F^{(n)}(X^{(n)}) - X^{(n)}\right\|_{L^2}^2 = \left\|G^{(n)}(Z^{(n)}) - Z^{(n)}\right\|_{L^2}^2 \leq \frac{2}{n^2}|h_g^{(n)}(X^{(n)})|.$$

(6) We have

$$\|F^{(n)} - \mathrm{id}\|_{\mathrm{Lip}} \leq d^{1/2}(\max(C, 1/c)^3 - 1)\max(C, 1/c)^{1/2},$$

and $\|G^{(n)} - \mathrm{id}\|_{\mathrm{Lip}}$ is bounded by some constant depending only on $\max(C, 1/c)$ and d, which goes to zero as $\max(C, 1/c) \to 1$.

Proof. For each $k = 1, \ldots, d$, let $V_k^{(n)}$ be the potential corresponding to the marginal law of $(X_1^{(n)}, \ldots, X_k^{(n)})$, given by the formula in Proposition 15.1.8. By that proposition, $(V_k^{(n)})_{n\in\mathbb{N}}$ satisfies Assumption 15.1.1 with the same constants c and C. Thus, we can Theorem 17.1.1 for $X_k^{(n)}$ conditioned on $X_1^{(n)}$, $\ldots$, $X_{k-1}^{(n)}$ to find a function $F_k^{(n)}(x_1, \ldots, x_k)$ such that

$$\left(X_1^{(n)}, \ldots, X_{k-1}^{(n)}, F_k^{(n)}\left(X_1^{(n)}, \ldots, X_{k-1}^{(n)}\right)\right) \sim \left(X_1^{(n)}, \ldots, X_{k-1}^{(n)}, Z_k^{(n)}\right).$$

Then set

$$F^{(n)}(x) := \left(F_1^{(n)}(x_1), F_2^{(n)}(x_1, x_2), \ldots, F_d^{(n)}(x_1, \ldots, x_d)\right).$$

The inverse function $G^{(n)}$ is slightly more complicated to express. Let $G_k^{(n)}$ be the function obtained from Theorem 17.1.1 to transport $Z_k^{(n)}$ to $X_k^{(n)}$ conditioned on $X_1^{(n)}$, $\ldots$, $X_{k-1}^{(n)}$. Then define

$$G^{(n)}(x_1, \ldots, x_d) = \left(H_1^{(n)}(x_1), H_2^{(n)}(x_1, x_2), \ldots, H_d^{(n)}(x_1, \ldots, x_d)\right),$$

where the $H_k^{(n)}$'s are defined by induction on k to satisfy

$$H_k^{(n)}(x_1, \ldots, x_k) = G_k^{(n)}\left(H_1^{(n)}(x_1), \ldots, H_{k-1}^{(n)}(x_1, \ldots, x_{k-1}), x_k\right).$$

Let us check that the claims of the theorem hold.

(1) By construction and a bit of computation, $F^{(n)}$ and $G^{(n)}$ are inverses of each other.

(2) holds by construction.

(3) Denote $Y_k^{(n)} := F_k^{(n)}(X_1^{(n)}, \ldots, X_k^{(n)})$ and write $Z^{(n)} = Z_1^{(n)}, \ldots, Z_d^{(n)}$. We will check by backwards induction on k that

$$\left(X_1^{(n)}, \ldots, X_k^{(n)}, Y_{k+1}^{(n)}, \ldots Y_d^{(n)}\right) \sim \left(X_1^{(n)}, \ldots, X_k^{(n)}, Z_{k+1}^{(n)}, \ldots, Z_d^{(n)}\right).$$

The base case $k = d$ is trivial. For the induction step, suppose the claim holds for k. Since $Y_k^{(n)}$ is a function $X_1^{(n)}$, $\ldots$, $X_k^{(n)}$, the claim for k implies that

$$\left(X_1^{(n)}, \ldots, X_{k-1}^{(n)}, Y_k^{(n)}, Y_{k+1}^{(n)}, \ldots Y_d^{(n)}\right) \sim \left(X_1^{(n)}, \ldots, X_{k-1}^{(n)}, Y_k^{(n)}, Z_{k+1}^{(n)}, \ldots Z_d^{(n)}\right).$$

But then note that $Z_{k+1}^{(n)}$, $\ldots$, $Z_d^{(n)}$ are independent of the $X_j^{(n)}$'s and the $Y_j^{(n)}$'s and $Z_k^{(n)}$. So using our choice of $F_k^{(n)}$, we have

$$\left(X_1^{(n)}, \ldots, X_{k-1}^{(n)}, Y_k^{(n)}, Z_{k+1}^{(n)}, \ldots Z_d^{(n)}\right) \sim \left(X_1^{(n)}, \ldots, X_{k-1}^{(n)}, Z_k^{(n)}, Z_{k+1}^{(n)}, \ldots Z_d^{(n)}\right).$$

This completes the induction and shows that $F^{(n)}(X^{(n)}) \sim Z^{(n)}$ as desired. Since $G^{(n)}$ is the inverse of $F^{(n)}$, this also implies that $G^{(n)}(Z^{(n)}) \sim X^{(n)}$.

(4) Theorem 17.1.1 guarantees that $F_k^{(n)}$ and $G_k^{(n)}$ are asymptotically approximable by trace polynomials and uniformly $\|\cdot\|_2$-Lipschitz. Since $F^{(n)}$ and $G^{(n)}$ are obtained from

these functions by iterated composition, they are also asymptotically approximable by trace polynomials and uniformly $\|\cdot\|_2$-Lipschitz.

(5) Note that the second term in (5) is equivalent to the first because $F^{(n)}(Z^{(n)}) \sim X^{(n)}$ and $F^{(n)}$ is the inverse of $G^{(n)}$.

To prove the inequality, recall that Theorem 17.1.1 guarantees that

$$\left\| F_k^{(n)}(X_1^{(n)}, \dots, X_k^{(n)}) - X_k^{(n)} \right\|_{L^2}^2 \leq \frac{2}{n^2} |h_g^{(n)}(X_k^{(n)} | X_1^{(n)}, \dots, X_k^{(n)})|.$$

Since $F_k^{(n)}$ defines the k-th coordinate function of $F^{(n)}$, we have

$$\left\| F^{(n)}(X^{(n)}) - X^{(n)} \right\|_{L^2}^2 = \sum_{k=1}^{d} \left\| F_k^{(n)}(X_1^{(n)}, \dots, X_k^{(n)}) - X_k^{(n)} \right\|_{L^2}^2.$$

But by additivity of entropy under conditioning,

$$h_g^{(n)}(X_1^{(n)}, \dots, X_d^{(n)}) = \sum_{k=1}^{d} h_g^{(n)}(X_k^{(n)} | X_1^{(n)}, \dots, X_k^{(n)}),$$

and hence (5) holds.

(6) Let $\pi_k(x_1, \dots, x_d) = x_k$. Then by decomposing $F^{(n)}$ coordinate-wise and applying Theorem 17.1.1 (5),

$$\left\| F^{(n)} - \mathrm{id} \right\|_{\mathrm{Lip}}^2 \leq \sum_{k=1}^{d} \left\| F_k^{(n)} - \pi_k \right\|_{\mathrm{Lip}}^2 \leq d(\max(C, 1/c)^3 - 1)^2 \max(C, 1/c),$$

and thus we have the desired estimate for $\|F^{(n)} - \mathrm{id}\|_{\mathrm{Lip}}$. The construction of $G^{(n)}$ is more complicated than that of $F^{(n)}$, but nonetheless since $\|G_k^{(n)} - \pi_k\|_{\mathrm{Lip}}$ is bounded by a constant that goes to zero as $\max(C, 1/c) \to 1$, one inductively obtains similar bounds for $\|H_k^{(n)} - \pi_k\|_{\mathrm{Lip}}$ and can thus estimate $\|G^{(n)} - \mathrm{id}\|_{\mathrm{Lip}}$. ☐

The final result is the W*-algebraic version of Theorem 17.1.8 which we obtain in the large-n limit. The arguments to deduce Theorem 17.1.9 from Theorem 17.1.8 are exactly the same as those used to deduce Theorem 17.1.4 from Theorem 17.1.1, so we leave the details of the proof to the reader.

Theorem 17.1.9. *Continue the notation of Theorem 17.1.8. Let X be a d-tuple of non-commutative random variables realizing the free Gibbs law which is the large-n limit of $\mu^{(n)}$, and let Z be a free semicircular d-tuple. Let $F, G \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\mathrm{sa}}^d$ be the functions such that $F^{(n)} \rightsquigarrow F$ and $G^{(n)} \rightsquigarrow G$. Then the following hold.*

(1) F and G are inverses of each other.

(2) $F(x_1, \dots, x_d) = (F_1(x_1), F_2(x_1, x_2), \dots, F_d(x_1, \dots, x_d))$*, where* $F_j \in C_{\mathrm{tr,app}}(\mathbb{R}^{*j}, \|\cdot\|_2)_{\mathrm{sa}}$*, and a similar expression holds for* G*.*

(3) $F(X) \sim Z$ *and* $G(Z) \sim X$ *in non-commutative law.*

(4) There is a unique isomorphism $\alpha : \mathrm{W}^*(X) \to \mathrm{W}^*(Z)$ *such that* $\alpha(F(X)) = Z$*. Moreover, for each* $k = 1, \dots, d$*, we have*

$$\phi(\mathrm{W}^*(X_1, \dots, X_k)) = \mathrm{W}^*(Z_1, \dots, Z_k).$$

(5) F *and* G *witness the free Talagrand inequality relative to the law of a free semicircular family, that is,*

$$\|F(X) - X\|_2^2 = \|G(Z) - Z\|_2^2 \leq 2|\chi_g^*(X)| = |\chi_g(X)|.$$

(6) We have

$$\|F - \mathrm{id}\|_{\mathrm{Lip}} \leq d^{1/2}(\max(C, 1/c)^3 - 1)\max(C, 1/c)^{1/2},$$

and $\|G - \mathrm{id}\|_{\mathrm{Lip}}$ *is bounded by some constant depending only on* $\max(C, 1/c)$ *and* d*, which goes to zero as* $\max(C, 1/c) \to 1$*.*

(7) We have

$$\|F(X) - X\|_\infty = \|G(Z) - Z\|_\infty \leq \Theta(\max(C, 1/c)^3 - 1)\max(C, 1/c) + \max_j |\tau(X_j)|,$$

where Θ *is the universal constant from Lemma 11.5.2.*

Remark 17.1.10. (4) of Theorem 17.1.9 can be stated as saying that the two sequences of W*-algebra inclusions

$$\mathrm{W}^*(X_1) \subseteq \mathrm{W}^*(X_1, X_2) \subseteq \dots \subseteq \mathrm{W}^*(X_1, \dots, X_d)$$

and

$$\mathrm{W}^*(Z_1) \subseteq \mathrm{W}^*(Z_1, Z_2) \subseteq \dots \subseteq \mathrm{W}^*(Z_1, \dots, Z_d).$$

are isomorphic. This second sequence of inclusions is of course isomorphic to

$$L(\mathbb{F}_1) \subseteq L(\mathbb{F}_2) \subseteq \dots \subseteq L(\mathbb{F}_d),$$

where $\mathbb{F}_j$ is the free group on j generators and the inclusion $\mathbf{F}_j \to \mathbb{F}_{j+1}$ is the standard one which sends the first j generators to the first j generators. In particular, all von Neumann algebraic properties of the sequence of inclusions coming from $(X_1, \dots, X_d)$ are the same as the sequence of inclusions for the free group von Neumann algebras.

17.2 Strategy

We will employ Lemma 12.3.1 as in Otto and Villani's argument, while also using the explicit form that the transport takes when the background measure is Gaussian or Lebesgue (see §12.5). Our end goal is to study the transport maps with Gaussian as the target measure, but we will perform part of the argument using the transport functions for Lebesgue measure, which are computationally more tractable and are related to the Gaussian transport functions by a change of variables.

Consider a sequence of potentials $V^{(n)} : M_n(\mathbb{C})_{\text{sa}}^{d_1+d_2} \to \mathbb{R}$ satisfying Assumption 15.1.1, and let $\mu^{(N)}$ be the corresponding probability measure and $(X^{(n)}, Y^{(n)})$ the associated random variable. Let $Z_t^{(n)}$ an independent d_1-tuple with the Gaussian distribution $\sigma_t^{(n)}$, let $X_t^{(n)} = X^{(n)} + Z_t^{(n)}$, and let $\mu_t^{(n)}$ be the law of $(X_t^{(n)}, Y^{(n)})$. Recall that the density $\rho_t^{(n)}$ of $\mu_t^{(n)}$ evolves according the normalized flat heat equation conditioned on $Y^{(n)}$, that is,

$$\partial_t \rho_t^{(n)}(x,y) = \frac{1}{2n^2} \Delta_x \rho_t^{(n)}(x,y).$$

(Of course, $\rho_t^{(n)}$ can be obtained from the standard flat heat evolution with respect to Δ_x rather than $(1/2n^2)\Delta_x$ by slowing down time by a factor of $2n^2$.) Let $V_t^{(n)}$ be the potential given by

$$V_t^{(n)}(x,y) = -\frac{1}{n^2} \log \int e^{-n^2 V^{(n)}(x-z,y)} \, d\sigma_t^{(n)}(z),$$

so that $V_t^{(n)}(x,y)$ equals $-\frac{1}{n^2} \log \rho_t^{(n)}(x,y)$ up to an additive constant and $V_0^{(n)} = V^{(n)}$.

Define a function $F_{s,t}^{(n)} : M_n(\mathbb{C})_{\text{sa}}^{d_1+d_2} \to M_n(\mathbb{C})_{\text{sa}}^{d_1}$ by solving the ODE

$$\partial_s F_{s,t}^{(n)}(x,y) = \frac{1}{2} \nabla_x V_s^{(n)}(F_{s,t}^{(n)}(x,y), y). \tag{17.1}$$

By Lemma 12.3.1 with Lebesgue measure as the background measure, $F_{s,t}(\cdot, y)$ pushes forward the conditional distribution of $X_t^{(n)}$ given $Y^{(n)} = y$ to the conditional distribution of $X_s^{(n)}$ given $Y^{(n)} = y$ for any $s, t > 0$. Thus, the function $(F_{s,t}^{(n)}, \pi_2)$ given by $(x,y) \mapsto (F_{s,t}^{(n)}(x,y), y)$ satisfies $(F_{s,t}^{(n)}, \pi_2)_* \mu_t^{(n)} = \mu_s^{(n)}$ per the discussion on conditional transport in §12.4. Note also that for $s, t, u \in [0, \infty)$, we have

$$F_{s,t}^{(n)}(F_{t,u}^{(n)}(x,y), y) = F_{s,u}^{(n)}(x,y),$$

which follows from the existence and uniqueness theory of ODE.

Remark 17.2.1. To verify that the hypotheses of Lemma 12.3.1 are satisfied for $\rho_t^{(n)}$, note that $\rho_t^{(n)}$ is C^∞ in (x,t) for $t > 0$ by the standard theory of the heat equation. Moreover, Lemma 16.4.3 implies that $c/(1+ct) \leq H_x V_t^{(n)} \leq C/(1+Ct)$ and hence $\nabla_x V_t^{(n)}$ is $C/(1+Ct)$-Lipschitz by Lemma 11.1.4. The claim $(F_{s,t}^{(n)}, \pi_2)_* \mu_t^{(n)} = \mu_s^{(n)}$ for s or t equal to zero will be justified later by a limiting argument.

Using the change of variables explained in §12.5, we can obtain the transport functions associated the heat equation with respect to Gaussian measure rather than Lebesgue measure. Let $\tilde{\mu}_t^{(n)}$ be the law of $(e^{-t/2}X_{e^t-1}^{(n)}, Y^{(n)})$, and let

$$\tilde{V}_t^{(n)}(x,y) := V_{e^t-1}^{(n)}(e^{t/2}x, y),$$

which is the potential corresponding to $\tilde{\mu}_t^{(n)}$. The density of $\tilde{\mu}_t^{(n)}$ with respect to the Gaussian measure is then

$$\tilde{\rho}_t(x,y) = e^{-n^2(\tilde{V}_t^{(n)}(x,y)-(1/2)\|x\|_2^2-(1/2)\|y\|^2)},$$

which evolves according to the normalized heat equation for Gaussian measure with respect to x conditioned on y:

$$\partial_t \tilde{\rho}_t(x,y) = \frac{1}{2n^2}\Delta_x \tilde{\rho}_t(x,y) - \frac{1}{2}\langle x, \nabla \tilde{\rho}_t(x)\rangle_2.$$

Then let

$$\tilde{F}_{s,t}^{(n)}(x,y) = e^{-s/2}\tilde{F}_{e^s-1,e^t-1}^{(n)}(e^{t/2}x,y),$$

so that $(\tilde{F}_{s,t}^{(n)}, \pi_2)_* \mu_s^{(n)} = \mu_t^{(n)}$ for $s,t > 0$. Moreover, $\tilde{F}_{s,t}^{(n)}$ satisfies

$$\partial_s \tilde{F}_{s,t}^{(n)}(x,y) = -\frac{1}{2}(\nabla_x \tilde{V}_t^{(n)} - \pi_1)\left(\tilde{F}_{s,t}^{(n)}(x,y)\right) = -\frac{1}{2n^2}(\nabla_x \log \tilde{\rho}_t)\left(\tilde{F}_{s,t}^{(n)}(x,y)\right). \tag{17.2}$$

We also have by $s,t,u \in [0,\infty)$ that

$$\tilde{F}_{s,t}^{(n)}(\tilde{F}_{t,u}^{(n)}(x,y),y) = \tilde{F}_{s,u}^{(n)}.$$

The details of these computations were explained in §12.5, except without the normalizing factor of $1/n^2$ in the Laplacian, and we leave the necessary modifications as an exercise.

To prove Theorem 17.1.1, we will show that $\tilde{F}_{s,t}^{(n)}$ has a limit as s or $t \to \infty$. Since $\tilde{\mu}_t$ is the law of $(\tilde{X}_t^{(n)}, Y^{(n)}) := (e^{-t/2}X^{(n)} + e^{-t/2}Z_{e^t-1}^{(n)}, Y^{(n)}) \sim (e^{-t/2}X^{(n)} + (1-e^{-t})^{1/2}Z_1^{(n)}, Y^{(n)})$, it is natural to denote by $\tilde{\mu}_\infty$ the law of $(Z_1^{(n)}, Y^{(n)})$. Then we will extend the equality $(\tilde{F}_{s,t}^{(n)})_* \mu_t^{(n)} = \mu_s^{(n)}$ to s and t in $[0,\infty]$. Then the functions $F^{(n)}$ and $G^{(n)}$ in the theorem will be given by $F^{(n)} = F_{\infty,0}^{(n)}$ and $G^{(n)} = F_{0,\infty}^{(n)}$.

To carry out this construction, we require finer control over the functions $\tilde{F}_{s,t}^{(n)}$ than in Otto and Villani's proof of the Talagrand inequality (Theorem 12.3.5). In fact, we will give explicit estimates for the Lipschitz norms of $\nabla_x V_t^{(n)}(x,y)$ and $F_{s,t}^{(n)}(x,y)$, which then translate into estimates for $\tilde{F}_{s,t}^{(n)}$. From there, we will prove explicit and dimension-independent estimates for the rates of convergence of $F_{s,t}^{(n)}$ as s or t goes to ∞ (rather than using the softer L^2 estimates in our proof sketch of Theorem 12.3.5).

After controlling the behavior of $\tilde{F}_{s,t}^{(n)}$ as s or t tends to ∞, we turn our attention to the behavior as $n \to \infty$. To show that $F_{s,t}^{(n)}$ (and hence $\tilde{F}_{s,t}^{(n)}$) is asymptotically approximable by trace polynomials, we use Proposition 14.3.7 together with the fact that $V_t^{(n)}$ is asymptotically approximable by trace polynomials by Lemma 16.4.3.

17.3 Lipschitz estimates for conditional transport

As the first step in our proof of Theorem 17.1.1, this section proves the technical estimate Lemma 17.3.2 on the Lipschitz seminorm of $F_{s,t}^{(n)}$. Our Lipschitz estimates for $F_{s,t}^{(n)}$ depend in turn upon the following technical estimates for $V_t^{(n)}(x,y)$ that come from the convexity properties of $V_t^{(n)}$ in Lemma 16.4.3 together with Lemma 11.1.5.

Lemma 17.3.1. *We have*

$$\left\langle \nabla_x V_t^{(n)}(x,y) - \nabla_x V_t^{(n)}(x',y'), x-x' \right\rangle_2 \leq \frac{C}{1+Ct}\|x-x'\|_2^2 + \frac{C-c}{(1+Ct)^{1/2}(1+ct)^{1/2}}\|x-x'\|_2\|y-y'\|_2$$

and

$$\left\langle \nabla_x V_t^{(n)}(x,y) - \nabla_x V_t^{(n)}(x',y'), x-x' \right\rangle_2 \geq \frac{c}{1+ct}\|x-x'\|_2^2 - \frac{C-c}{(1+Ct)^{1/2}(1+ct)^{1/2}}\|x-x'\|_2\|y-y'\|_2. \quad (17.3)$$

Proof. First, note that

$$\begin{aligned}\left\langle \nabla_x V_t^{(n)}(x,y) - \nabla_x V_t^{(n)}(x',y'), x-x' \right\rangle_2 \\ = \left\langle \nabla_x V_t^{(n)}(x,y) - \nabla_x V_t^{(n)}(x',y), x-x' \right\rangle_2 + \left\langle \nabla_x V_t^{(n)}(x',y) - \nabla_x V_t^{(n)}(x',y'), x-x' \right\rangle_2\end{aligned} \quad (17.4)$$

By Lemma 11.1.4, the first term on the right-hand side of (17.4) can be estimated by

$$\frac{c}{1+ct}\|x-x'\|_2^2 \leq \langle \nabla_x V_t^{(n)}(x,y) - \nabla_x V_t^{(n)}(x',y), x-x'\rangle_2 \leq \frac{C}{1+Ct}\|x-x'\|_2^2.$$

To handle the second term on the right-hand side of (17.4), define

$$\begin{aligned}\overline{V}_t^{(n)}(x,y) &= V_t^{(n)}(x,y) - \frac{c}{2(1+ct)}\|x\|_2^2 - \frac{c}{2}\|y\|_2^2 \\ \underline{V}_t^{(n)}(x,y) &= V_t^{(n)}(x,y) - \frac{C}{2(1+Ct)}\|x\|_2^2 - \frac{C}{2}\|y\|_2^2.\end{aligned}$$

By Lemma 16.4.3, $\overline{V}_t^{(n)}$ is convex and $\underline{V}_t^{(n)}$ is concave, and in particular

$$\begin{aligned}0 \leq H\overline{V}_t^{(n)} &\leq \left(\frac{C}{1+Ct} - \frac{c}{1+ct}\right) I_{n^2 d_1} \oplus (C-c) I_{n^2 d_2} \\ &= \frac{C-c}{(1+Ct)(1+ct)} I_{n^2 d_1} \oplus (C-c) I_{n^2 d_2}.\end{aligned}$$

Note that

$$\begin{aligned}\nabla_x V_t^{(n)}(x',y) - \nabla_x V_t^{(n)}(x',y') &= \left(\nabla_x V_t^{(n)}(x',y) - \frac{c}{1+ct}x'\right) - \left(\nabla_x V_t^{(n)}(x',y) - \frac{c}{1+ct}x'\right) \\ &= \nabla_x \overline{V}_t^{(n)}(x',y) - \nabla_x \overline{V}_t^{(n)}(x',y').\end{aligned}$$

Therefore,

$$\begin{aligned}\left\langle \nabla_x V_t^{(n)}(x',y) - \nabla_x V_t^{(n)}(x',y'), x - x' \right\rangle_2 &= \left\langle \nabla_x \overline{V}_t^{(n)}(x',y) - \nabla_x \overline{V}_t^{(n)}(x',y'), x - x' \right\rangle_2 \\ &= \left\langle \nabla \overline{V}_t^{(n)}(x',y) - \nabla \overline{V}_t^{(n)}(x',y'), (x - x', 0) \right\rangle_2.\end{aligned}$$

Now we apply Lemma 11.1.5 to $\overline{V}_t$ with the matrix $A = \frac{C-c}{(1+Ct)(1+ct)} I_m \oplus (C-c) I_n$ and conclude that

$$\begin{aligned}&\left| \left\langle \nabla \overline{V}_t^{(n)}(x',y) - \nabla \overline{V}_t^{(n)}(x',y'), (x - x', 0) \right\rangle_2 \right| \\ &\leq \left((C-c)\|y-y'\|_2^2 \right)^{1/2} \left(\frac{C-c}{(1+Ct)(1+ct)} \|x-x'\|_2^2 \right)^{1/2} \\ &\leq \frac{C-c}{(1+Ct)^{1/2}(1+ct)^{1/2}} \|x-x'\| \|y-y'\|.\end{aligned}$$

Combining this estimate for the second term of (17.4) with our earlier estimate for the first term completes the proof. □

Lemma 17.3.2. *We have*

$$\left\| F_{s,t}^{(n)} \right\|_{\mathrm{Lip},dx} \leq \begin{cases} \frac{(1+Cs)^{1/2}}{(1+Ct)^{1/2}}, & s \geq t \\ \frac{(1+cs)^{1/2}}{(1+ct)^{1/2}} & s \leq t. \end{cases} \tag{17.5}$$

and

$$\left\| F_{s,t}^{(n)} \right\|_{\mathrm{Lip},dy} \leq \begin{cases} (C/c-1)(1+Cs)^{1/2} \left(\frac{1}{(1+Ct)^{1/2}} - \frac{1}{(1+Cs)^{1/2}} \right), & s \geq t, \\ (C/c-1)(1+cs)^{1/2} \left(\frac{1}{(1+Cs)^{1/2}} - \frac{1}{(1+Ct)^{1/2}} \right) & s \leq t. \end{cases} \tag{17.6}$$

Proof. Fix $t \geq 0$ and (x,y) and (x',y') in $M_n(\mathbb{C})_{\mathrm{sa}}^{d_1+d_2}$ and define

$$\phi(s) = \|F_{s,t}(x,y) - F_{s,t}(x',y')\|_2.$$

Note that ϕ is locally Lipschitz, hence absolutely continuous. Also,

$$\begin{aligned}
2\phi(s)\phi'(s) &= \partial_s[\phi(s)^2] \\
&= 2\Big\langle \partial_s F_{s,t}^{(n)}(x,y) - \partial_s F_{s,t}^{(n)}(x',y'), F_{s,t}^{(n)}(x,y) - F_{s,t}^{(n)}(x',y') \Big\rangle_2 \\
&= \Big\langle \nabla_x V_s^{(n)}\left(F_{s,t}^{(n)}(x,y), y\right) - \nabla_x V_s^{(n)}\left(F_{s,t}^{(n)}(x',y'), y'\right), F_{s,t}^{(n)}(x,y) - F_{s,t}^{(n)}(x',y') \Big\rangle_2 \\
&\leq \frac{C}{1+Ct}\left\| F_{s,t}^{(n)}(x,y) - F_{s,t}^{(n)}(x',y') \right\|_2^2 \\
&\quad + \frac{C-c}{(1+Ct)^{1/2}(1+ct)^{1/2}} \left\| F_{s,t}^{(n)}(x,y) - F_{s,t}^{(n)}(x',y') \right\|_2 \|y-y'\|_2 \\
&= \frac{C}{1+Cs}\phi(s)^2 + \frac{C-c}{(1+Cs)^{1/2}(1+cs)^{1/2}}\phi(s)\|y-y'\|_2,
\end{aligned}$$

where in the last step we have applied Lemma 17.3.1. It follows that whenever $\phi(s) > 0$,

$$\phi'(s) \leq \frac{C}{2(1+Cs)}\phi(s) + \frac{C-c}{2(1+Cs)^{1/2}(1+cs)^{1/2}}\|y-y'\|_2.$$

On the other hand, since $\phi(s) \geq 0$, any point where ϕ is zero and ϕ is differentiable must be a critical point, so when $\phi(s) = 0$ the estimate is vacuously true. This inequality implies

$$\begin{aligned}
\frac{d}{ds}\left[\frac{1}{(1+Cs)^{1/2}}\phi(s)\right] &\leq \frac{C-c}{2(1+Cs)(1+cs)^{1/2}}\|y-y'\|_2 \\
&\leq \frac{C(C-c)}{2c(1+Cs)^{3/2}}\|y-y'\|_2,
\end{aligned}$$

where in the last line we have observed that $(1+cs)^{1/2} \geq (c/C)^{1/2}(1+Cs)^{1/2} \geq (c/C)(1+Cs)^{1/2}$. Hence for $s \geq t$

$$\frac{1}{(1+Cs)^{1/2}}\phi(s) - \frac{1}{(1+Ct)^{1/2}}\phi(t) \leq \frac{C-c}{c}\left(\frac{1}{(1+Ct)^{1/2}} - \frac{1}{(1+Cs)^{1/2}}\right)\|y-y'\|_2.$$

Now we substitute $\phi(s) = \left\| F_{s,t}^{(n)}(x,y) - F_{s,t}^{(n)}(x',y') \right\|_2$ and $\phi(t) = \|x-x'\|_2$ and rearrange to obtain

$$\begin{aligned}
&\frac{1}{(1+Cs)^{1/2}}\left\| F_{s,t}^{(n)}(x,y) - F_{s,t}^{(n)}(x',y') \right\|_2 \\
&\qquad \leq \frac{1}{(1+Ct)^{1/2}}\|x-x'\|_2 + \frac{C-c}{c}\left(\frac{1}{(1+Ct)^{1/2}} - \frac{1}{(1+Cs)^{1/2}}\right)\|y-y'\|_2.
\end{aligned}$$

This proves the asserted estimates in the case where $s \geq t$. The argument for the case $s \leq t$ is similar. Here we use the lower bound rather than the upper bound in Lemma 17.3.1 and get

$$\phi'(s) \geq \frac{c}{2(1+cs)}\phi(s) - \frac{C-c}{2(1+Cs)^{1/2}(1+cs)^{1/2}}\|y-y'\|_2$$

so that

$$\frac{d}{ds}\left[\frac{1}{(1+cs)^{1/2}}\phi(s)\right] \geq -\frac{C-c}{2(1+Cs)^{1/2}(1+cs)}\|y-y'\|_2$$
$$\geq -\frac{C(C-c)}{2c(1+Cs)^{3/2}}\|y-y'\|_2.$$

Now we take $s \leq t$ and obtain

$$\frac{1}{(1+ct)^{1/2}}\|x-x'\|_2 - \frac{1}{(1+cs)^{1/2}}\left\|F_{s,t}^{(n)}(x,y) - F_{s,t}^{(n)}(x',y')\right\|_2$$
$$\geq -\frac{C-c}{c}\left(\frac{1}{(1+Cs)^{1/2}} - \frac{1}{(1+Ct)^{1/2}}\right)\|y-y'\|_2,$$

which yields the desired estimates. □

17.4 Transport in the large-t limit

In this section, we estimate the Lipschitz norms of the renormalized transport functions $\tilde{F}_{s,t}^{(n)}$, and demonstrate convergence as s or t goes to 0 or ∞. As the first step, we deduce from Lemma 17.3.2 the following Lipschitz estimates on $\tilde{F}_{s,t}^{(n)}$ which are uniform in s and t. Note also that the coefficient of $\|y-y'\|_2$ goes to zero as $s,t \to \infty$.

Lemma 17.4.1. *We have*

$$\left\|\tilde{F}_{s,t}^{(n)}\right\|_{\mathrm{Lip},dx} \leq \max(C,1/c)^{1/2} \tag{17.7}$$

and

$$\left\|\tilde{F}_{s,t}^{(n)}\right\|_{\mathrm{Lip},dy} \leq (C/c-1)\max(C,1/C)^{3/2}|e^{-s/2}-e^{-t/2}|. \tag{17.8}$$

In particular,

$$\left\|\tilde{F}_{s,t}^{(n)}\right\|_{\mathrm{Lip}} \leq \max(C,1/c)^{7/2}. \tag{17.9}$$

Proof. For the first estimate, for the case where $s \geq t$, direct substitution of (17.2) into (17.5) of Lemma 17.3.2 shows that

$$\left\|\tilde{F}_{s,t}^{(n)}\right\|_{\mathrm{Lip},dx} \leq e^{-s/2}\frac{(1+C(e^s-1))^{1/2}}{(1+C(e^t-1))^{1/2}}e^{t/2} = \frac{(C+(1-C)e^{-s})^{1/2}}{(C+(1-C)e^{-t})^{1/2}}.$$

The function $C+(1-C)e^{-s}$ is either increasing or decreasing for $s \in [0,\infty)$ and achieves the values 1 and C at 0 and ∞ respectively, and hence is between $\min(1,C)$ and $\max(1,C)$. Hence,

$$\left\|\tilde{F}_{s,t}^{(n)}\right\|_{\mathrm{Lip},dx} \leq \frac{\max(1,C)^{1/2}}{\min(1,C)^{1/2}} = \max(C,1/C)^{1/2} \leq \max(C,1/c)^{1/2}.$$

The case where $s \le t$ follows by the same argument, this time using the bound $\max(c, 1/c) \le \max(C, 1/c)$.

For the second estimate, we apply (17.6). Note in (17.6), in the case $s \le t$, we may use $(1+cs)^{1/2} \le (1+Cs)^{1/2}$ and thus in both cases $s \ge t$ or $s \le t$,

$$\begin{aligned}\left\|F^{(n)}_{s,t}\right\|_{\mathrm{Lip},dy} &\le (C/c-1)(1+Cs)^{1/2}\left|\frac{1}{(1+Cs)^{1/2}} - \frac{1}{(1+Ct)^{1/2}}\right| \\ &= (C/c-1)(1+Cs)^{1/2}\left|\int_s^t \frac{C}{2(1+Cu)^{3/2}}\,du\right|\end{aligned}$$

This implies that

$$\begin{aligned}\left\|\tilde{F}^{(n)}_{s,t}\right\|_{\mathrm{Lip},dy} &\le e^{-s/2}(C/c-1)(1+C(e^s-1))^{1/2}\left|\int_{e^s-1}^{e^t-1} \frac{C}{2(1+Cu)^{3/2}}\,du\right| \\ &= (C/c-1)e^{-s/2}(1+C(e^s-1))^{1/2}\left|\int_s^t \frac{Ce^w}{2(1+C(e^w-1))^{3/2}}\,dw\right| \\ &\le (C/c-1)\max(1,C)^{1/2}\left|\int_s^t \frac{Ce^w}{2\min(1,C)^{3/2}e^{3w/2}}\,dw\right| \\ &\le (C/c-1)\frac{\max(1,C)^{1/2}C}{\min(1,C)^{3/2}}|e^{-t/2}-e^{-s/2}| \\ &\le (C/c-1)\max(C,1/C)^{3/2}|e^{-t/2}-e^{-s/2}|.\end{aligned}$$

where we have again applied $\min(1,C)e^s \le 1 + C(e^s-1) \le \max(1,C)e^s$.

For the last estimate (17.9), observe that

$$\begin{aligned}\left\|\tilde{F}^{(n)}_{s,t}\right\|_{\mathrm{Lip}} &\le \left\|\tilde{F}^{(n)}_{s,t}\right\|_{\mathrm{Lip},dx} + \left\|\tilde{F}^{(n)}_{s,t}\right\|_{\mathrm{Lip},dy} \\ &\le \max(C,1/c)^{1/2} + (C/c-1)\max(C,1/C)^{3/2}|e^{-s/2}-e^{-t/2}| \\ &\le \max(C,1/c)^{3/2} + (\max(C,1/c)^2-1)\max(C,1/c)^{3/2} \\ &= \max(C,1/c)^{7/2}.\end{aligned}$$

□

Lemma 17.4.2. *Let* π_1 *denote the function* $\pi_1(x,y) = x$. *Then*

$$\left\|\tilde{F}^{(n)}_{s,t} - \pi_1\right\|_{\mathrm{Lip},dx} \le \frac{1}{2}(\max(C,1/c)-1)\max(C,1/C)^{1/2}|e^{-s}-e^{-t}| \tag{17.10}$$

and

$$\left\|\tilde{F}^{(n)}_{s,t} - \pi_1\right\|_{\mathrm{Lip}} \le (\max(C,1/c)^3-1)\max(C,1/c)^{1/2}|e^{-s/2}-e^{-t/2}|. \tag{17.11}$$

Proof. Let $U_s^{(n)}(x,y) = \tilde{V}_s^{(n)}(x,y) - (1/2)\|x\|_2^2$. Then (17.2) says that

$$\partial_s \tilde{F}_{s,t}^{(n)}(x,y) = \frac{1}{2}\nabla_x U_s^{(n)}\left(\tilde{F}_{s,t}^{(n)}(x,y), y\right).$$

Moreover, we have

$$\frac{ce^s}{1 + c(e^s - 1)} \leq H_x \tilde{V}_s^{(n)} \leq \frac{Ce^s}{1 + C(e^s - 1)}.$$

We can bound $H_x U_s$ above and below by subtracting 1 from both sides, which after some computation reduces to

$$\frac{c-1}{1 + c(e^s - 1)} \leq H_x U_s^{(n)} \leq \frac{C-1}{1 + C(e^s - 1)}.$$

Therefore, we have $-LI \leq H_x U_s \leq LI$ where

$$L := \max\left(-\frac{c-1}{1 + c(e^s - 1)}, \frac{C-1}{1 + C(e^s - 1)}\right).$$

We claim that $L \leq L' := (\max(C, 1/c) - 1)e^{-s}$. If the first term $(1-c)/(1 + c(e^s - 1))$ is negative, then it is $\leq L'$ automatically, but if it is positive, then $c \leq 1$ and hence

$$\frac{1-c}{1 + c(e^s - 1)} \leq \frac{1-c}{c + c(e^s - 1)} = (1/c - 1)e^{-s} \leq (\max(C, 1/c) - 1)e^{-s}.$$

Similarly, if $(C-1)/(1 + C(e^s - 1))$ is negative, then it is $\leq L'$ automatically, but otherwise $C \geq 1$, and hence

$$\frac{C-1}{1 + C(e^s - 1)} \leq \frac{C-1}{1 + (e^s - 1)} = (C-1)e^{-s} \leq (\max(C, 1/c) - 1)e^{-s}.$$

But $-L'I \leq H_x U_s^{(n)} \leq L'I$ implies that $\nabla_x U_s^{(n)}$ is L'-Lipschitz in x. Therefore,

$$\begin{aligned}\left\|\partial_s \tilde{F}_{s,t}^{(n)}(x,y) - \partial_s \tilde{F}_{s,t}^{(n)}(x',y)\right\|_2 &= \frac{1}{2}\left\|\nabla_x U_s^{(n)}(\tilde{F}_{s,t}^{(n)}(x,y), y) - \nabla_x U_s(\tilde{F}_{s,t}^{(n)}(x',y))\right\|_2 \\ &\leq \frac{1}{2}(\max(C, 1/c) - 1)e^{-s}\left\|\tilde{F}_{s,t}^{(n)}(x,y) - \tilde{F}_{s,t}^{(n)}(x',y)\right\|_2.\end{aligned}$$

Applying (17.7) in the case where $s \geq t$, we get

$$\left\|\partial_s \tilde{F}_{s,t}^{(n)}(x,y) - \partial_s \tilde{F}_{s,t}^{(n)}(x',y)\right\|_2 \leq \frac{1}{2}(\max(C, 1/c) - 1)\max(C, 1/c)^{1/2} e^{-s}\|x - x'\|_2.$$

Hence,

$$\begin{aligned}&\left\|\tilde{F}_{s,t}^{(n)}(x,y) - \tilde{F}_{s,t}^{(n)}(x',y) - (x - x')\right\|_2 \\ &\quad\leq \left|\int_t^s \left\|\partial_u \tilde{F}_{u,t}^{(n)}(x,y) - \partial_u \tilde{F}_{u,t}^{(n)}(x',y)\right\|_2 du\right| \\ &\quad\leq \frac{1}{2}(\max(C, 1/c) - 1)\max(C, 1/c)^{1/2}|e^{-s} - e^{-t}|\|x - x'\|_2.\end{aligned}$$

which proves the desired estimate (17.10).

To check the second estimate (17.11), first observe

$$\frac{1}{2}|e^{-s}-e^{-t}| = \int_{\min(s,t)}^{\max(s,t)} \frac{1}{2}e^{-u}\,du \le \int_{\min(s,t)}^{\max(s,t)} \frac{1}{2}e^{-u/2}\,du = |e^{-s/2}-e^{-t/2}|,$$

Moreover, $\|\tilde{F}^{(n)}_{s,t}-\pi_1\|_{\mathrm{Lip},dy} = \|\tilde{F}^{(n)}_{s,t}\|_{\mathrm{Lip},dy}$. Therefore, using (17.8) and (17.10),

$$\begin{aligned}
&\left\|\tilde{F}^{(n)}_{s,t}-\pi_1\right\|_{\mathrm{Lip}} \\
&\le \|\tilde{F}^{(n)}_{s,t}-\pi_1\|_{\mathrm{Lip},dx} + \|\tilde{F}^{(n)}_{s,t}-\pi_1\|_{\mathrm{Lip},dy} \\
&\le (\max(C,1/c)-1)\max(C,1/C)^{1/2}\frac{1}{2}|e^{-s}-e^{-t}| + (C/c-1)\max(C,1/C)^{3/2}|e^{-s/2}-e^{-t/2}| \\
&\le \left[(\max(C,1/c)-1)\max(C,1/c)^{1/2} + (\max(C,1/c)^2-1)\max(C,1/c)^{3/2}\right]|e^{-s/2}-e^{-t/2}| \\
&= (\max(C,1/c)^3-1)\max(C,1/c)^{1/2}|e^{-s/2}-e^{-t/2}|. \qquad \square
\end{aligned}$$

In the following proposition, we use the notation $\mathrm{Var}(X^{(n)})$ for the total variance $E\|X^{(n)}-E(X^{(n)})\|_2^2$.

Proposition 17.4.3. *The limits $\tilde{F}^{(n)}_{s,\infty} := \lim_{t\to\infty}\tilde{F}^{(n)}_{s,t}$ and $\tilde{F}^{(n)}_{\infty,t} = \lim_{s\to\infty}\tilde{F}^{(n)}_{s,t}$ exist for $s,t\ge 0$. More precisely, let $(X^{(n)},Y^{(n)})$ and be a pair of random variables with the laws $\mu^{(n)}$. Then*

$$\left\|\tilde{F}^{(n)}_{s,\infty}(x,y)-\tilde{F}^{(n)}_{s,t}(x,y)\right\|_2 \le \max(C,1/c)^{1/2}e^{-t/2}\|E(X)\|_2$$
$$+\, e^{-t/2}(\max(C,1/c)^3-1)\max(C,1/c)\left(\left\|(x,y-E(Y^{(n)}))\right\|_2 + (d_1+\mathrm{Var}(Y^{(n)}))^{1/2}\right) \quad (17.12)$$

and

$$\left\|\tilde{F}^{(n)}_{s,t}(x,y)-\tilde{F}^{(n)}_{\infty,t}(x,y)\right\|_2 \le \frac{1}{2}(\max(C,1/c)-1)\max(C,1/c)^{1/2}e^{-s}$$
$$\Bigg(e^{-s/2}\|E(X^{(n)})\|_2 + \max(C,1/c)^{7/2}\Big(\|(x-e^{-t/2}E(X^{(n)}),y-E(Y^{(n)}))\|_2$$
$$+\,(e^{-t}\,\mathrm{Var}(X^{(n)}) + (1-e^{-t})d_1 + \mathrm{Var}(Y^{(n)}))^{1/2}\Big)\Bigg) \quad (17.13)$$

The estimates of Lemmas 17.4.1 and 17.4.2 extend to the cases where s or t is infinite, where we define $\tilde{F}^{(n)}_{\infty,\infty}(x,y)=x$. Moreover, the relation $(\tilde{F}^{(n)}_{s,t},\pi_2)_*\tilde{\mu}^{(n)}_t = \tilde{\mu}^{(n)}_s$ holds for all $s,t\in[0,\infty]$

Proof. We first consider the case where s is fixed and $t \to \infty$. Note that by (17.7),

$$\left\|\tilde{F}^{(n)}_{s,t'}(x,y) - \tilde{F}^{(n)}_{s,t}(x,y)\right\|_2 = \left\|\tilde{F}^{(n)}_{s,t'}(\tilde{F}^{(n)}_{t',t}(x,y),y) - \tilde{F}^{(n)}_{s,t'}(x,y)\right\|_2 \tag{17.14}$$
$$\leq \left\|\tilde{F}^{(n)}_{s,t'}\right\|_{\mathrm{Lip},dx}\left\|\tilde{F}^{(n)}_{t,t'}(x,y) - x\right\|_2$$
$$\leq \max(C,1/c)^{1/2}\left\|\tilde{F}^{(n)}_{t,t'}(x,y) - x\right\|_2.$$

By Lemma 17.4.2,

$$\left\|\tilde{F}^{(n)}_{t,t'} - \pi_1\right\|_{\mathrm{Lip}} \leq L|e^{-t/2} - e^{t'/2}|,$$

where $L = \max(C,1/c)^3 - 1)\max(C,1/c)^{1/2}$. Then we apply Lemma 11.2.7 to the function $G^{(n)}(x,y) = \tilde{F}^{(n)}_{t,t'}(x,y) - x$ and a random variable $(\tilde{X}^{(n)}_{t'}, Y^{(n)})$ which has law $\tilde{\mu}^{(n)}_t$. Note that $(\tilde{X}^{(n)}_{t'}, Y^{(n)})$ has mean $(e^{-t'/2}E(X^{(n)}), E(Y))$ and variance $e^{-t'}\operatorname{Var}(X^{(n)}) + (1-e^{-t'})d_1 + \operatorname{Var}(Y^{(n)})$. Moreover,

$$E[G(\tilde{X}^{(n)}_{t'}, Y^{(n)})] = E[\tilde{X}^{(n)}_t] - E[\tilde{X}^{(n)}_{t'}] = (e^{-t/2} - e^{-t'/2})E(X^{(n)}).$$

Thus, by Lemma 11.2.7,

$$\left\|\tilde{F}^{(n)}_{t,t'}(x,y) - x\right\|_2 \leq |e^{-t/2} - e^{-t'/2}|\left\|E(X^{(n)})\right\|_2$$
$$+ L|e^{-t/2} - e^{t'/2}|\Bigg(\left\|(x - e^{-t'/2}E(X^{(n)}), y - E(Y^{(n)}))\right\|_2 + (e^{-t'}\operatorname{Var}(X^{(n)})$$
$$+ (1-e^{-t'})d_1 + \operatorname{Var}(Y^{(n)}))^{1/2}\Bigg). \tag{17.15}$$

Plugging this into (17.14), we see that $\tilde{F}^{(n)}_{s,t}$ is Cauchy in t as $t \to \infty$. Moreover, we obtain the estimate (17.12) by taking $t' \to \infty$ in (17.15) and multiplying by $\left\|\tilde{F}^{(n)}_{s,t}\right\|_{\mathrm{Lip},dx} \leq \max(c,1/c)^{1/2}$.

Now let us fix t and consider when s' and s approach ∞. The argument for this case is similar but antisymmetrical. We estimate

$$\left\|\tilde{F}^{(n)}_{s',t}(x,y) - \tilde{F}^{(n)}_{s,t}(x,y)\right\|_2 = \left\|\tilde{F}^{(n)}_{s',s}(\tilde{F}^{(n)}_{s,t}(x,y),y) - \tilde{F}^{(n)}_{s,t}(x,y)\right\|_2$$
$$\leq \left\|\tilde{F}^{(n)}_{s',s} - \pi_1\right\|_{\mathrm{Lip},dx}\left\|\tilde{F}^{(n)}_{s,t}(x,y)\right\|_2$$
$$\leq \frac{1}{2}(\max(C,1/c) - 1)\max(C,1/c)^{1/2}|e^{-s} - e^{-s'}|\left\|\tilde{F}^{(n)}_{s,t}(x,y)\right\|_2,$$

where the last line follows from (17.10). Then by applying Lemma 11.2.7 to the function

$\tilde{F}_{s,t}^{(n)}(x,y)$ and the random variable $(\tilde{X}_t^{(n)}, Y^{(n)})$, together with (17.9), we obtain

$$\begin{aligned}\left\|\tilde{F}_{s,t}^{(n)}(x,y)\right\|_2 \le e^{-s/2}\|E(X^{(n)})\|_2 \\ + \max(C,1/c)^{7/2}\Big(\|(x - e^{-t/2}E(X^{(n)}), y - E(Y^{(n)}))\|_2 + (e^{-t}\operatorname{Var}(X^{(n)}) \\ + (1-e^{-t})d_1 + \operatorname{Var}(Y^{(n)}))^{1/2}\Big)\end{aligned}$$

This produces an estimate on $\left\|\tilde{F}_{s',t}^{(n)} - \tilde{F}_{s,t}^{(n)}\right\|_2$ which shows that $\tilde{F}_{s,t}^{(n)}$ is Cauchy as $s \to \infty$, so that $\tilde{F}_{\infty,t}^{(n)}$ is well-defined. The explicit bound on the rate of convergence follows fixing s and t, combining the above estimates, and taking $s' \to \infty$.

Finally, since we have established convergence of $\tilde{F}_{s,t}^{(n)}$ as s or t approaches ∞, a routine argument with limits will extend the estimates of Lemmas 17.4.1 and 17.4.2, and the transport relations, to the cases where s or t is $+\infty$. Similarly, because $\tilde{F}_{s,t}^{(n)}$ depends continuously on s and t with the explicit estimates given in the foregoing argument, a straightforward approximation argument shows that the relation $(F_{s,t}^{(n)}, \pi_2)_* \tilde{\mu}_t^{(n)} = \tilde{\mu}_s^{(n)}$ extends to the case where s or t is zero. Of course, in working out the details of these limiting arguments, it is also helpful to use the explicit Lipschitz estimates we have for $\tilde{F}_{s,t}^{(n)}$. □

17.5 Transport in the large-n limit

Now we show that the maps $F_{s,t}^{(n)}$ and $\tilde{F}_{s,t}^{(n)}$ constructed above are asymptotically approximable by trace polynomials and finish the proof of Theorem 17.1.1.

Proposition 17.5.1. *Continue with the setup from §17.2. For any $s,t \in [0,\infty]$, the sequences $(F_{s,t}^{(n)})_{n\in\mathbb{N}}$ and $(\tilde{F}_{s,t}^{(n)})_{n\in\mathbb{N}}$ are asymptotically approximable by trace polynomials.*

Proof. Since $F_{s,t}^{(n)}$ is defined by solving the ODE (17.1), we will use Proposition 14.3.7 to show asymptotic approximability by trace polynomials, and thus we should check that $\nabla_x V_t^{(n)}(x,y)$ satisfies Assumption 14.3.6.

(1) We already know from Lemma 16.4.3 that $(\nabla_x V_t^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials.

(2) By Lemma 16.4.3, we have $0 \le HV_t^{(n)} \le CI$, and hence $\nabla_x V_t^{(n)}$ is C-Lipschitz for all t.

(3) Finally, to estimate the modulus of continuity of $\nabla_x V_t^{(n)}$ with respect to t, fix $s \le t$. We can then apply (16.9) with the initial potential $V_s^{(n)}$ and time step $t-s$. Since

$c/(1+cs) \leq HV_s^{(n)} \leq C$, we get

$$\begin{aligned}
&\left\|\nabla V_t^{(n)}(x,y) - \nabla V_s^{(n)}(x,y)\right\| \\
&\leq C\left(\frac{2d_1(t-s)}{1+c(t-s)/(1+cs)} + \frac{(t-s)^2}{(1+c(t-s)/(1+cs))^2}\|\nabla V_s^{(n)}(x,y)\|_2^2\right)^{1/2} \\
&= C\left(\frac{2d_1(t-s)(1+cs)}{1+ct} + \frac{(t-s)^2(1+cs)^2}{(1+ct)^2}\|\nabla V_s^{(n)}(x,y)\|_2^2\right)^{1/2} \\
&\leq C\sqrt{2d_1(t-s)(1+cs)} + C(t-s)(1+cs)\left\|\nabla V_s^{(n)}(x,y)\right\|_2.
\end{aligned}$$

But using the same estimate for times 0 and s and using the triangle inequality again

$$\left\|\nabla V_s^{(n)}(x,y)\right\|_2 \leq \left\|\nabla V^{(n)}(x,y)\right\|_2 + C\sqrt{2d_1 s} + Cs\left\|\nabla V^{(n)}(x,y)\right\|_2.$$

Thus,

$$\begin{aligned}
&\left\|\nabla V_t^{(n)}(x,y) - \nabla V_s^{(n)}(x,y)\right\| \\
&\leq C\sqrt{2d_1(t-s)(1+cs)} + C^2(t-s)(1+cs)\left(\sqrt{2d_1 s} + (1+Cs)\|\nabla V^{(n)}(x,y)\|_2\right).
\end{aligned}$$

Since $(\nabla V^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials, it is in particular uniformly bounded in $\|\cdot\|_2$ on operator norm balls. Hence, we have an estimate for $\|\nabla V_t^{(n)} - \nabla V_s^{(n)}\|_2$ that goes to zero uniformly on operator norm balls as $t-s \to 0$, as long as s and t remain in some given compact time interval.

Since Assumption 14.3.6 holds, Proposition 14.3.7 implies that $(F_{s,t}^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials for any $s,t \in [0,\infty)$. This in turn implies that $(\tilde{F}_{s,t}^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials for any $s,t \in [0,\infty)$, since $\tilde{F}_{s,t}^{(n)}$ only differs from $F_{s,t}^{(n)}$ by rescaling. Finally, we can extend the conclusion to the cases where s or t is ∞ by taking limits, relying on Lemma 13.5.13 and Proposition 17.4.3. □

Proof of Theorem 17.1.1. Define $F^{(n)} = \tilde{F}_{\infty,0}^{(n)}$ and $G^{(n)} = \tilde{F}_{0,\infty}^{(n)}$. By Proposition 17.4.3, $(F^{(n)},\pi_2)$ pushes forward $\tilde{\mu}_0 = \mu$ to $\tilde{\mu}_\infty$, which is the law of $(Z_1^{(n)}, Y^{(n)})$. Meanwhile, $(G^{(n)},\pi_2)$ does the reverse, and it is the inverse function of $(F^{(n)},\pi_2)$. Hence, (1) and (2) of the theorem hold.

(3) follows from Proposition 17.5.1.

(4) Let $\tilde{\rho}_t$ be the density corresponding to $\tilde{V}_t$ as in §17.2. Note that $\tilde{\rho}_{n^2 t}$ satisfies the equation

$$\partial_t[\tilde{\rho}_{n^2 t}] = \frac{1}{2}\Delta_x \tilde{\rho}_{n^2 t} - \frac{1}{2}\langle n^2 x, \nabla \tilde{\rho}_{n^2 t}\rangle_2,$$

which is exactly the heat equation relative to the measure $\sigma_t^{(n)}$. It follows the maps $\tilde{F}_{s,t}$ constructed by (17.2) are the same maps used in Otto and Villani's proof of the Talagrand inequality (see the proof of Theorem 12.3.5). Although $V^{(n)}$ might not satisfy the smoothness hypotheses of the theorem, certainly $\tilde{\rho}_t(\cdot|y)$ is smooth because it is the solution to the heat equation. Hence, if we apply that result to the conditional density of $X^{(n)} + Z_t^{(n)}$ given $Y^{(n)}$ and then integrate with respect to y, we obtain for $t > 0$ that

$$\left\| F_{\infty,t}^{(n)}(\tilde{X}_t^{(n)}, Y^{(n)}) - \tilde{X}_t^{(n)} \right\|_{L^2}^2 \leq -\frac{2}{n^2} h_g^{(n)}(\tilde{X}_t^{(n)} | Y^{(n)}).$$

This is because $\sigma^{(n)}$ satisfies the log-Sobolev inequality with constant n^2. By limiting arguments as in Lemma 12.1.4 (for instance, using the conversion between entropy relative to Gaussian and Lebesuge measure together with Lemma 12.1.2), we can obtain

$$\lim_{t \to \infty} h_g^{(n)}(\tilde{X}_t^{(n)} | Y^{(n)}) = h_g^{(n)}(X^{(n)} | Y^{(n)}).$$

Therefore, in the limit,

$$\left\| \tilde{F}_{\infty,0}^{(n)}(X^{(n)}, Y^{(n)}) - X^{(n)} \right\|_{L^2}^2 \leq -\frac{2}{n^2} h_g^{(n)}(\tilde{X}_t^{(n)} | Y^{(n)}).$$

Moreover, using the push-forward and composition relations for $\tilde{F}_{s,t}$, we have

$$\left\| \tilde{F}_{0,\infty}^{(n)}(Z^{(n)}, Y^{(n)}) - Z^{(n)} \right\|_{L^2} = \left\| X^{(n)} - \tilde{F}_{\infty,0}^{(n)}(X^{(n)}, Y^{(n)}) \right\|_{L^2},$$

which completes the proof of (4).

(5) follows from Lemma 17.4.2. □

Remark 17.5.2. Of course, we can apply the same reasoning to $\tilde{F}_{s,t}^{(n)}$ that we did to $F^{(n)}$ and $G^{(n)}$ in the proof of Theorem 17.1.4 (5). Let X, Y, and Z_t be the non-commutative random variables from Theorem 17.1.4, let $\tilde{X}_t = e^{-t/2} X + e^{-t/2} Z_{e^t - 1}$, let $\tilde{F}_{s,t}^{(n)} \rightsquigarrow \tilde{F}_{s,t}$. Then

$$\tilde{F}_{s,t}(\tilde{X}_s, Y) \sim (\tilde{X}_t, Y).$$

Moreover,

$$\left\| \tilde{F}_{s,t}(\tilde{X}_s, Y) - \tilde{X}_s \right\|_2^2 \leq 2 |\chi_g^*(\tilde{X}_t | Y) - \chi_g^*(\tilde{X}_s | Y)|.$$

and

$$\begin{aligned} \left\| \tilde{F}_{s,t}(\tilde{X}_s, Y) - \tilde{X}_s \right\|_\infty &\leq \Theta(\max(C, 1/c)^3 - 1) \max(C, 1/c) + |\tau(\tilde{X}_t) - \tau(\tilde{X}_s)| \\ &= \Theta(\max(C, 1/c)^3 - 1) \max(C, 1/c) + |\tau(X)(e^{-t/2} - e^{-s/2})|. \end{aligned}$$

CHAPTER 18

Examples and applications II

In this chapter, we describe some examples to which our previous results apply. In particular, we show that our triangular transport result Theorem 17.1.9 applies to free Gibbs laws obtained from perturbations of a quadratic potential on an operator-norm ball, as well as perturbations of a semicircular family by transport. These results are not intended to be as general as possible, but rather as concrete illustrations of the applicability of the general theory we have developed thus far.

18.1 Functional calculus and the free difference quotient

In order to perform operator-norm cut-off tricks for our applications in this chapter, we must consider application of a smooth functions ϕ to self-adjoint operators, and get some control over their non-commutative derivatives. We also have several claims about functional calculus with smooth functions that we did not justify earlier in the text. In Lemma 13.3.6 (3), we claimed that the $\|\cdot\|_2$-Lipschitz norm of ϕ can be estimated using the Fourier transform $\widehat{\phi}$, and in the proof of Proposition 16.2.4, we claimed that if $\phi \in C_c^\infty(\mathbb{R})$, then ϕ defines a C^∞ function $M_n(\mathbb{C})_{\text{sa}} \to M_n(\mathbb{C})_{\text{sa}}$.

These claims come from a more general theory of how Voiculescu's free difference quotients can be extended to smooth functions, and how they can be estimated in a tensor norm by using the Fourier transform. The work of Peller and Aleksandrov [Pel06, AP10b, AP10a, AP17a] has developed sharp estimates and shown how Besov spaces of functions $\mathbb{R} \to \mathbb{R}$ are the correct spaces for this purpose. However, we will settle for giving a "baby version" that can be explained in a self-contained way without introducing too many new definitions, yet is still sufficiently powerful for our applications.

18.1.1 Setup of $C_{\text{nc}}^k(\mathbb{R})$

Let $\mathcal{D} : \mathbb{C}[X] \to \mathbb{C}[X] \otimes \mathbb{C}[X]$ be Voiculescu's free difference quotient in one variable (see §14.1). Let us define for k,

$$\mathcal{D} : \mathbb{C}[X]^{\otimes k} \to \mathbb{C}[X]^{\otimes (k+1)}$$

by

$$\mathcal{D} = \sum_{j=1}^{k} \mathrm{id}^{\otimes(j-1)} \otimes \mathcal{D} \otimes \mathrm{id}^{\otimes(n-j)},$$

so that $\mathcal{D}^k$ maps $\mathbb{C}[X]$ into $\mathbb{C}[X]^{\otimes(k+1)}$. We also define a multiplication operation $\mathbb{C}[X]^{\otimes(k+1)} \times \mathbb{C}[X]^{\otimes(\ell+1)} \to \mathbb{C}[X]^{\otimes(k+\ell+1)}$ by

$$(f_0 \otimes \cdots \otimes f_\ell)(g_0 \otimes \cdots \otimes g_\ell) = f_0 \otimes \cdots \otimes f_{k-1} \otimes f_\ell g_0 \otimes g_1 \otimes \cdots \otimes g_\ell.$$

Lemma 18.1.1. *The operator $\mathcal{D}$ satisfies the Leibniz product rule in the sense that for $f \in \mathbb{C}[X]^{\otimes(k+1)}$ and $g \in \mathbb{C}[X]^{\otimes(\ell+1)}$, we have*

$$\mathcal{D}(fg) = \mathcal{D}f \cdot g + f \cdot \mathcal{D}g.$$

Proof. First, consider the case where $k = \ell = 0$. Then we have to show that

$$\mathcal{D}[X^m X^n] = \mathcal{D}[X^m]X^n + X^m \mathcal{D}[X^n].$$

This is a direct computation using the fact that

$$\mathcal{D}[X^n] = \sum_{j=0}^{n} X^j \otimes X^{n-j}.$$

Now we prove the claim for general k and ℓ. It suffices to check it for simple tensors $f = f_0 \otimes \cdots \otimes f_k$ and $g = g_0 \otimes \cdots \otimes g_\ell$. Now

$$\mathcal{D}f = \sum_{j=0}^{k} f_0 \otimes \cdots \otimes f_{j-1} \otimes \mathcal{D}f_j \otimes f_{j+1} \otimes \cdots \otimes f_k,$$

and

$$\mathcal{D}g = \sum_{j=0}^{\ell} g_0 \otimes \cdots \otimes g_{j-1} \otimes \mathcal{D}g_j \otimes g_{j+1} \otimes \cdots \otimes g_k.$$

Meanwhile,

$$\begin{aligned}\mathcal{D}[fg] = &\sum_{j=0}^{k-1} f_0 \otimes \cdots \otimes f_{j-1} \otimes \mathcal{D}f_j \otimes f_{j+1} \otimes \cdots \otimes f_k g \\ &+ f_0 \otimes \cdots \otimes f_{k-1} \otimes \mathcal{D}[f_k g_0] \otimes g_1 \otimes \cdots \otimes g_\ell \\ &+ \sum_{j=0}^{\ell} f g_0 \otimes \cdots \otimes g_{j-1} \otimes \mathcal{D}g_j \otimes g_{j+1} \otimes \cdots \otimes g_k.\end{aligned}$$

Substituting $\mathcal{D}[f_k g_0] = \mathcal{D}f_k \cdot g_0 + f_k \cdot \mathcal{D}g_0$ completes the argument. □

We want to define $C^k_{\mathrm{nc}}(\mathbb{R})$ as a certain Fréchet-space completion of non-commutative polynomials. The question is how to measure the norm of $\mathcal{D}^k p$ over $|x| \leq R$. To do this, we use the *projective tensor product* $C([-R,R])^{\widehat{\otimes} k}$ defined as follows. For f in the algebraic tensor product, we define the norm as

$$\|f\| = \inf \left\{ \sum_{j=1}^N \|f_{j,1}\|_{C([-R,R])} \dots \|f_{j,k}\|_{C([-R,R])} : f = \sum_{j=1}^N f_{j,1} \otimes \dots \otimes f_{j,k} \right\}.$$

Then $C([-R,R])^{\widehat{\otimes} k}$ is the completion of the algebraic tensor product with respect to this norm.

Definition 18.1.2. We define $C^k_{\mathrm{nc}}(\mathbb{R})$ as the completion of $\mathbb{C}[X]$ with respect to the family of seminorms

$$\{\|\mathcal{D}^j p\|_{C([-R,R])^{\widehat{\otimes}(j+1)}} : j = 0, \dots, k; R > 0\}.$$

Similarly, $C^\infty_{\mathrm{nc}}(\mathbb{R})$ is defined using the same family of seminorms but for $j \in \mathbb{N}_0$.

Note that $\mathcal{D}^k$ extends to a well-defined map $C^k_{\mathrm{nc}}(\mathbb{R}) \to C([-R,R])^{\widehat{\otimes}(j+1)}$ for each $j \leq k$. We also observe that there is a well-defined multiplication map

$$C([-R,R])^{\widehat{\otimes}(k+1)} \times C([-R,R])^{\widehat{\otimes}(\ell+1)} \to C([-R,R])^{\widehat{\otimes}(k+\ell+1)},$$

defined similarly as in the polynomial case, and it satisfies

$$\|fg\|_{C([-R,R])^{\widehat{\otimes}(k+\ell+1)}} \leq \|f\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} \|g\|_{C([-R,R])^{\widehat{\otimes}(\ell+1)}}.$$

Now the following lemma describes the behavior of multiplication on $C^k_{\mathrm{nc}}(\mathbb{R})$.

Lemma 18.1.3. *There is a unique multiplication operation $C^k_{\mathrm{nc}}(\mathbb{R}) \times C^k_{\mathrm{nc}}(\mathbb{R})$ extending the multiplication on polynomials, and we have*

$$\mathcal{D}^k[fg] = \sum_{k_1,k_2 : k_1+k_2=k} \frac{k!}{k_1! k_2!} \mathcal{D}^{k_1} f \cdot \mathcal{D}^{k_2} g.$$

Proof. First, one checks the asserted formula in the case of polynomials. This is a standard induction argument that is identical to the one used in one-variable calculus. This formula implies that

$$\|\mathcal{D}^k[fg]\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} \leq \sum_{k_1,k_2 : k_1+k_2=k} \frac{k!}{k_1! k_2!} \|\mathcal{D}^{k_1} f\|_{C([-R,R])^{\widehat{\otimes}(k_1+1)}} \|\mathcal{D}^{k_2} g\|_{C([-R,R])^{\widehat{\otimes}(k_2+1)}}.$$

This implies that the multiplication operation extends to the completion $C^k_{\mathrm{nc}}(\mathbb{R})$ and the asserted bounds also hold in the completion. □

18.1.2 $C^k_{\mathrm{nc}}(\mathbb{R})$ estimates through the Fourier transform

Lemma 18.1.4. *For each $t \in \mathbb{R}$, the power series $\psi_t(X) = e^{2\pi i t X} = \sum_{m=0}^{\infty}(1/m!)(2\pi i t X)^m$ converges in $C^\infty_{\mathrm{nc}}(\mathbb{R})$, and we have*

$$\|\mathcal{D}^k \psi_t\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} \leq |2\pi t|^k \text{ for all } k, R, t.$$

Proof. Because $\mathcal{D}$ satisfies the Leibniz rule, we have

$$\mathcal{D}^k[X^\ell] = \sum_{\substack{k_1,\dots,k_\ell \geq 0 \\ k_1+\dots+k_\ell = k}} \frac{k!}{k_1! \dots k_\ell!} \mathcal{D}^{k_1}[X] \dots \mathcal{D}^{k_\ell}[X].$$

Now $\mathcal{D}[X] = 1 \otimes 1$ and $\mathcal{D}^2[X] = 0$, so that the only terms that contribute to the sum are those where $k_j = 0$ or 1. Choosing the integers k_1, ..., k_ℓ that sum up to k is equivalent to choosing a subset of $\{1, \dots, \ell\}$ of cardinality k, and the number of terms is ℓ choose k. Each term can be bounded by $R^{k-\ell}1^k$ since $\|\mathcal{D}[X]\|_{C([-R,R])^{\widehat{\otimes}2}} = 1$ and $\|X\|_{C([-R,R])} = R$. Thus, we have

$$\left\|\mathcal{D}^k[X^\ell]\right\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} \leq R^{\ell-k} \frac{\ell!}{(\ell-k)!} \text{ for } k \leq \ell,$$

and it is zero if $k > \ell$. This implies that for every k,

$$\sum_{m=0}^{\infty} \frac{1}{m!} |2\pi t|^m \|\mathcal{D}^k[X^m]\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} \leq \sum_{m=k}^{\infty} \frac{1}{(m-k)!} |2\pi t|^m R^{m-k} \leq |2\pi t|^k e^{|2\pi t| R}.$$

We also have $\|\psi_t\|_{C([-R,R])} \leq 1$, and so we may write

$$\|\mathcal{D}^k \psi_t\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} \leq |2\pi t|^k e^{|2\pi t| Rk}.$$

To finish the proof, note that $\psi_t = \psi_{t/n}^n$ for each $n \in \mathbb{N}$. Thus, we have

$$\mathcal{D}^k[\psi_t] = \sum_{\substack{k_1,\dots k_n \geq 0 \\ k_1+\dots+k_n = k}} \frac{k!}{k_1! \dots k_n!} \mathcal{D}^{k_1}\psi_{t/n} \dots \mathcal{D}^{k_n}\psi_{t/n}.$$

Hence,

$$\begin{aligned} \|\mathcal{D}^k[\psi_t]\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} &\leq \sum_{\substack{k_1,\dots k_n \geq 0 \\ k_1+\dots+k_n = k}} \prod_{j=1}^{n} |2\pi t/n|^{k_j} e^{|2\pi t| R k_j / n} \\ &= |2\pi t|^k e^{|2\pi t| Rk/n} \frac{1}{n^k} \sum_{\substack{k_1,\dots k_n \geq 0 \\ k_1+\dots+k_n = k}} \frac{k!}{k_1! \dots k_\ell!}. \end{aligned}$$

Now

$$\sum_{\substack{k_1,\dots k_n \geq 0 \\ k_1+\dots+k_n=k}} \frac{1}{k_1!\dots k_\ell!}$$

is the kth power series coefficient in $(e^t)^n = e^{nt}$ computed using the Cauchy product of power series, and hence the sum evaluates to $n^k/k!$. Hence,

$$\|\mathcal{D}^k[\psi_t]\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} \leq |2\pi t|^k e^{|2\pi t|Rk/n}.$$

But since n was arbitrary, we can take $n \to \infty$ on the right hand side, which completes the proof of the asserted inequality. □

We use the following normalization of the Fourier transform:

$$\widehat{\phi}(t) = \int_{\mathbb{R}} e^{-2\pi i s t} \phi(s)\, ds.$$

Proposition 18.1.5. *Let $\phi : \mathbb{R} \to \mathbb{C}$ such that $\widehat{\phi}$ is continuous with $\int (1+|t|^k)|\widehat{\phi}(t)|\, dt < \infty$. Then there exists a function $f \in C^k_{\mathrm{nc}}(\mathbb{R})$ such that the projection of f onto $C([-R,R])$ agrees with ϕ for all R, and we have*

$$\left\|\mathcal{D}^k f\right\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} \leq (2\pi)^k \int_{\mathbb{R}} |t^k \widehat{\phi}(t)|\, dt.$$

Proof. Let $\psi_t(X) = e^{2\pi i t X}$ in $C^k_{\mathrm{nc}}(\mathbb{R})$. Note that $t \mapsto \psi_t$ is a continuous function $\mathbb{R} \to C^k_{\mathrm{nc}}(\mathbb{R})$ because $\psi_{t+\epsilon} = \psi_t \psi_\epsilon$ and $\|\mathcal{D}^k(\psi_\epsilon - 1)\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} = O(\epsilon)$. Therefore, $t \mapsto \widehat{\phi}(t)\psi_t$ is continuous, and hence the Riemann integral

$$f_M := \int_{-M}^{M} \widehat{\phi}(t)\psi_t\, dt$$

is well-defined in $C^k_{\mathrm{nc}}(\mathbb{R})$. Moreover,

$$\|\mathcal{D}^k f_M\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} \leq \int_{-M}^{M} |\widehat{\phi}(t)| \|\psi_t\|_{C([-R,R])^{\widehat{\otimes}(k+1)}}\, dt \leq (2\pi)^k \int_{-M}^{M} |t^k \widehat{\phi}(t)|\, dt.$$

By the same token, if $M' \geq M$, then

$$\|\mathcal{D}^k f_M - \mathcal{D}^k f_{M'}\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} \leq (2\pi)^k \int_{M \leq |t| \leq M'} |t^k \widehat{\phi}(t)|\, dt.$$

This shows that (f_M) is Cauchy as $M \to \infty$, and hence converges to some f in $C^k_{\mathrm{nc}}(\mathbb{R})$. The projection of f onto $C([-R,R])$ is given by

$$f(s) = \int_{\mathbb{R}} \widehat{\phi}(t)\psi_t(s)\, dt = \phi(s) \text{ for } s \in [-R,R],$$

using the Fourier inversion formula. □

18.1.3 Application of $C^k_{\mathrm{nc}}(\mathbb{R})$ to functional calculus

Let $x \in M_n(\mathbb{C})_{\mathrm{sa}}$ with $\|x\|_\infty \leq R$. Then we define the evaluation map

$$C([-R,R])^{\widehat{\otimes}(k+1)} \to M_n(\mathbb{C})^{\otimes(k+1)} : f \mapsto f(x)$$

by

$$(f_0 \otimes \dots \otimes f_k)(x) = f_0(x) \otimes \dots \otimes f_k(x).$$

This map is well-defined for the following reasons. First, $f_j(x)$ is defined by functional calculus. By the spectral theorem, $\|f_j(x)\|_\infty \leq \|f_j\|_{C([-R,R])}$. This implies that

$$\|f(x)\|_{M_n(\mathbb{C})^{\widehat{\otimes}(k+1)}} \leq \|f\|_{C([-R,R])^{\widehat{\otimes}(k+1)}},$$

for a simple tensor, and hence for every element in the algebraic tensor product, and thus the map extends to the completion. Here $M_n(\mathbb{C})^{\widehat{\otimes}(k+1)}$ denotes the projective tensor product; since $M_n(\mathbb{C})_{\mathrm{sa}}$ is finite-dimensional, this just amounts to a certain choice of norm on the algebraic tensor product.

Moreover, we define the hash operation

$$\# : M_n(\mathbb{C})^{\otimes(k+1)} \times M_n(\mathbb{C})^k \to M_n(\mathbb{C})$$

by

$$z_0 \otimes \dots \otimes z_k \#(y_1, \dots, y_k) = z_0 y_1 z_1 \dots y_k z_k.$$

In particular, for every $f \in C^k_{\mathrm{nc}}(\mathbb{R})$ and $x \in M_n(\mathbb{C})_{\mathrm{sa}}$ and $y_1, \dots, y_k$ in $M_n(\mathbb{C})_{\mathrm{sa}}$, the expression $\mathcal{D}^k f(x)\#(y_1, \dots, y_k)$ is well-defined in $M_n(\mathbb{C})$. Moreover, we have

$$\|\mathcal{D}^k f(x)\#(y_1,\dots,y_k)\|_\infty \leq \|\mathcal{D}^k f\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} \|y_1\|_\infty \dots \|y_j\|_\infty. \tag{18.1}$$

Proposition 18.1.6. *Suppose that $f \in C^k_{\mathrm{nc}}(\mathbb{R})$. Let $f^{(n)} : M_n(\mathbb{C})_{\mathrm{sa}} \to M_n(\mathbb{C})$ be the evaluation of f on a matrix through functional calculus. For $y \in M_n(\mathbb{C})_{\mathrm{sa}}$, let ∂_y denote the directional derivative in direction y. Then $f^{(n)}$ is a C^k function and*

$$\partial_{y_1} \dots \partial_{y_j} f^{(n)}(x) = \frac{1}{j!} \sum_{\sigma \in S_j} \mathcal{D}^j f(x)\#(y_{\sigma(1)}, \dots, y_{\sigma(j)}),$$

where S_j is the permutation group on $\{1, \dots, j\}$.

Proof. First, consider the case where f is a monomial $f(X) = X^m$. Then similar to our previous computations,

$$\frac{1}{j!}\mathcal{D}^j f(X) = \sum_{\substack{j_1,\dots,j_m \in \{0,1\} \\ j_1+\dots+j_m = j}} \mathcal{D}^{j_1}[X] \dots \mathcal{D}^{j_m}[X] = \sum_{\substack{i_0,\dots,i_k \geq 0 \\ i_0+\dots+i_k = m-k}} X^{i_0} \otimes \dots \otimes X^{i_k},$$

where the last equality follows from some elementary combinatorics. Thus, for matrices x and y_i,

$$\frac{1}{j!}\sum_{\sigma \in S_j} \mathcal{D}^j f(x) \# (y_1, \dots, y_j) = \sum_{\sigma \in S_j} \sum_{\substack{i_0, \dots, i_k \geq 0 \\ i_0 + \dots + i_k = m-k}} x^{i_0} y_{\sigma(1)} x^{i_1} \dots y_{\sigma(k)} x^{i_k}.$$

On the other hand, $\partial_{y_1} \dots \partial_{y_j} f^{(n)}$ is the $t_1 \dots t_j$ term in the Taylor expansion of $f^{(n)}(x + t_1 y_1 + \dots + t_j y_j)$ in the variables $(t_1, \dots, t_j)$ since this function is polynomial in $(t_1, \dots, t_j)$. Expanding out $(x + t_1 y_1 + \dots + t_j y_j)^m$, we see that the $t_1 \dots t_j$ term is exactly

$$\sum_{\sigma \in S_j} \sum_{\substack{i_0, \dots, i_k \geq 0 \\ i_0 + \dots + i_k = m-k}} x^{i_0} y_{\sigma(1)} x^{i_1} \dots y_{\sigma(k)} x^{i_k},$$

which proves the asserted formula. It follows by linearity that this holds for all polynomials $p \in \mathbb{C}[X]$.

Finally, suppose $f \in C^k_{\mathrm{nc}}(\mathbb{R})$ and $(p_k)_{k \in \mathbb{N}}$ is a sequence of polynomials converging to f in $C^k_{\mathrm{nc}}(\mathbb{R})$. Then because of (18.1), we see that $\partial_{y_1} \dots \partial_{y_j} p_k^{(n)}$ converge uniformly on $\|X\|_\infty \leq R$ for each R and each $y_1, \dots, y_k$. Therefore, $f^{(n)}$ has continuous directional derivatives up to order k, and thus is a C^k function. □

Corollary 18.1.7. *Let $f \in C^1_{\mathrm{nc}}(\mathbb{R})$, and let $f^{(n)}$ be the evaluation of f on $n \times n$ self-adjoint matrices. Then*

$$Jf^{(n)}(x)[y] = \mathcal{D}f(x) \# y. \tag{18.2}$$

Moreover,

$$\|f^{(n)}(x) - f^{(n)}(x')\|_2 \leq \|\mathcal{D}f\|_{C([-R,R])^{\hat{\otimes} 2}} \|x - x'\|_2 \text{ for } \|x\|_\infty, \|x'\|_\infty \leq R. \tag{18.3}$$

Similarly, if $f^{(n)} \in C^2_{\mathrm{nc}}(\mathbb{R})$, then

$$\|Jf^{(n)}(x)[y] - Jf^{(n)}(x')[y]\|_2 \leq \|\mathcal{D}^2 f\|_{C([-R,R])^{\hat{\otimes} 3}} \|y\|_\infty \|x - x'\|_2 \text{ for } \|x\|_\infty, \|x'\|_\infty \leq R. \tag{18.4}$$

Proof. The claim (18.2) follows because $Jf^{(n)}(x)[y]$ is the directional derivative in direction y. Next, note that for $g = g_0 \otimes g_1 \in C([-R,R])^{\hat{\otimes} 2}$ and $\|x\|_\infty \leq R$, we have

$$\|g(x) \# y\|_2 \leq \|g_0(x)\|_\infty \|Y\|_2 \|g_1(x)\|_\infty \leq \|g_0\|_{C([-R,R])} \|g_1\|_{C([-R,R])} \|y\|_2,$$

which implies by definition of the projective tensor product norm that for all $g \in C([-R,R])^{\hat{\otimes} 2}$,

$$\|g(x) \# y\|_2 \leq \|g\|_{C([-R,R])^{\hat{\otimes} 2}} \|y\|_2.$$

When we apply this to $g = \mathcal{D}f$, we obtain that the operator norm of $Jf^{(n)}(x)$ as a map $(M_n(\mathbb{C})_{\mathrm{sa}}, \|\cdot\|_2) \to (M_n(\mathbb{C}), \|\cdot\|_2$ is bounded by $\|\mathcal{D}f\|_{C([-R,R])^{\hat{\otimes} 2}}$ for $\|x\|_\infty \leq R$. Since this ball is convex, a uniform estimate on the Jacobian is equivalent to a Lipschitz estimate, which proves (18.3).

Finally, to prove (18.4), we apply the same argument to the function $\mathcal{D}f(x)\#y$ instead of $f(x)$. The relevant inequality in this case is that for $\|x\|_\infty \leq R$,

$$\|\mathcal{D}^2 f(x)\#(y,z)\|_2 \leq \|\mathcal{D}^2 f\|_{C([-R,R])^{\widehat{\otimes} 3}} \|y\|_\infty \|z\|_2. \qquad \square$$

Remark 18.1.8. In particular, suppose that $\phi : \mathbb{R} \to \mathbb{C}$ with $(1+|t|)\widehat{\phi}(t)$ continuous and integrable. Then using Proposition 18.1.5 and Corollary 18.1.7, we obtain

$$\|\phi(x) - \phi(x')\|_2 \leq 2\pi \int_{\mathbb{R}} |t\widehat{\phi}(t)|\, dt \cdot \|x - x'\|_2,$$

for $x, x' \in M_n(\mathbb{C})_{\text{sa}}$. Here we do not need to impose restrictions on $\|x\|_\infty$ and $\|x'\|_\infty$ because Proposition 18.1.5 has a uniform estimate for all R. This proves Lemma 13.3.4 (3).

Remark 18.1.9. Besides (18.1) and the estimates in Corollary 18.1.7, we more generally have the following inequality. Let $(\mathcal{M}, \tau)$ be a tracial W*-algebra. Let $\|x\|_\alpha = \tau((x^* x)^{\alpha/2})^{1/\alpha}$ for $\alpha \geq 1$. Then if $x \in \mathcal{M}_{\text{sa}}$ with $\|x\| \leq R$ and if $y_1, \dots, y_k \in \mathcal{M}$, then

$$\|\mathcal{D}^k f(x)\#(y_1, \dots, y_k)\|_\alpha \leq \|\mathcal{D}^k f\|_{C([-R,R])^{\widehat{\otimes}(k+1)}} \|y_1\|_{\alpha_1} \dots \|y_j\|_{\alpha_k},$$

whenever $\alpha, \alpha_0, \dots, \alpha_k \in [1, \infty]$ with $1/\alpha = 1/\alpha_0 + \dots + 1/\alpha_k$. This follows from the non-commutative Hölder's inequality; see for instance [Sim05, Theorem 1.15 and 2.8], [da 18, Theorems 2.4 - 2.6], [PX03, §2].

Remark 18.1.10. This theory of non-commutative differential calculus is closely related to the differentiation of fully matricial functions in §3.

Remark 18.1.11. Future work should study a theory of "trace C^k functions" in variables X_1, ..., X_d that combines the ideas of this section with those of §13 and §14.1.

18.2 Perturbations of the quadratic potential on a $\|\cdot\|_\infty$-ball

All our previous results used Assumption 15.1.1, which in particular supposes that $HV^{(n)} \leq CI$ everywhere. This restriction is inconvenient because one of the main motivating examples is the case when $V^{(n)}$ is a trace polynomial. The only trace polynomials that satisfy $cI \leq HV^{(n)} \leq CI$ are of degree 2. Furthermore, since the random matrix tuples $X^{(n)}$ are almost surely bounded in $\|\cdot\|_\infty$ as $n \to \infty$, why should the potential $V^{(n)}$ have to be globally defined?

This section will partially address these issues by studying random matrix models supported in a $\|\cdot\|_\infty$-ball with a potential which is a perturbation of $(1/2)\|x\|_2^2$. This is based on [Jek18, §8].

Theorem 18.2.1. *Let $\epsilon > 0$. Let*

$$W^{(n)} : \{x \in M_n(\mathbb{C})_{\text{sa}}^d : \|x\|_\infty \leq 2 + 2\epsilon\} \to \mathbb{R}$$

satisfy

(A) $W^{(n)}$ is unitarily invariant.

(B) $W^{(n)}$ is C^1 and $\nabla W^{(n)}$ is K-Lipschitz on $\{\|x\|_\infty \leq 2 + 2\epsilon\}$.

(C) $\|\nabla W^{(n)}(0)\|_\infty \leq M$.

(D) $\nabla W^{(n)}$ is asymptotically approximable by trace polynomials on the $\|\cdot\|_\infty$-ball of radius $2 + 2\epsilon$, that is, there exists $f \in C_{\mathrm{tr,app}}(\mathbb{R}^{*d}, \|\cdot\|_2)_{\mathrm{sa}}^d$ with

$$\lim_{n\to\infty} \|\nabla W^{(n)} - f\|_{2,2+\epsilon}^{(n)} = 0.$$

For $\delta > 0$, let

$$V_\delta^{(n)}(x) = \frac{1}{2}\|x\|_2^2 + \delta W^{(n)}(x),$$

let

$$d\mu_\delta^{(n)}(x) = \frac{1}{\int_{\|x\|_\infty \leq 2+\epsilon} e^{-n^2 V_\delta^{(n)}}} 1_{\|x\|_\infty \leq 2+\epsilon} e^{-n^2 V_\delta^{(n)}(x)}\, dx,$$

and let $X_\delta^{(n)}$ be a random matrix tuple chosen according to $\mu_\delta^{(n)}$. There exist constants A_1 and A_2 depending only on d such that whenever

$$\delta < \frac{\min(\epsilon, \epsilon^2)}{A_1(2 + 2\epsilon)K + A_2 M},$$

the following conclusions hold:

(1) There is a non-commutative law λ_δ such that the non-commutative law of $X_\delta^{(n)}$ converges to λ almost surely.

(2) Let X_δ be a d-tuple of non-commutative random variables realizing the law λ_δ. Then we have

$$\chi(X_\delta) = \underline{\chi}(X_\delta) = \chi^*(X_\delta) = \lim_{n\to\infty} \left(\frac{1}{n^2} h(X_\delta^{(n)}) + d \log n \right).$$

(3) Let Z be a free semicircular d-tuple with law σ_1. There exists a tracial W^*-isomorphism $\mathrm{W}^*(X_\delta) \to \mathrm{W}^*(Z)$ that maps $\mathrm{W}^*(X_{\delta,1}, \dots, X_{\delta,k})$ to $\mathrm{W}^*(Z_1, \dots, Z_k)$ for every $k = 1, \dots, d$.

Remark 18.2.2. Although we have not kept track of the constants A_1 and A_2 explicitly, they can be mined from the proof. We did not try to optimize the constants, or even their dependence on d, but we did work to optimize the nature of the dependence of δ on ϵ.

To prove Theorem 18.2.1, we will extend $V_\delta^{(n)}$ to a potential defined everywhere using a cut-off argument. Specifically, we consider $\phi \in C_c^\infty(\mathbb{R})$ with $\phi(t) = t$ for $|t| \leq 2 + \epsilon$ and $|\phi(t)| \leq 2 + 2\epsilon$ everywhere, and set

$$\tilde{V}_\delta^{(n)}(x) = \frac{1}{2}\|x\|_2^2 + \delta W^{(n)}(\phi(x)),$$

where $\phi(x) = (\phi(x_1), \dots, \phi(x_d))$ evaluated through functional calculus. This is now a potential defined everywhere, which is equal to $V_\delta^{(n)}(x)$ when $\|x\|_\infty \leq 2 + \epsilon$. We will show that for the correct choice of ϕ and for δ sufficiently small, $\tilde{V}_\delta^{(n)}$ satisfies Assumption 15.1.1, and hence the associated random variables $\tilde{X}_\delta^{(n)}$ have a large-n limit described by a free Gibbs law. Next, we will show that $\|\tilde{X}_\delta^{(n)}\|_\infty$ is less than $2 + \epsilon$ with high probability provided that δ is small enough. Since $\mu_\delta^{(n)}$ is the truncation to the operator-norm ball of radius $2 + \epsilon$ of the measure $\tilde{\mu}_\delta^{(n)}$ given by $\tilde{V}_\delta^{(n)}$, the asymptotic behavior of $\mu_\delta^{(n)}$ is described by the same free Gibbs law.

The first step is to choose a good cut-off function.

Lemma 18.2.3. *For each $\epsilon > 0$, there exists $\phi \in C_c^\infty(\mathbb{R})$ such that*

(1) $|\phi| \leq 2 + 2\epsilon$,

(2) $|\phi'| \leq 1$,

(3) $\phi(t) = t$ *for* $|t| \leq 2 + \epsilon$,

(4) $2\pi \int |t\widehat{\phi}(t)|\, dt \leq B_1 (2 + 2\epsilon)^{1/2} \epsilon^{-1/2}$,

(5) $(2\pi)^2 \int |t^2 \widehat{\phi}(t)|\, dt \leq B_2 \epsilon^{-1}$,

where B_1 and B_2 are universal constants.

Proof. Let $R = 2 + 2\epsilon$. Define the function

$$\psi_R(xt) = \int_0^t (2\chi_{[-R,R]}(s) - \chi_{[-2R,2R]}(s))\, ds.$$

This is a piecewise linear function which is equal to zero outside $[-2R, 2R]$ and equal to t on $[-R, R]$. Let ρ be a C^∞ probability density supported in $[-1, 1]$ with mean zero, let $\rho_\epsilon(t) = \epsilon^{-1} \rho(\epsilon^{-1} t)$. Then define $\phi = \psi * \rho_\epsilon$.

(1) Note that $|\psi_R| \leq R = 2 + 2\epsilon$, hence $|\phi| = |\psi_R * \rho_\epsilon| \leq 2 + 2\epsilon$.

(2) Note that $|\psi_R'| \leq 1$, hence $|\phi'| = |\psi_R' * \rho_\epsilon| \leq 1$.

(3) Now ρ_ϵ has mean zero and is supported in $[-\epsilon, \epsilon]$, while $\psi(t) = t$ for $|t| \leq 2 + 2\epsilon$. Thus, $\phi(t) = t$ for $|t| \leq 2 + \epsilon$.

(4) Using standard facts about Fourier transforms (see e.g. [Fol99, §8 - 9]),

$$2\pi \int |t\widehat{\phi}(t)|\, dt = \|\widehat{\phi'}\|_{L^1} = \left\| \widehat{\psi_R' * \rho_\epsilon} \right\|_{L^1} = \left\| \widehat{\psi_R'} \widehat{\rho}_\epsilon \right\|_{L^1} \leq \left\| \widehat{\psi_R'} \right\|_{L^2} \|\widehat{\rho}_\epsilon\|_{L^2} = \|\psi_R'\|_{L^2} \|\rho_\epsilon\|_{L^2}.$$

But note that $\|\psi_R'\|_{L^2} = (4R)^{1/2} = 2R^{1/2}$ and by scaling $\|\rho_\epsilon\|_{L^2} = \epsilon^{-1/2} \|\rho_1\|_{L^2}$. Thus, we can take $B_1 = 2\|\rho_1\|_{L^2}$.

(5) A direct computation shows that

$$2\pi it\widehat{\psi_R}(t) = \widehat{\psi'_R}(t) = \frac{1}{2\pi t}(2\sin 2\pi Rt - \sin 4\pi Rt),$$

and hence

$$|(2\pi it)^2\widehat{\psi_R}(t)| = |2\sin 2\pi Rt - \sin 4\pi Rt| \leq 3.$$

Therefore,

$$(2\pi)^2 \int |t^2\widehat{\phi}(t)|\,dt = (2\pi)^2 \int |t^2\widehat{\psi_R}(t)\widehat{\rho}_\epsilon(t)|\,dt \leq 3\|\widehat{\rho}_\epsilon\|_{L^1} = 3\epsilon^{-1}\|\widehat{\rho}_1\|_{L^1},$$

where the last inequality follows from scaling properties. Thus, (5) holds with $B_2 = 3\|\widehat{\rho}_1\|_{L^1}$.
□

Lemma 18.2.4. *Fix $\epsilon > 0$, let $W^{(n)}$ be as in Theorem 18.2.1, and let ϕ be as in Lemma 18.2.3. Define $\tilde{W}^{(n)}(x) = W^{(n)}(\phi(x), \dots, \phi(x))$.*

(1) $(\nabla\tilde{W}^{(n)})_{n\in\mathbb{N}}$ is asymptotically approximable by trace polynomials.

(2) $\|\nabla\tilde{W}^{(n)}\|_{\mathrm{Lip}} \leq \epsilon^{-1}[C_1(2+2\epsilon)K + C_2 M]$ for some constants C_1 and C_2 which only depend on d.

(3) $|\tau_n(\nabla_{x_j}\tilde{W}^{(n)}(x))| \leq (2+2\epsilon)K + M$.

Proof. (1) Note that the composition $\tilde{W}^{(n)}$ makes sense because $|\phi| \leq 2+2\epsilon$, which implies that $\|\phi(x)\|_\infty \leq 2+2\epsilon$ for $x \in M_n(\mathbb{C})_{\mathrm{sa}}$ by the spectral mapping theorem.

Because of our estimates on $\widehat{\phi}$ from Lemma 18.2.3, we can apply Proposition 18.1.6 to see that ϕ defines an element of $C^2_{\mathrm{nc}}(\mathbb{R})$ (which we continue to denote by ϕ as an abuse of notation).

Let $x \in M_n(\mathbb{C})^d_{\mathrm{sa}}$ with $\|x\|_\infty \leq R$ and let $y \in M_n(\mathbb{C})_{\mathrm{sa}}$. Let us denote $\phi(x) = (\phi(x_1), \dots, \phi(x_d))$. By the chain rule and (18.2), we have

$$\begin{aligned}\left.\frac{d}{dt}\right|_{t=0} \tilde{W}^{(n)}(x_1, \dots, x_j + ty, x_{j+1}, \dots, x_d) &= \left\langle \nabla_{x_j} W^{(n)}(\phi(x)), \mathcal{D}\phi(x_j)\#y \right\rangle_2 \\ &= \tau_n[\nabla_{x_j} W^{(n)}(x)(\mathcal{D}\phi(x_j)\#y)].\end{aligned}$$

Now $\mathcal{D}\phi$ is an element of $C([-R,R])\widehat{\otimes}C([-R,R])$ for each R, and we claim that $\mathcal{D}\phi$ is invariant under the tensor flip $f \otimes g \mapsto g \otimes f$. By density, it suffices to check this when ϕ is a polynomial, and by linearity, it suffices to check it when ϕ is a monomial. However, we computed this in the proof of Proposition 18.1.6, and it is clearly flip-invariant. Because of flip-invariance and traciality,

$$\tau_n\left[\nabla_{x_j}\tilde{W}^{(n)}(x)(\mathcal{D}\phi(x_j)\#y)\right] = \tau_n[(\mathcal{D}\phi(x_j)\#\nabla_{x_j}W^{(n)}(\phi(x)))y].$$

Since this holds for all y,

$$\nabla_{x_j} \tilde{W}^{(n)}(x) = \mathcal{D}\phi(x_j) \# \nabla_{x_j} W^{(n)}(\phi(x)).$$

Now fix $R > 0$. By definition of $C^2_{\mathrm{nc}}(\mathbb{R})$, there is a sequence of polynomials p_k such that $\mathcal{D}p_k \to \mathcal{D}\phi$ in $C([-R,R]) \widehat{\otimes} C([-R,R])$. Since $\|\phi\|_{C([-R,R])} \leq 2 + 2\epsilon$, by making a small adjustment by scaling, we may assume that $\|p_k\|_{C([-R,R])} \leq 2 + 2\epsilon$. Now

$$\mathcal{D}p_k(x_j) \# W^{(n)}(\phi(x))$$

is asymptotically approximable by trace polynomials as $n \to \infty$ using a variant of Proposition 13.5.12. Indeed, $(\nabla W^{(n)})_{n \in \mathbb{N}}$ is asymptotically approximable on the $\|\cdot\|_\infty$-ball of radius $2+2\epsilon$, and ϕ and p_k map the $\|\cdot\|_\infty$-ball of radius R into the $\|\cdot\|_\infty$-ball of radius $2 + 2\epsilon$, so the argument of Proposition 13.5.12 shows that $(\nabla W^{(n)} \circ \phi)_{n \in \mathbb{N}}$ is asymptotically approximable. By applying $\mathcal{D}p_k(x_j)\#$ just amounts to multiplying by polynomials of x_j on the left and right and taking linear combinations, which will obviously preserve asymptotic approximability.

Therefore, for each k, the sequence $\mathcal{D}p_k(x_j, x_j) \# W^{(n)}(\phi(x))$ is asymptotically approximable by trace polynomials as $n \to \infty$. Next, note that for $x \in M_n(\mathbb{C})^d_{\mathrm{sa}}$ with $\|x\|_\infty \leq R$,

$$\begin{aligned} &\left\| \mathcal{D}p_k(x_j, x_j) \# W^{(n)}(\phi(x)) - \mathcal{D}\phi(x_j, x_j) \# W^{(n)}(\phi(x)) \right\|_2 \\ &\qquad \leq \|\mathcal{D}p_k - \mathcal{D}\phi\|_{C([-R,R]) \widehat{\otimes} C([-R,R])} \|W^{(n)}(\phi(x))\|_2. \end{aligned}$$

Because $\mathcal{D}p_k$ converges to $\mathcal{D}\phi$ in $C([-R,R]) \widehat{\otimes} C([-R,R])$, we know $\mathcal{D}p_k(x_j, x_j) \# W^{(n)}(\phi(x))$ converges to $\mathcal{D}\phi(x_j, x_j) \# W^{(n)}(\phi(x))$ as $k \to \infty$ uniformly on the operator norm ball with the rate of convergence independent of n. Thus, by Lemma 13.5.13, $(\nabla \tilde{W}^{(n)})_{n \in \mathbb{N}}$ is asymptotically approximable by trace polynomials.

(2) Let us write for shorthand

$$\mathcal{D}\phi(x) \# \nabla W^{(n)}(\phi(x)) = \big(\mathcal{D}\phi(x_1) \# \nabla_{x_j} W^{(n)}(\phi(x)), \dots, \mathcal{D}\phi(x_d) \# \nabla_{x_d} W^{(n)}(\phi(x)) \big).$$

Fix R, and let $\|x\|_\infty$, $\|x'\|_\infty \leq R$. Then we write

$$\begin{aligned} &\mathcal{D}\phi(x) \# \nabla W^{(n)}(\phi(x)) - \mathcal{D}\phi(x') \# \nabla W^{(n)}(\phi(x')) \\ &\quad = (\mathcal{D}\phi(x) - \mathcal{D}\phi(x')) \# \nabla W^{(n)}(\phi(x)) + \mathcal{D}\phi(x') \# (\nabla W^{(n)}(\phi(x)) - \nabla W^{(n)}(\phi(x'))). \end{aligned} \tag{18.5}$$

To estimate the first term on the right hand side of (18.5), we apply (18.4) to obtain that

$$\left\| (\mathcal{D}\phi(x_j) - \mathcal{D}\phi(x_j')) \# \nabla_{x_j} W^{(n)}(\phi(x)) \right\|_2 \leq \|\mathcal{D}^2 \phi\|_{C([-R,R])^{\widehat{\otimes} 3}} \|\nabla_{x_j} W^{(n)}(\phi(x))\|_\infty \|x_j - x_j'\|_2.$$

By Proposition 18.1.5 and Lemma 18.2.3 (5),

$$\|\mathcal{D}^2 \phi\|_{C([-R,R])^{\widehat{\otimes} 3}} \leq (2\pi)^2 \int_{\mathbb{R}} |t^2 \widehat{\phi}(t)| \, dt \leq \frac{B_2}{\epsilon}$$

for a universal constant B_2. By Lemma 11.5.4, since $\nabla_{x_j} W^{(n)}$ is K-Lipschitz, we have

$$\|\nabla_{x_j} W^{(n)}(y) - \tau_n(\nabla_{x_j} W^{(n)}(y))\|_\infty \leq 2\sqrt{2} d^{1/2} \Theta K(2+2\epsilon) \text{ for } \|y\|_\infty \leq 2+2\epsilon.$$

Furthermore,

$$|\tau_n(\nabla_{x_j} W^{(n)}(y))| \leq \left\|\nabla_{x_j} W^{(n)}(y)\right\|_2 \leq K\|y\|_2 + \|\nabla_{x_j} W^{(n)}(0)\|_2 \leq K(2+2\epsilon) + M.$$

Therefore,

$$\|\nabla_{x_j} W^{(n)}(y)\|_\infty \leq (1 + 2\sqrt{2} d^{1/2}\Theta)(2+2\epsilon)K + M.$$

And therefore,

$$\begin{aligned}
&\left\|(\mathcal{D}\phi(x) - \mathcal{D}\phi(x'))\#\nabla W^{(n)}(\phi(x))\right\|_2 \\
&= \left(\sum_{j=1}^d \left\|(\mathcal{D}\phi(x_j) - \mathcal{D}\phi(x_j'))\#\nabla_{x_j} W^{(n)}(\phi(x))\right\|_2^2\right)^{1/2} \\
&\leq \left(\sum_{j=1}^d \|\mathcal{D}^2\phi\|_{C([-R,R])^{\hat{\otimes}3}}^2 \|\nabla_{x_j} W^{(n)}(\phi(x))\|_\infty^2 \|x_j - x_j'\|_2^2\right)^{1/2} \\
&\leq \frac{B_2}{\epsilon}\left[\left(1 + 2\sqrt{2} d^{1/2}\Theta\right)(2+2\epsilon)K + M\right]\|x - x'\|_2
\end{aligned}$$

To estimate the second term on the right-hand side of (18.5), note that

$$\begin{aligned}
&\left\|\mathcal{D}\phi(x_j')\#(\nabla_{x_j} W^{(n)}(\phi(x)) - \nabla_{x_j} W^{(n)}(\phi(x')))\right\|_2 \\
&\qquad\qquad \leq \|\mathcal{D}\phi\|_{C([-R,R])^{\hat{\otimes}2}} \left\|\nabla_{x_j} W^{(n)}(\phi(x)) - \nabla_{x_j} W^{(n)}(\phi(x'))\right\|_2.
\end{aligned}$$

Moreover, by (18.3),

$$\|\phi(x) - \phi(x')\|_2 \leq \|\mathcal{D}\phi\|_{C([-R,R])^{\hat{\otimes}2}} \|x - x'\|_2.$$

Therefore, since $\|\nabla W^{(n)}\|_{\text{Lip}} \leq K$, we have

$$\begin{aligned}
&\left\|\mathcal{D}\phi(x')\#(\nabla W^{(n)}(\phi(x)) - \nabla W^{(n)}(\phi(x')))\right\|_2 \\
&= \left(\sum_{j=1}^d \left\|\mathcal{D}\phi(x_j')\#(\nabla_{x_j} W^{(n)}(\phi(x)) - \nabla_{x_j} W^{(n)}(\phi(x')))\right\|_2^2\right)^{1/2} \\
&\leq \|\mathcal{D}\phi\|_{C([-R,R])^{\hat{\otimes}2}} \left\|\nabla W^{(n)}(\phi(x)) - \nabla W^{(n)}(\phi(x'))\right\|_2 \\
&\leq \|\mathcal{D}\phi\|_{C([-R,R])^{\hat{\otimes}2}} K \left(\sum_{j=1}^d \|\phi(x_j) - \phi(x_j')\|_2^2\right)^{1/2} \\
&\leq \|\mathcal{D}\phi\|_{C([-R,R])^{\hat{\otimes}2}}^2 K \|x - x'\|_2 \\
&\leq \frac{B_1^2}{\epsilon}(2+2\epsilon)K\|x - x'\|_2,
\end{aligned}$$

where the last inequality follows from Lemma 18.2.3 (4).

Altogether, putting in the estimates for both terms of (18.5), we obtain

$$\begin{aligned}\left\|\mathcal{D}\phi(x)\#\nabla W^{(n)}(\phi(x)) - \mathcal{D}\phi(x')\#\nabla W^{(n)}(\phi(x'))\right\|_2 \\ \leq \frac{1}{\epsilon}\left(\left[B_2\left(1+2\sqrt{2}d^{1/2}\Theta\right) + B_1^2\right](2+2\epsilon)K + B_2 M\right)\|x-x'\|_2,\end{aligned}$$

which is a Lipschitz bound for $\nabla\tilde{W}^{(n)}$ of the desired form.

(3) Let $\mathfrak{m} : C([-R,R])\widehat{\otimes}C([-R,R])$ denote the flipped multiplication map (which is the same as the usual multiplication map since $C([-R,R])$ is commutative). Then we have

$$\tau_n(\nabla_{x_j}\tilde{W}^{(n)}(x)) = \tau_n(\mathcal{D}\phi(x_j)\nabla_{x_j}W^{(n)}(\phi(x))) = \tau_n\left(\mathfrak{m}(\mathcal{D}\phi)(x_j)\nabla_{x_j}W^{(n)}(\phi(x))\right).$$

Now we claim that $\mathfrak{m}(\mathcal{D}\phi)(x_j) = \phi'(x_j)$. This is true by direct computation when ϕ is a monomial, hence by linearity when ϕ is a polynomial. As remarked earlier, the multiplication map $C([-R,R])\widehat{\otimes}C([-R,R]) \to C([-R,R])$ is continuous. Thus, if $(\phi_k)_{k\in\mathbb{N}}$ is a sequence of polynomials converging to ϕ in $C^1_{\mathrm{nc}}(\mathbb{R})$, then $\phi_k' \to \phi'$ uniformly on compact sets. Hence, the equality $\mathfrak{m}(\mathcal{D}\phi) = \phi'$ extends to all of $C^1_{\mathrm{nc}}(\mathbb{R})$. In particular,

$$\tau_n(\nabla\tilde{W}^{(n)}(x)) = \tau_n\left(\phi'(x_j)\nabla_{x_j}W^{(n)}(\phi(x))\right).$$

By Lemma 18.2.3, $|\phi'| \leq 1$, so by the spectral mapping theorem, $\|\phi'(x_j)\|_\infty \leq 1$. On the other hand, for $\|y\|_\infty \leq 2+2\epsilon$, we have

$$\left\|\nabla_{x_j}W^{(n)}(y)\right\|_2 \leq K\|y\|_2 + \left\|\nabla_{x_j}W^{(n)}(0)\right\|_2 \leq (2+2\epsilon)K + M.$$

Thus, by Cauchy-Schwarz, we get $|\tau_n(\nabla\tilde{W}^{(n)}(x))| \leq (2+2\epsilon)K + M$. □

Corollary 18.2.5. *Let $\tilde{W}^{(n)}$ and C_1 and C_2 be as in the previous lemma. If*

$$\delta < \frac{\epsilon}{C_1(2+2\epsilon) + C_2 M},$$

then the potential $\tilde{V}_\delta^{(n)}(x) = (1/2)\|x\|_2^2 + \delta\tilde{W}^{(n)}(x)$ satisfies Assumption 15.1.1 with

$$\left(1 - \frac{\delta}{\epsilon}[C_1(2+2\epsilon)K + C_2 M]\right) I \leq H\tilde{V}_\delta^{(n)} \leq \left(1 + \frac{\delta}{\epsilon}[C_1(2+2\epsilon)K + C_2 M]\right) I.$$

Hence, letting $\tilde{\mu}_\delta^{(n)}$ be the measure on $M_n(\mathbb{C})_{\mathrm{sa}}^d$ given by the potential $\tilde{V}_\delta^{(n)}$ and $\tilde{X}_\delta^{(n)}$ be the corresponding random variable, Theorem 15.1.5 furnishes a d-tuple of self-adjoint non-commutative random variables X_δ such that $\tilde{X}_\delta^{(n)}$ converges in non-commutative law to X_δ almost surely.

The next stage of the proof is to estimate $\|X_\delta^{(n)}\|_\infty$ with high probability. Using Lemma 11.5.2, we could obtain an asymptotic bound of $c^{-1/2}\Theta$, where $c = \left(1 - \frac{\delta}{\epsilon}[C_1(2+2\epsilon)K + C_2 M]\right)$ is the lower bound on the Hessian. However, we want this Θ to be replaced by 2.

For this reason, we will use a sharper operator-norm bound than Lemma 11.5.2 (although it is less general since it only estimates $\|X\|_\infty$ rather than $\|f(X)\|_\infty$ for arbitrary unitarily equivariant and Lipschitz f). This estimate was shown in the proof of [GM06, Theorem 3.4], and it is based on a convex/log-concave correlation inequality due to Hargé [Har04, Theorem 1.1].

Theorem 18.2.6 ([Har04, Theorem 1.1])**.** *Let $V : \mathbb{R}^d \to \mathbb{R}$ with $HV \geq cI$ and let $d\mu(x) = (1/\int e^{-V})e^{-V(x)}\,dx$, and a be the mean $a = \int x\,d\mu(x)$. Let $\gamma_{c^{-1}}$ be the Gaussian measure with density $\text{const}e^{-c\|x\|^2/2}$. If $f : \mathbb{R}^d \to \mathbb{R}$ is convex, then*

$$\int f(x-a)\,d\mu(x) \leq \int f(y)\,d\gamma_{c^{-1}}(y).$$

Proposition 18.2.7 ([GM06])**.** *Let $V^{(n)} : M_n(\mathbb{C})_{\text{sa}}^d \to \mathbb{R}$ with $HV^{(n)} \geq cI$, let $\mu^{(n)}$ be the associated measure, and let $X^{(n)}$ be the associated random variable. Then*

$$\limsup_{n\to\infty} E\|X^{(n)} - E(X^{(n)})\|_\infty \leq 2c^{-1/2}, \tag{18.6}$$

and

$$\limsup_{n\to\infty} \|X^{(n)} - E(X^{(n)})\|_\infty \leq 2c^{-1/2} \text{ almost surely.} \tag{18.7}$$

Proof. Note that $\|\cdot\|_\infty$ is a convex function on $M_n(\mathbb{C})_{\text{sa}}$. Therefore, by the previous theorem,

$$E\left\|X^{(n)} - E(X^{(n)})\right\|_\infty \leq E\left\|Z^{(n)}\right\|_\infty,$$

where $Z^{(n)}$ is a Gaussian random matrix tuple with distribution $\sigma_{c^{-1}}^{(n)}$. Therefore, to prove the first claim, it suffices to show that

$$\limsup_{n\to\infty} E\left\|Z^{(n)}\right\|_\infty \leq 2c^{-1/2}.$$

This is a standard result in random matrix theory; see for instance the proof of [AGZ09, Theorem 2.1.22]. For the second claim, we apply Herbst's concentration inequality as in the proof of Lemma 11.5.2 to conclude that

$$P\left(\left\|X^{(n)} - E(X^{(n)})\right\|_\infty \geq E\left\|X^{(n)} - E(X^{(n)})\right\|_\infty \geq \delta\right) \leq e^{-cn\delta^2/2}.$$

This estimate together with (18.6) implies (18.7) using a standard argument from probability theory with the Borel-Cantelli lemma. ∎

Lemma 18.2.8. *Continue with the setup from Theorem 18.2.1, Lemma 18.2.3, Lemma 18.2.4, and Corollary 18.2.5. Suppose that*

$$\delta < \frac{\epsilon}{C_1(2+2\epsilon)K + C_2 M}.$$

Then

$$\limsup_{n\to\infty} \left\| \tilde{X}_\delta^{(n)} \right\|_\infty \leq \delta[(2+2\epsilon)K + M] + \frac{2}{\sqrt{1-\delta\epsilon^{-1}[C_1(2+2\epsilon)K + C_2 M]}}.$$

Proof. In light of Corollary 18.2.5 and (18.7), we have

$$\limsup_{n\to\infty} \left\| \tilde{X}_\delta^{(n)} - E(\tilde{X}_\delta^{(n)}) \right\|_\infty \leq \frac{2}{\sqrt{1-\delta\epsilon^{-1}[C_1(2+2\epsilon)K + C_2 M]}} \text{ almost surely.}$$

Thus, it suffices to show that

$$\left\| E(\tilde{X}_\delta^{(n)}) \right\|_\infty \leq \delta[(2+2\epsilon)K + M]$$

We have from Corollary 11.2.6

$$0 = E\left[\nabla \tilde{V}_\delta^{(n)}(\tilde{X}_\delta^{(n)})\right] = E\left[\tilde{X}_\delta^{(n)} + \delta \nabla \tilde{W}_\delta^{(n)}(\tilde{X}_\delta^{(n)})\right].$$

Therefore,

$$E\left[\tilde{X}_{\delta,j}^{(n)}\right] = -\delta E\left[\nabla_{x_j} \tilde{W}_\delta^{(n)}(\tilde{X}_\delta^{(n)})\right] = -\delta E \circ \tau_n \left[\nabla_{x_j} \tilde{W}_\delta^{(n)}(\tilde{X}_\delta^{(n)})\right],$$

using unitary equivariance. But we showed in Lemma 18.2.4 (3) that $|\tau_n(\nabla_{x_j} \tilde{W}(x))| \leq (2+2\epsilon)K + M$. □

Proof of Theorem 18.2.1. Continue with all the notation from above. Suppose that

$$\delta < \frac{\epsilon}{2(C_1(2+2\epsilon)K + C_2 M)}.$$

Using convexity, we have that

$$(1-t)^{-1/2} - 1 \leq 2\left(\sqrt{2}-1\right)t \leq t \text{ for } t \in [0,1/2].$$

Therefore,

$$\begin{aligned} &\delta[(2+2\epsilon)K + M] + \frac{2}{\sqrt{1-\delta\epsilon^{-1}[C_1(2+2\epsilon)K + C_2 M]}} \\ &\qquad \leq 2 + \delta\left[(2+2\epsilon)K + M + 2\epsilon^{-1}[C_1(2+2\epsilon)K + C_2 M]\right]. \end{aligned}$$

In order to guarantee that this is less than $2+\epsilon$, it suffices to choose

$$\delta < \frac{\epsilon}{(2+2\epsilon)K + M + 2\epsilon^{-1}[C_1(2+2\epsilon)K + C_2 M]} = \frac{\epsilon^2}{(2C_1+\epsilon)(2+2\epsilon)K + (2C_2+\epsilon)M}.$$

Both our conditions on δ will be met if we guarantee that

$$\delta < \frac{\min(\epsilon, \epsilon^2)}{A_1(2+2\epsilon)K + A_2 M},$$

where $A_1 = 1 + \max(1, 2C_1)$ and $A_2 = 1 + \max(1, 2C_2)$; this claim is checked directly using the cases $\epsilon \leq 1$ and $\epsilon \geq 1$.

Since $\tilde{V}_\delta^{(n)}$ satisfies Assumption 15.1.1, Theorem 15.1.5 implies that $\tilde{X}_\delta^{(n)}$ converges in non-commutative law almost surely to some X_δ. We also have that almost surely $\limsup_{n\to\infty} \|\tilde{X}_\delta\|_\infty < 2+\epsilon$. Note that $\mu_\delta^{(n)}$ is the truncation of $\mu^{(n)}$ to the $\|\cdot\|_\infty$-ball of radius $2+\epsilon$. Therefore, the random variable $X_\delta^{(n)}$ given by $\mu_\delta^{(n)}$ also converges almost surely in non-commutative law to X_δ, hence (1) of Theorem 18.2.1 is proved.

(2) It follows from Theorem 16.4.1 and Corollary 16.4.2 that

$$\chi(X_\delta) = \underline{\chi}(X_\delta) = \chi^*(X_\delta) = \lim_{n\to\infty} \left(\frac{1}{n^2} h(\tilde{X}_\delta^{(n)}) + d \log n \right).$$

But we also have

$$\lim_{n\to\infty} \frac{1}{n^2} \left(h(\tilde{X}_\delta^{(n)}) - h(X_\delta^{(n)}) \right) = 0.$$

This follows from the truncation arguments done in the proof of Proposition 16.1.4 together with the fact that $\limsup_{n\to\infty} \|\tilde{X}_\delta^{(n)}\|_\infty < 2+\epsilon$ almost surely. We leave the details to the reader.

(3) This follows from Theorem 17.1.9 applied to $\tilde{V}_\delta^{(n)}$. □

Remark 18.2.9. Theorem 18.2.1 can be generalized by replacing the starting potential $(1/2)\|x\|_2^2$ with some other chosen $V_0^{(n)}(x)$ satisfying Assumption 15.1.1 for some $0 < c \leq C$. Then δ has to be chosen small enough that $\delta \|\nabla \tilde{W}^{(n)}\|_{\mathrm{Lip}} < c$ in order to ensure that $V_0^{(n)} + \delta \tilde{W}^{(n)}$ is uniformly convex, while in the quadratic case we have studied, $c = 1$.

18.3 Perturbations of a semicircular family by transport

Our next theorem is an application of triangular transport that can be stated in purely W*-algebraic terms. It says that for a free semicircular family $Z = (Z_1, \dots, Z_d)$, the isomorphism class of the sequence of inclusions

$$\mathrm{W}^*(Z_1) \subseteq \mathrm{W}^*(Z_1, Z_2) \subseteq \cdots \subseteq \mathrm{W}^*(Z_1, \dots, Z_d)$$

is stable under small polynomial perturbations of the generators $Z_1, \dots, Z_d$.

Theorem 18.3.1. *Let $Z = (Z_1, \dots, Z_d)$ be a d-tuple of standard freely independent semicirculars, and let $f = (f_1, \dots, f_d)$ be a d-tuple of self-adjoint non-commutative polynomials. Then for sufficiently small δ, there is an automorphism α of $\mathrm{W}^*(Z_1, \dots, Z_d)$ such that*

$$\alpha(\mathrm{W}^*(Z_1, \dots, Z_k)) = \mathrm{W}^*(Z_1 + \delta f_1(Z), \dots, Z_k + \delta f_k(Z)) \text{ for } k = 1, \dots, d.$$

The idea of the proof is to construct random matrix models $(\mathrm{id} + \delta p)(Z)$ to which we can apply Theorem 18.2.1. The canonical random matrix models for Z are those given by the Gaussian measure $d\sigma_1^{(n)}(x) = (2\pi n^2)^{-dn^2/2} e^{-n^2 \|x\|_2^2/2}\, dx$. After we truncate to a suitable domain where $\mathrm{id} + \delta p$ is an invertible function, the push-forward by $\mathrm{id} + \delta p$ should have density

$$\begin{aligned}\frac{1}{(2\pi n^2)^{dn^2/2}} & e^{-n^2 \|(\mathrm{id} + \delta f)^{-1}(x)\|_2^2/2} |\det J(\mathrm{id} + \delta f)^{-1}(x)| \\ &= \frac{1}{(2\pi n^2)^{dn^2/2}} \exp\left(-n^2 \left(\frac{1}{2}\|(\mathrm{id} + \delta f)^{-1}(x)\|_2^2 - \frac{1}{n^2} \operatorname{Tr}\log J(\mathrm{id} + \delta f)^{-1}(x)\right)\right).\end{aligned}$$

We will first arrange that $(\mathrm{id} + \delta p)^{-1}$ is given by a non-commutative power series in a certain radius, and then we will compute the function in the exponent (and its derivatives) in terms of free difference quotients and cyclic derivatives of $(\mathrm{id} + \delta p)^{-1}$, and thus show that it can be approximated by trace polynomials.

We remark that the theorem and approach should work in much greater generality. Like Theorem 18.2.1, the semicirculars can be replaced by another non-commutative d-tuple arising from random matrix models satisfying Assumption 15.1.1. Moreover, in principle there is no need to restrict to power series, and we conjecture that the result is true for "trace C^3 functions" with the range of values of δ only depending on the first three derivatives. However, the computation of the log of the Jacobian for trace C^k functions (and even trace polynomials) is more complicated and thus requires more preparation than we have time to undertake here. Thus, we will be content for the present with power series methods.

Let $\mathcal{P}$ be the basis for TrP_d consisting of functions of the form

$$p = \operatorname{tr}(p_1) \dots \operatorname{tr}(p_\ell) p_0,$$

where $p_1, \dots, p_\ell$ are monomials considered up to cyclic equivalence and p_0 is a monomial in $\mathbb{C}\langle X_1, \dots, X_d \rangle$. Given a formal series of the form

$$f = \sum_{p \in \mathcal{P}} a_p p, \text{ where } a_p \in \mathbb{C},$$

we define

$$\|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R)} = \sum_{p \in \mathcal{P}} |a_p| R^{\deg(p)},$$

where deg(p) is the total degree of p (the grading given in Definition 13.1.4). We define

$$\mathcal{S}_{\mathrm{tr}}(d,R) = \{f : \|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R)} < \infty\}.$$

We also denote by $\mathcal{S}(R)$ the subspace consisting of power series in only the non-commutative monomials, with no trace terms. We leave the verification of the next observation as an (easy) exercise.

Observation 18.3.2. *Define addition, multiplication, and the $*$-operations for $\mathcal{S}_{\mathrm{tr}}(d,R)$ in the same way as for* TrP_d*. Then $\mathcal{S}_{\mathrm{tr}}(d,R)$ is a Banach $*$-algebra and $\mathcal{S}(d,R)$ is a Banach $*$-subalgebra. Also, $\|\mathrm{tr}(f)\|_{\mathcal{S}_{\mathrm{tr}}(d,R)} \leq \|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R)}$.*

Composition of formal trace power series is defined just like the composition for trace polynomials in Definition 13.4.1. The power series norms behave as follows under composition. The norm of a d-tuple $g = (g_1,\dots,g_d) \in \mathcal{S}_{\mathrm{tr}}(d,R)^d$ is defined as follows. If $g_j = \sum_{p \in \mathcal{P}} a_{p,j} p$, then we define

$$\|g\|_{\mathcal{S}_{\mathrm{tr}}(d,R)^d} := \max_j \|g_j\|_{\mathcal{S}_{\mathrm{tr}}(d,R)}.$$

Lemma 18.3.3.

(1) If $f \in \mathcal{S}_{\mathrm{tr}}(d,R_1)^{d_1}$ and $g \in \mathcal{S}(d,R)^d$ with $\|g\|_{\mathcal{S}_{\mathrm{tr}}(d,R)^d} \leq R_1$, then $\|f \circ g\|_{\mathcal{S}_{\mathrm{tr}}(d,R)} \leq \|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)}$.

(2) Suppose $R_1 < R_2$ and $f \in \mathcal{S}_{\mathrm{tr}}(d,R_1)^{d_1}$ and $g,h \in \mathcal{S}_{\mathrm{tr}}(d,R)$ with $\|g\|_{\mathcal{S}_{\mathrm{tr}}(d,R)^d} \leq R_1$ and $\|h\|_{\mathcal{S}_{\mathrm{tr}}(d,R)^d} \leq R_1$. Then

$$\|f \circ g - f \circ h\|_{\mathcal{S}_{\mathrm{tr}}(d,R)} \leq \frac{1}{R_1 e \log(R_2/R_1)} \|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_2)} \|g-h\|_{\mathcal{S}_{\mathrm{tr}}(d,R)^d}.$$

Proof. (1) Let us write $f = \sum_{p \in \mathcal{P}} a_p p$, where a_p is the vector $(a_{p,1},\dots,a_{p,d_1})$. Using Observation 18.3.2, we have for $p \in \mathcal{P}$ that

$$\|p(g)\|_{\mathcal{S}_{\mathrm{tr}}(d,R)} \leq R_1^{\deg(p)}.$$

Thus,

$$\|f(g)\|_{\mathcal{S}_{\mathrm{tr}}(d,R)} \leq \sum_{p \in \mathcal{P}} \|a_p\| R_1^{\deg(p)} = \|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)}.$$

(2) Given a monomial $p(X) = X_{i(1)} \dots X_{i(k)}$, we have

$$p(g) - p(h) = \sum_{j=1}^k g_{i(1)} \dots g_{i(j-1)} (g_{i(j)} - h_{i(j)}) h_{i(j+1)} \dots h_{i(k)},$$

so that

$$\|p(g) - p(h)\|_{\mathcal{S}_{\mathrm{tr}}(d,R)} \leq k R_1^{k-1} \|g-h\|_{\mathcal{S}_{\mathrm{tr}}(d,R)}$$

A similar argument applies to $p = \operatorname{tr}(p_1)\dots\operatorname{tr}(p_\ell)p_0$, showing that

$$\|p(g) - p(h)\|_{\mathcal{S}_{\mathrm{tr}}(d,R)} \leq \deg(p) R_1^{\deg(p)-1} \|g - h\|_{\mathcal{S}_{\mathrm{tr}}(d,R)}.$$

An elementary computation shows that function $tR_1^{t-1}/R_2^t = R_1^{-1}t(R_1/R_2)^t$ is maximized when $t = 1/\log(R_2/R_1)$, resulting in

$$\frac{1}{R_1} t \left(\frac{R_1}{R_2}\right)^t \leq \frac{1}{R_1 e \log(R_2/R_1)}.$$

Hence,

$$\begin{aligned}
\|f(g) - f(h)\|_{\mathcal{S}_{\mathrm{tr}}(d,R)} &\leq \sum_{p \in \mathcal{P}} \|a_p\| \deg(p) R_1^{\deg(p)-1} \|g - h\|_{\mathcal{S}_{\mathrm{tr}}(d,R)} \\
&\leq \frac{1}{R_1 e \log(R_2/R_1)} \sum_{p \in \mathcal{P}} \|a_p\| R_2^{\deg(p)} \|g - h\|_{\mathcal{S}_{\mathrm{tr}}(d,R)} \\
&= \frac{1}{R_1 e \log(R_2/R_1)} \|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_2)} \|g - h\|_{\mathcal{S}_{\mathrm{tr}}(d,R)}. \qquad \square
\end{aligned}$$

Now as the first step to proving Theorem 18.3.1, we give an easy perturbative inverse function theorem for $\mathcal{S}_{\mathrm{tr}}(d,R)$.

Lemma 18.3.4. *Let $R_1 < R_2$, and suppose that $f \in \mathcal{S}_{\mathrm{tr}}(d,R_2)^d$ with*

$$\|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_2)^d} < \min\left(R_2 - R_1, R_1 e \log(R_2/R_1)\right).$$

Then there exists $g \in \mathcal{S}_{\mathrm{tr}}(d,R_1)^d$ with $(\mathrm{id} + f) \circ (\mathrm{id} - g) = \mathrm{id}$ and

$$\|g\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d} \leq \|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d}.$$

Moreover, if f_1, ..., f_d are self-adjoint, then so are g_1, ..., g_d, and if f_1, ..., f_d are in $\mathcal{S}(d,R_2)^d$, then g_1, ..., g_d are in $\mathcal{S}(d,R_1)^d$.

Proof. Consider the map $\Phi : g \mapsto f \circ (\mathrm{id} - g)$ defined for any formal trace power series g. We claim that Φ is a contraction mapping from $\{g : \|g\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d} \leq R_2 - R_1\}$ into itself. If $\|g\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d} \leq R_2 - R_1$, then $\|\mathrm{id} - g\|_{\mathcal{S}_{|tr}(d,R_1)^d} \leq R_1 + (R_2 - R_1) = R_2$, and hence by Lemma 18.3.3 (1),

$$\|f \circ (\mathrm{id} - g)\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d} \leq \|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_2)^d} \leq R_2 - R_1.$$

Moreover, for g, h with $\|g\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d} \leq R_2 - R_1$ and $\|h\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d} \leq R_2 - R_1$, we have by Lemma 18.3.3 (2) that

$$\|f \circ (\mathrm{id} - g) - f \circ (\mathrm{id} - h)\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d} \leq \frac{\|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_2)^d}}{R_1 e \log(R_2/R_1)} \|g - h\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d},$$

and we assumed that $\|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_2)} < R_1 e \log(R_2/R_1)$, hence Φ is a contraction mapping.

Thus, by the Banach fixed-point theorem, there is a unique g with $\|g\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d} \leq R_2 - R_1$ and $g = f \circ (\mathrm{id} - g)$. Since $g = f \circ (\mathrm{id} - g)$, we get $\|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d} \leq \|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_2)^d}$. Being a fixed point means that $\mathrm{id} = \mathrm{id} - g + f \circ (\mathrm{id} - g) = (\mathrm{id} + f) \circ (\mathrm{id} - g)$.

Finally, the claim about self-adjointness follows because Φ restricts to a mapping on self-adjoint power series g, since self-adjoints are closed under composition. □

A fairly standard argument, similar to the one for Lemma 18.3.3 (2), allows us to estimate the norms of various derivatives associated to the power series. It will be convenient for us to treat the non-commutative polynomial part and the trace part separately. If f is a power series of non-commutative monomials (with no trace part), then we define $\mathcal{D}_{X_j} f$ as the termwise application of the difference quotient operator $\mathcal{D}_{X_j}$, which takes values in the tensor product of two copies of the non-commutative formal power series ring. Similarly, if $f \in \mathcal{S}(d,R)^d$, then let

$$\mathcal{D} f = \begin{pmatrix} \mathcal{D}_{X_1} f_1 & \dots & \mathcal{D}_{X_d} f_1 \\ \vdots & \ddots & \vdots \\ \mathcal{D}_{X_1} f_d & \dots & \mathcal{D}_{X_d} f_d \end{pmatrix}.$$

Lemma 18.3.5. *Let $R_1 < R_2$. If $f \in \mathcal{S}(d,R_2)$, then $\mathcal{D} f$ may be viewed as an element of the projective tensor product $M_d(\mathbb{C}) \widehat{\otimes} \mathcal{S}(d,R_1) \widehat{\otimes} \mathcal{S}(d,R_1)$ (where $M_n(\mathbb{C})$ is equipped with the operator norm), and we have*

$$\|\mathcal{D} f\|_{M_n(\mathbb{C}) \widehat{\otimes} \mathcal{S}(d,R_1) \widehat{\otimes} \mathcal{S}(d,R_1)} \leq \frac{d}{R_1 e \log(R_2/R_1)} \|f\|_{\mathcal{S}(d,R_2)^d}.$$

Proof. Let us write $f_j = \sum_{p \in \mathcal{P}} a_{p,j} p$. Then let $a_p = (a_{p,1}, \dots, a_{p,d})$, so that $f = \sum_{p \in \mathcal{P}} a_p p$. Let $a_p e_j^\dagger$ denote the matrix in $M_n(\mathbb{C})$ with a_p in the jth column and zeroes in the other columns, and note $\|a_p e_j^\dagger\|_\infty = \|a_p\| = \left(\sum_j |a_{p,j}|^2\right)^{1/2} \leq \sum_j |a_{p,j}|$. Then

$$\mathcal{D}(a_p p) = \sum_{j=1}^d (a_p e_j^\dagger) \otimes \mathcal{D}_{X_j} p.$$

Therefore,

$$\|\mathcal{D}(a_p p)\|_{M_d(\mathbb{C}) \widehat{\otimes} \mathcal{S}(d,R_1) \widehat{\otimes} \mathcal{S}(d,R_1)} \leq \|a_p\| \sum_{j=1}^d \|\mathcal{D}_{X_j} p\|_{\mathcal{S}(d,R_1) \widehat{\otimes} \mathcal{S}(d,R_1)}.$$

A straightforward computation shows that for a monomial p,

$$\sum_{j=1}^d \|\mathcal{D}_{X_j} p\|_{\mathcal{S}(d,R_1) \widehat{\otimes} \mathcal{S}(d,R_1)} \leq \deg(p) R_1^{\deg(p)-1}.$$

As we argued above,

$$\deg(p) R_1^{\deg(p)-1} \leq \frac{1}{R_1 e \log(R_2/R_1)} R_2^{\deg(p)}.$$

Thus,

$$\begin{aligned}\sum_{p \in \mathcal{P}} \|\mathcal{D}(a_p p)\|_{M_d(\mathbb{C}) \widehat{\otimes} \mathcal{S}(d,R_1) \widehat{\otimes} \mathcal{S}(d,R_1)} &\leq \frac{1}{R_1 e \log(R_2/R_1)} \sum_{p \in \mathcal{P}} \sum_{j=1}^d |a_{p,j}| R_2^{\deg(p)} \\ &= \frac{1}{R_1 e \log(R_2/R_1)} \sum_{j=1}^d \|f_j\|_{\mathcal{S}(d,R_2)}.\end{aligned}$$

This shows that the series expansion for $\mathcal{D}f$ converges absolutely in $M_d(\mathbb{C}) \widehat{\otimes} \mathcal{S}(d,R_1) \widehat{\otimes} \mathcal{S}(d,R_1)$, and hence $\mathcal{D}f$ makes sense as an element of that space, and the asserted estimate holds. $\square$

Next, let $\mathcal{S}_{\mathrm{tr}}^0(d,R)$ denote the subspace of $\mathcal{S}_{\mathrm{tr}}(d,R)$ consisting of power series in terms of the form $\mathrm{tr}(p_1) \dots \mathrm{tr}(p_\ell)$ with no non-commutative polynomial terms. Let $D_{X_j}^0 : \mathcal{S}_{\mathrm{tr}}^0(d,R) \to \mathcal{S}_{\mathrm{tr}}(d,R)$ the operator defined in (14.2), extended from trace polynomials to trace power series in the obvious way. Let $D^0 f := (D_{X_1}^0 f, \dots, D_{X_d}^0 f)$. The following lemma is proved in much the same way as the previous one, so we leave the details as an exercise.

Lemma 18.3.6. *If $R_1 < R_2$ and if $f \in \mathcal{S}_{\mathrm{tr}}^0(d,R_2)$, then $D^0 f \in \mathcal{S}_{\mathrm{tr}}(d,R_1)^d$ with*

$$\|D^0 f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d} \leq \frac{1}{R_1 e \log(R_2/R_1)} \|f\|_{\mathcal{S}_{\mathrm{tr}}^0(d,R_2)}.$$

If $f \in \mathcal{S}_{\mathrm{tr}}(d,R)$ and $x \in M_n(\mathbb{C})^d$ with $\|x\|_\infty \leq R$, then the evaluation $f(x)$ is defined because the series $\sum_{p \in \mathcal{P}} a_p p(x)$ converges absolutely in $\|\cdot\|_\infty$. Note that for $p \in \mathcal{P}$ and $\|x\|_\infty \leq R_1$, $\|y\|_\infty \leq R_1$, we have

$$\begin{aligned}\|p(x) - p(y)\|_\infty &\leq \deg(p) R_1^{\deg(p)-1} \|x - y\|_\infty \\ \|p(x) - p(y)\|_2 &\leq \deg(p) R_1^{\deg(p)-1} \|x - y\|_2,\end{aligned}$$

by the typical telescoping sum argument for products. This easily implies the following lemma, whose proof we leave as an exercise.

Lemma 18.3.7. *Suppose $R_1 < R_2$ and $f \in \mathcal{S}_{\mathrm{tr}}(d,R_1)^d$ and $x, y \in M_n(\mathbb{C})^d$ with $\|x\|_\infty$, $\|y\|_\infty \leq R_1$. Then*

$$\begin{aligned}\|f(x) - f(y)\|_\infty &\leq \frac{1}{R_1 e \log(R_2/R_1)} \|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d} \|x - y\|_\infty, \\ \|f(x) - f(y)\|_2 &\leq \frac{d}{R_1 e \log(R_2/R_1)} \|f\|_{\mathcal{S}_{\mathrm{tr}}(d,R_1)^d} \|x - y\|_2.\end{aligned}$$

Now we are ready for the main proof. In fact, we will prove a more refined version of the theorem. Clearly, Theorem 18.3.1 follows from Theorem 18.3.8 because if f is a polynomial, then $\|f\|_{\mathcal{S}(d,2+6\epsilon)^d}$ is finite and hence $\|\delta f\|_{\mathcal{S}(d,2+6\epsilon)^d}$ can be made arbitrarily small by choosing δ small enough.

Theorem 18.3.8. *Let $Z = (Z_1, \dots, Z_d)$ be a standard free semicircular family, and let $f \in \mathcal{S}(d, 2+6\epsilon)^d$ be a d-tuple of self-adjoint power series. If $\|f\|_{\mathcal{S}(d,2+6\epsilon)^d}$ is sufficiently small (depending on ϵ and d), then there is an automorphism α of $\mathrm{W}^*(Z_1, \dots, Z_d)$ such that*

$$\alpha(\mathrm{W}^*(Z_1, \dots, Z_k)) = \mathrm{W}^*(Z_1 + \delta f_1(Z), \dots, Z_k + \delta f_k(Z)) \text{ for } k = 1, \dots, d.$$

Proof. For $j = 1, \dots, 6$, let $R_j = 2 + j\epsilon$. Let $B_\infty^{(n)}(R_j)$ denote the ball of radius R_j in $\|\cdot\|_\infty$ in $M_n(\mathbb{C})_{\mathrm{sa}}^d$. Note that $\log(R_{j+1}/R_j)$ is increasing in j, and hence can be bounded by $\log(R_6/R_5)$. Assume that

$$\|f\|_{\mathcal{S}(d,R_6)^d} < \min\left(\epsilon, R_2 e \log \frac{R_6}{R_5}\right).$$

By Lemma 18.3.7, we have

$$\|(\mathrm{id}+f)(x) - (\mathrm{id}+f)(y)\|_\infty \geq \left(1 - \frac{\|f\|_{\mathcal{S}(d,R_6)^d}}{R_5 e \log(R_6/R_5)}\right) \|x - y\|_\infty \text{ for } \|x\|_\infty, \|y\|_\infty \leq R_5.$$

Since $R_2 < R_5$, we have $\|f\|_{\mathcal{S}(d,R_6)^d} < R_5 e \log(R_6/R_5)$, hence the above equation implies that $\mathrm{id}+f$ is injective on $B_\infty^{(n)}(R_5)$.

Next, by Lemma 18.3.4, there is a self-adjoint tuple $g \in \mathcal{S}(d, R_5)^d$ satisfying $(\mathrm{id}+f) \circ (\mathrm{id}-g) = \mathrm{id}$ and

$$\|g\|_{\mathcal{S}(d,R_5)^d} \leq \|f\|_{\mathcal{S}(d,R_6)^d} < \left(\epsilon, R_2 e \log \frac{R_6}{R_5}\right).$$

But note that $\|\mathrm{id}-g\|_{\mathcal{S}(d,R_4)^d} \leq R_4 + \epsilon = R_5$, and hence $\mathrm{id}-g$ maps $B_\infty^{(n)}(R_4)$ into $B_\infty^{(n)}(R_5)$, and thus the image $(\mathrm{id}+f)(B_\infty^{(n)}(R_5))$ contains $B_\infty^{(n)}(R_4)$.

Let $Z^{(n)}$ be a random variable in $M_n(\mathbb{C})_{\mathrm{sa}}^d$ with probability distribution given by

$$\frac{1}{\int_{\|z\|_\infty \leq R_5} 1_{\|z\|_\infty \leq R_5} e^{-n^2\|z\|_2^2/2}} e^{-n^2\|z\|_2^2/2}\, dz,$$

that is, the truncation of a Gaussian random matrix d-tuple to $B^{(n)}(R_5)$. Since $R_5 > 2$, we know that a Gaussian random matrix d-tuple has norm eventually less than R_5 almost surely as $n \to \infty$. Thus, the measure of $B^{(n)}(R_5)$ under the Gaussian distribution tends to 1 as $n \to \infty$. So the random variable $Z^{(n)}$ with the truncated distribution still converges in non-commutative law almost surely to a free semicircular family, and $\lim_{n\to\infty}\|Z^{(n)}\|_\infty = 2$ almost surely.

Let $f^{(n)}$ and $g^{(n)}$ denote the evaluation of f and g on $M_n(\mathbb{C})_{\mathrm{sa}}^d$. Let $Y^{(n)} = (\mathrm{id} + f^{(n)})(Z^{(n)})$. Since $\mathrm{id} + f^{(n)}$ is injective on $B_\infty^{(n)}(R_5)$ and the inverse function is given by $\mathrm{id} - g$, the probability distribution of $Y^{(n)}$ is computed by change of variables as

$$\frac{1}{\int_{\|z\|_\infty \le R_5} e^{-n^2\|z\|_2^2/2}} 1_{y \in (\mathrm{id} + f^{(n)})(B_\infty^{(n)}(R_5))} e^{-n^2\|y - g^{(n)}(y)\|_2^2/2} |\det J(\mathrm{id} - g^{(n)})(y)|\, dy.$$

By Corollary 13.4.4, $Y^{(n)}$ converges in non-commutative law almost surely to $(\mathrm{id} + f)(Z)$. Moreover, since $\limsup_{n\to\infty} \|Z^{(n)}\|_\infty \le 2$ almost surely, we have $\limsup_{n\to\infty} \|Y^{(n)}\|_\infty \le 2 + \|f\|_{\mathcal{S}(d,R_6)^d} < 2 + \epsilon = R_1$ almost surely. This implies that $P(Y^{(n)} \in B_\infty^{(n)}(R_1)) \to 1$ as $n \to \infty$.

Let $X^{(n)}$ be a random variable whose probability distribution is the truncation of the distribution of $Y^{(n)}$ to $B_\infty^{(n)}(R_1)$. That is, $X^{(n)}$ has the probability distribution

$$A^{(n)} 1_{\|x\|_\infty \le R_1} e^{-n^2\|x - g^{(n)}(x)\|_2^2/2} |\det J(\mathrm{id} - g^{(n)})(x)|\, dx,$$

where $A^{(n)}$ is the normalizing constant to make this a probability measure. From the preceding claims about $Y^{(n)}$, it follows that $X^{(n)}$ converges in non-commutative law to Z almost surely. We want to apply Theorem 18.2.1 to $X^{(n)}$ and thus obtain triangular transport for Z. The density of $X^{(n)}$ can be described as $e^{-n^2 V^{(n)}(x)}$, where

$$V^{(n)}(x) = \frac{1}{2}\|x - g^{(n)}(x)\|_2^2 - \frac{1}{n^2} \log |\det J(\mathrm{id} - g^{(n)})(x)|,$$

so we have to estimate the Lipschitz norm of $\nabla(V^{(n)}(x) - (1/2)\|x\|_2^2)$.

Let us consider the log-determinant term first. By Lemma 14.1.3 (or rather its generalization to power series),

$$Jg^{(n)}(x)[y] = \mathcal{D}g(x) \# y,$$

where $\mathcal{D}g(x)\#y$ is given by

$$(\mathcal{D}g(x)\#y)_i = \sum_{j=1}^d \mathcal{D}_{X_j} g(x) \# y_j.$$

It is easy to show that for $\|x\|_\infty \le R_4$,

$$\left\| Jg^{(n)}(x) \right\| \le \|\mathcal{D}g\|_{M_d(\mathbb{C}) \hat{\otimes} \mathcal{S}(d,R_5) \hat{\otimes} \mathcal{S}(d,R_5)}.$$

And by Lemma 18.3.5 and Lemma 18.3.4, we have

$$\|\mathcal{D}g\|_{M_d(\mathbb{C}) \hat{\otimes} \mathcal{S}(d,R_4) \hat{\otimes} \mathcal{S}(d,R_4)} \le \frac{d}{R_4 e \log(R_5/R_4)} \|g\|_{\mathcal{S}(d,R_5)^d} \le \frac{d}{R_4 e \log(R_5/R_4)} \|f\|_{\mathcal{S}(d,R_6)^d}.$$

We are allowed to make $\|f\|_{\mathcal{S}(d,R_6)^d}$ is small as we wish depending on ϵ and d. Thus, we may assume that the right-hand side is less than or equal to $1 - 1/e$. In particular, this implies it is strictly less than 1, which implies convergence of the power series

$$\log(I - Jg^{(n)}(x)) = \sum_{k=1}^\infty \frac{1}{k} Jg^{(n)}(x)^k.$$

It is well known (and easy to demonstrate using the Jordan canonical form) that since $\|Jg^{(n)}(x)\| < 1$, we have $\det(I - Jg^{(n)}(x)) > 0$ and

$$\log\det(I - Jg^{(n)}(x)) = \operatorname{Tr}\log(I - Jg^{(n)}(x)).$$

Let us equip $M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)$ with the "hash multiplication" operation

$$(A_1 \otimes f_1 \otimes g_1)\#(A_2 \otimes f_2 \otimes g_2) = A_1A_2 \otimes f_1f_2 \otimes g_2g_1.$$

Since

$$\begin{aligned}\|(A_1 \otimes f_1 \otimes g_1)\#(A_2 \otimes f_2 \otimes g_2)\|_{M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)} \\ \leq \|A_1\|_{M_d(\mathbb{C})}\|f_1\|_{\mathcal{S}(d,R_4)}\|g_1\|_{\mathcal{S}(d,R_4)}\|A_2\|_{M_d(\mathbb{C})}\|f_2\|_{\mathcal{S}(d,R_4)}\|g_2\|_{\mathcal{S}(d,R_4)},\end{aligned}$$

the universal property of projective tensor products implies that hash multiplication extends to a well-defined map

$$[M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)] \times [M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)] \to M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4),$$

with

$$\|F_1\#F_2\|_{M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)} \leq \|F_1\|_{M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)}\|F_2\|_{M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)}.$$

In particular,

$$\log_{\#}(I - \mathcal{D}g) = \sum_{k=1}^{\infty} \frac{1}{k}(\mathcal{D}g)^{\#k}$$

converges in $M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)$ when $\|\mathcal{D}g\|_{M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)} < 1$, and we have

$$\begin{aligned}\left\|\log_{\#}(I - \mathcal{D}g)\right\|_{M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)} &\leq -\log\left(1 - \|\mathcal{D}g\|_{M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)}\right) \\ &\leq e\|\mathcal{D}g\|_{M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)},\end{aligned}$$

because $-\log(1-t) \leq et$ for $t \in [0, 1 - 1/e]$, and we have arranged that

$$\|\mathcal{D}g\|_{M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)} \leq 1 - 1/e.$$

Of course, for an $n \times n$ matrix tuple x with $\|x\|_\infty \leq R_4$, we have

$$\log(I - Jg^{(n)}(x))[y] = \log_{\#}(I - \mathcal{D}g)\#y.$$

To compute the trace of this matrix, let $\mathcal{B}$ be the orthonormal basis for $M_n(\mathbb{C})_{\mathrm{sa}}$ given by (14.1). Let us view $\log(I - Jg^{(n)}(x))$ as a $d \times d$ block matrix, each block being a linear transformation $M_n(\mathbb{C})_{\mathrm{sa}} \to M_n(\mathbb{C})_{\mathrm{sa}}$, and denote the (j,j) block by $\log(I - Jg^{(n)}(x))_{j,j}$. Then using Lemma 14.1.8 as in the proof of Lemma 14.1.9, we obtain

$$\begin{aligned}\frac{1}{n^2}\operatorname{Tr}\log(I - Jg^{(n)}(x)) &= \sum_{j=1}^d \sum_{b\in\mathcal{B}} \tau_n(b\log(I - Jg^{(n)}(x))_{j,j})[b]) \\ &= \frac{1}{n^2}\sum_{j=1}^d \sum_{b\in\mathcal{B}} \tau_n(b\log(I - \mathcal{D}g)_{j,j}(x)\# b) \\ &= \sum_{j=1}^d \tau_n \otimes \tau_n[\log(I - \mathcal{D}g)_{j,j}(x)] \\ &= \sum_{j=1}^d \operatorname{tr}\otimes\operatorname{tr}[\log(I - \mathcal{D}g)_{j,j}](x) \\ &= [\operatorname{Tr}\otimes\operatorname{tr}\otimes\operatorname{tr}[\log(I - \mathcal{D}g)]](x).\end{aligned}$$

Here $\mathcal{D}g_{j,j}$ stands for the (j,j) entry of $\mathcal{D}g$, viewed as a $d \times d$ matrix with entries in $\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)$. Moreover, tr stands for the map $\mathcal{S}(d,R_4) \to \mathcal{S}_{\mathrm{tr}}^0(d,R_4)$ that applies tr to each monomial p in the power series. Finally, we define $\operatorname{Tr}\otimes\operatorname{tr}\otimes\operatorname{tr} : M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4) \to \mathcal{S}_{\mathrm{tr}}^0(d,R_4)$ by

$$[\operatorname{Tr}\otimes\operatorname{tr}\otimes\operatorname{tr}](A\otimes f\otimes g) = \operatorname{Tr}(A)\operatorname{tr}(f)\operatorname{tr}(g).$$

The upshot is that $(1/n^2)\operatorname{Tr}\log(I - Jg^{(n)}(x))$ is given by the evaluation on $n \times n$ matrices of the power series

$$h := \operatorname{Tr}\otimes\operatorname{tr}\otimes\operatorname{tr}[\log(I - \mathcal{D}g)] \in \mathcal{S}_{\mathrm{tr}}^0(d,R_4).$$

Furthermore, it is straightforward to check that

$$\|h\|_{\mathcal{S}_{\mathrm{tr}}^0(d,R_4)} \le d\big\|\log_{\#}(I - \mathcal{D}g)\big\|_{M_d(\mathbb{C})\widehat{\otimes}\mathcal{S}(d,R_4)\widehat{\otimes}\mathcal{S}(d,R_4)},$$

and due to our previous estimates, this is bounded by a constant (depending on d and ϵ) times $\|f\|_{\mathcal{S}(d,R_6)}$ (assuming that $\|f\|_{\mathcal{S}(d,R_6)}$ is sufficiently small).

Letting $h^{(n)}$ be the evaluation of h on $n \times n$ matrix tuples, we have

$$V^{(n)}(x) = \frac{1}{2}\|x\|_2^2 - \langle x, g^{(n)}(x)\rangle_2 + \frac{1}{2}\|g^{(n)}(x)\|_2^2 - h^{(n)}(x).$$

Let

$$W = -\langle \mathrm{id}, g\rangle_2 + \frac{1}{2}\|g\|_2^2 - h \in \mathcal{S}_{\mathrm{tr}}^0(d,R_4),$$

and let $W^{(n)}$ be its evaluation on matrices. Then $\|g\|_{\mathcal{S}(d,R_4)}$ and $\|h\|_{\mathcal{S}_{\mathrm{tr}}^0(d,R_4)}$ are controlled by a constant times $\|f\|_{\mathcal{S}(d,R_6)}$, hence so is $\|W\|_{\mathcal{S}_{\mathrm{tr}}^0(d,R_4)}$. Note that $\nabla W^{(n)}(x) = D^0 W(x)$ for $x \in M_n(\mathbb{C})_{\mathrm{sa}}^d$ with $\|x\|_\infty < R_4$. By Lemma 18.3.6, $\|D^0 W\|_{\mathcal{S}_{\mathrm{tr}}(d,R_3)^d}$ is bounded by a constant times $\|W\|_{\mathcal{S}_{\mathrm{tr}}(d,R_4)}$. Finally, by Lemma 18.3.7, the Lipschitz norm of $\nabla W^{(n)}$ with respect to $\|\cdot\|_2$ on $B_\infty^{(n)}(R_2)$ is controlled by a constant times $\|D^0 W\|_{\mathcal{S}_{\mathrm{tr}}(d,R_3)^d}$.

By choosing $\|f\|_{\mathcal{S}(d,R_6)}$ small enough, we can make $\nabla W^{(n)}(0)$ and the Lipschitz norm of $\nabla W^{(n)}$ small enough that Theorem 18.2.1 can be applied with $\delta = 1$. This yields an isomorphism

$$\alpha : \mathrm{W}^*(Z) \to \mathrm{W}^*(Z + f(Z)).$$

with the desired triangular property. But since $Z = (\mathrm{id} - g)(\mathrm{id} + f)(Z)$, we see that $\mathrm{W}^*(Z + f(Z)) = \mathrm{W}^*(Z)$. □

REFERENCES

[ABF13] Michael Anshelevich, Serban T. Belinschi, Maxime F'evrier, and Alexandru Nica. "Convolution powers in the operator-valued framework." *Transactions of the American Mathematical Society*, **365**:2063–2097, 2013.

[AGZ09] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.

[AK13] Gulnara Abduvalieva and Dmitry S. Kaliuzhnyi-Verbovetskyi. "Fixed point theorems for noncommutative functions." *Journal of Mathematical Analysis and Applications*, **401**(1):436–446, 2013.

[AK15] Gulnara Abduvalieva and Dmitry S. Kaliuzhnyi-Verbovetskyi. "Implicit/inverse function theorems for free noncommutative functions." *Journal of Functional Analysis*, **269**(9):2813–2844, 2015.

[AM16] Jim Agler and John E. McCarthy. "The implicit function theorem and free algebraic sets." *Transactions of the American Mathematical Society*, **368**:3157–3175, 2016.

[AP10a] A. B. Aleksandrov and V. V. Peller. "Functions of perturbed unbounded self-adjoint operators. Operator Bernstein type inequalities." *Indiana University Mathematics Journal*, **59**:1451–1490, 04 2010.

[AP10b] A. B. Aleksandrov and V. V. Peller. "Operator Hölder-Zygmund functions." *Advances in Mathematics*, **224**(3):910–966, 2010.

[AP17a] A. B. Aleksandrov and V. V. Peller. "Multiple operator integrals, Haagerup and Haagerup-like tensor products, and operator ideals." *Bulletin of the London Mathematical Society*, **49**(3):463–479, 2017.

[AP17b] Claire Antharaman and Sorin Popa. "An Introduction to II1 Factors." Preprint at http://www.math.ucla.edu/ popa/Books/IIun-v10.pdf, 2017.

[AS18] Octavio Arizmendi and Mauricio Salazar. "A Berry-Esseen type limit theorem for Boolean convolution." *Archiv der Mathematik*, **111**(1):101–111, 2018.

[ASW19] Octavio Arizmendi, Mauricio Salazar, and Jiun-Chao Wang. "Berry-Esseen Type Estimate and Return Sequence for Parabolic Iteration in the Upper Half-Plane." *International Mathematics Research Notices*, 2019. rnz365.

[AW14] Michael D. Anshelevich and John D. Williams. "Limit Theorems for Monotonic Convolution and the Chernoff Product Formula." *International Mathematics Research Notices*, **2014**(11):2990–3021, 2014.

[AW16] Michael D. Anshelevich and John D. Williams. "Operator-Valued Monotone Convolution Semigroups and an Extension of the Bercovici-Pata Bijection." *Documenta Mathematica*, **21**:841–871, 2016.

[Bar96] Andrew R. Barron. "Entropy and the Central Limit Theorem." *Annals of Probability*, **14**:336–342, 01 1996.

[Bau04] Robert O. Bauer. "Löwner's Equation from a Noncommutative Probability Perspective." *Journal of Theoretical Probability*, **17**(2):435–457, 2004.

[Bau05] Robert O. Bauer. "Chordal Loewner families and univalent Cauchy transforms." *Journal of Mathematical Analysis and Applications*, **302**(2):484 – 501, 2005.

[BBH08] Itaï Ben Yaacov, Alexander Berenstein, C. Ward Henson, and Alexander Usvyatsov. "Model theory for metric structures." In Z. Chatzidakis et al., editor, *Model Theory with Applications to Algebra and Analysis, Vol. II*, volume 350 of *London Mathematical Society Lecture Notes Series*, pp. 315–427. Cambridge University Press, 2008.

[BC15] Rémi Boutonnet and Alessandro Carderi. "Maximal amenable von Neumann subalgebras arising from maximal amenable subgroups." *Geometric and Functional Analysis*, **25**:1688–1705, 2015.

[BCG03] P. Biane, M. Capitaine, and A. Guionnet. "Large deviation bounds for matrix Brownian motion." *Inventiones Mathematicae*, **152**:433–459, 2003.

[BE85] D. Bakry and Michel Émery. "Diffusions hypercontractives." In *Séminaire de Probabilités XIX, 1983/84*, volume 1123 of *Lecture Notes in Mathematics*, pp. 177–206. Springer, Berlin, 1985.

[Ber05] Hari Bercovici. "A remark on monotonic convolution." *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, **8**(01):117–120, 2005.

[Ber06] Hari Bercovici. "On Boolean convolutions." *Journal of Operator Theory*, **420**:7–13, 2006. Volume 6 of "Theta Ser. Adv. Math".

[BH18] Rémi Boutonnet and Cyril Houdayer. "Amenable absorption in amalgamated free product von Neumann algebras." *Kyoto Journal of Mathematics*, **58**:583–593, 2018.

[Bia97] Philippe Biane. "Segal-Bargmann Transform, functional calculus on matrix spaces and the theory of semi-circular and circular systems." *Journal of Functional Analysis*, **144**(1):232 – 286, 1997.

[Bia98] Philippe Biane. "Processes with free increments." *Mathematische Zeitschrift*, **227**(1):143–174, 1998.

[BKM05] V. I. Bogachev, A. V. Kolesnikov, and K. V. Medvedev. "Triangular transformations of measures." *Sbornik Mathematics*, **196**(3):309–335, 2005.

[BL76] Herm Jan Brascamp and Elliot H. Lieb. "On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation." *Journal of Functional Analysis*, **22**:366–389, 1976.

[BL00] S. G. Bobkov and M. Ledoux. "From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities." *Geometric and Functional Analysis*, **10**:1028–1052, 2000.

[Bla06] Bruce Blackadar. *Operator Algebras: Theory of C^*-algebras and von Neumann algebras*, volume 122 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, Heidelberg, 2006.

[BMS13] Serban T. Belinschi, Tobias Mai, and Roland Speicher. "Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem." *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 2013.

[BN08] Serban T. Belinschi and Alexandru Nica. "η-series and a Boolean Bercovici-Pata bijection for bounded k-tuples." *Advances in Mathematics*, **217**(1):1–41, 2008.

[BP99] Hari Bercovici and Vittorino Pata. "Stable laws and domains of attraction in free probability theory." *Annals of Mathematics*, **149**:1023–1060, 1999. With an appendix by Philippe Biane.

[BPV12] Serban T. Belinschi, Mihai Popa, and Victor Vinnikov. "Infinite divisibility and a non-commutative Boolean-to-free Bercovici-Pata bijection." *Journal of Functional Analysis*, **262**(1):94–123, 2012.

[BPV13] Serban T. Belinschi, Mihai Popa, and Victor Vinnikov. "On the operator-valued analogues of the semicircle, arcsine and Bernoulli laws." *Journal of Operator Theory*, **70**(1):239–258, 2013.

[BS02] A. Ben Ghorbal and M. Schürmann. "Non-commutative notions of stochastic independence." *Mathematical Proceedings of the Cambridge Philosophical Society*, **133**:531–561, 2002.

[BU10] Itaïï Ben Yaacov and Alexander Usvyatsov. "Continuous first order logic and local stability." *Transactions of the American Mathematical Society*, **362**(10):5213–5259, 10 2010.

[BV92] Hari Bercovici and Dan-Virgil Voiculescu. "Lèvy-Hincin type theorems for multiplicative and additive free convolution." *Pacific Journal of Mathematics*, **153**:217–248, 1992.

[BV93] Hari Bercovici and Dan-Virgil Voiculescu. "Free convolution of measures with unbounded Support." *Indiana University Mathematics Journal*, **42**(3):733–773, 1993.

[Cap10] Valerio Capraro. "A survey on Connes' embedding conjecture." arXiv:1003.2076, 2010.

[Ceb13] Guillaume Cébron. "Free convolution operators and free Hall transform." *Journal of Functional Analysis*, **265**(11):2645 – 2708, 2013.

[CG08] G. P. Christyakov and F. Götze. "Limit theorems in free probability theory. I." *Annals of Probability*, **36**(1):54–90, 03 2008.

[CN19] Ian Charlesworth and Brent Nelson. "Free Stein irregularity and dimension." arXiv:1902.02379, 2019.

[da 18] Ricardo Correa da Silva. "Lecture Notes on Non-commutative L_p-spaces." arXiv:1803.02390, 2018.

[Dab10] Yoann Dabrowski. "A non-commutative path space approach to stationary free stochastic differential equations." arxiv:1006.4351, 2010.

[Dab17] Yoann Dabrowksi. "A Laplace Principle for Hermitian Brownian Motion and Free Entropy I: the convex functional case." arXiv:1604.06420, 2017.

[DGS16] Yoann Dabrowski, Alice Guionnet, and Dimitri Shlyakhtenko. "Free Transport for Convex Potentials." arXiv:1701.00132, 2016.

[DHK13] Bruce K. Driver, Brian C. Hall, and Todd Kemp. "The large-N limit of the Segal-Bargmann transform on U_N." *Journal of Functional Analysis*, **265**(11):2585 – 2644, 2013.

[Dyk07] Kenneth J. Dykema. "Multilinear function series and transforms in free probability theory." *Advances in Mathematics*, **208**:351–407, Jan 2007.

[FHS13] Ilias Farah, Bradd Hart, and David Sherman. "Model theory of operator algebras I: stability." *Bulletin of the London Mathematical Society*, **45**(4):825–838, 2013.

[FHS14a] Ilias Farah, Bradd Hart, and David Sherman. "Model theory of operator algebras II: model theory." *Israel Journal of Mathematics*, **201**(1):477–505, 2014.

[FHS14b] Ilias Farah, Bradd Hart, and David Sherman. "Model theory of operator algebras III: elementary equivalence and II_1 factors." *Bulletin of the London Mathematical Society*, **46**(3):609–628, 2014.

[FHS18] Uwe Franz, Takahiro Hasebe, and Sebastian Schleißinger. "Monotone increment processes, classical Markov processes, and Loewner chains." arXiv:1811.02873, 2018.

[Fol99] Gerald B. Folland. *Real Analysis: Modern Techniques and their Applications*. Pure and Applied Mathematics. John Wiley & Sons, New York, 2 edition, 1999.

[GM06] Alice Guionnet and Edouard Maurel-Segala. "Combinatorial Aspects of Random Matrix Models." *Latin American Journal of Probability and Statistics (ALEA)*, **1**:241–279, 2006.

[Gro75] Leonard Gross. "Logarithmic Sobolev Inequalities." *American Journal of Mathematics*, **97**(4):1061–1083, 1975.

[GS09] Alice Guionnet and Dimitri Shlyakhtenko. "Free Diffusions and Matrix Models with Strictly Convex Interaction." *Geometric and Functional Analysis*, **18**(6):1875–1916, 03 2009.

[GS14] Alice Guionnet and Dimitri Shlyakhtenko. "Free monotone transport." *Inventiones Mathematicae*, **197**(3):613–661, 09 2014.

[GSS92] Peter Glockner, Michael Schürmann, and Roland Speicher. "Realization of free white noises." *Archiv der Mathematik*, **58**(4):407–416, 1992.

[GZ00] Alice Guionnet and Ofer Zeitouni. "Concentration of the spectral measure for large matrices." *Electronic Communications in Probability*, **5**:119–136, 2000.

[Har04] Gilles Hargé. "A convex/log-concave correlation inequality for Gaussian measure and an application to abstract Wiener spaces." *Probability Theory and Related Fields*, **130**:415–440, 2004.

[Has10a] Takahiro Hasebe. "Monotone convolution and monotone infinite divisibility from complex analytic viewpoints." *Infinite Dimensional Analysis, Quantum Probability, and Related Topics*, **13**(1):111–131, 2010.

[Has10b] Takahiro Hasebe. "Monotone Convolution Semigroups." *Studia Mathematica*, **200**:175–199, 2010.

[HJN19] Ben Hayes, David Jekel, Brent Nelson, and Thomas Sinclair. "A random matrix approach to absorption theorems for free products." arXiv:1912.11569, 2019.

[HMS15] J. William Helton, Tobias Mai, and Roland Speicher. "Applications of realizations (aka linearizations) to free probability." arxiv:1511.05330, 2015.

[HN18] Michael Hartglass and Brent Nelson. "Free transport for interpolated free group factors." *Journal of Functional Analysis*, **274**(1):222–251, 2018.

[HS14] Takahiro Hasebe and Hayato Saigo. "On operator-valued monotone independence." *Nagoya Mathematical Journal*, **215**:151–167, 2014.

[HT05] Uffe Haagerup and Steen Thorbjørnsen. "A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group." *Annals of Mathematics*, **162**:711–775, 2005.

[HU06] F. Hiai and Y. Ueda. "Free transportation cost inequalities for noncommutative multi-variables." *Infinite Dimensional Analysis, Quantum Probability, and Related Topics*, **9**:391–412, 2006.

[Jek18] David Jekel. "An elementary approach to free entropy theory for convex potentials." arXiv:1805.08814. To appear in Analysis & PDE Journal, 2018.

[Jek19] David Jekel. "Conditional expectation, entropy, and transport for convex Gibbs laws in free probability." arXiv:1906.10051, 2019.

[Jek20] David Jekel. "Operator-valued chordal Loewner chains and non-commutative probability." *Journal of Functional Analysis*, **278**(10):108452, 2020.

[JL19] David Jekel and Weihua Liu. "An operator of non-commutative independences defined by trees." arXiv:1901.09158, 2019.

[Kap53] Irving Kaplansky. "Modules over operator algebras." *American Journal of Mathematics*, **75**:839–858, 1953.

[Kas80] G. G. Kasparov. "Hilbert C*-modules: theorems of Stinespring and Voiculescu." *Journal of Operator Theory*, **4**(1):133–150, 1980.

[Kat78] Tosio Kato. "Trotter's product formula for an arbitrary pair of self-adjoint contraction semigroups." In I. Gohberg and M. Kac, editors, *Topics in Functional Analysis, Essays Dedicated to M.G. Krein*, volume 3 of *Advances in Mathematics, Supplementary Studies*. Academic Press, 1978.

[Kem16] Todd Kemp. "The large-N limits of Brownian motions on $GL(N)$." *International Mathematics Research Notices*, **2016**(13):4012–4057, 2016.

[Kem17] Todd Kemp. "Heat kernel empirical laws on $U(N)$ and $GL(N)$." *Journal of Theoretical Probability*, **30**(2):397–451, 2017.

[Kom67] Yukio Komura. "Nonlinear semi-groups in Hilbert space." *Journal of the Mathematical Society of Japan*, **19**(4):493–507, 1967.

[KV14] Dmitry S. Kaliuzhnyi-Verbovetskyi and Victor Vinnikov. *Foundations of Free Non-Commutative Function Theory*, volume 199 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2014.

[Lan95] E. Christopher Lance. *Hilbert C^*-Modules: A Toolkit for Operator Algebraists.* London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, 1995.

[Leh01] Franz Lehner. "On the computation of spectra in free probability." *Journal of Functional Analysis*, **183**(2):451 – 471, 2001.

[Lei72] László Leindler. "On a certain converse of Hölder's inequality II." *Acta Universitatis Szegediensis. Acta Scientiarum Mathematicarum*, **33**:217–223, 1972.

[Len07] Romuald Lenczewski. "Decompositions of the free additive convolution." *Journal of Functional Analysis*, **246**(2):330–365, 2007.

[Ler76] Uri Leron. "Trace identities and polynomial identities of $n \times n$ matrices." *Journal of Algebra*, **42**:369–377, 1976.

[Liu18] Weihua Liu. "Relations between convolutions and transforms in operator-valued free probability." arXiv:1809.05789, 2018.

[Lu97] Y.G. Lu. "An interacting free Fock space and the arcsine law." *Probability and Mathematical Statistics*, **17**(1):149–166, 1997.

[Maa92] Hans Maassen. "Addition of freely independent random variables." *Journal of Functional Analysis*, **106**:409–438, 1992.

[McD69] Dusa McDuff. "Uncountably many II_1 factors." *Annals of Mathematics Second Series*, **90**(2):372–377, 1969.

[Mec19] Elizabeth S. Meckes. *The Random Matrix Theory of the Classical Compact Groups.* Cambridge Tracts in Mathematics. Cambridge University Press, 2019.

[Mur97] Naofumi Muraki. "Noncommutative Brownian Motion in monotone Fock Space." *Communications in Mathematical Physics*, **183**:557–570, 1997.

[Mur00] Naofumi Muraki. "Monotonic convolution and monotone Lévy-Hinčin formula." Preprint at https://www.math.sci.hokudai.ac.jp/ thasebe/Muraki2000.pdf, 2000.

[Mur01] Naofumi Muraki. "Monotonic independence, monotonic central limit theorem, and monotonic law of small numbers." *Infinite Dimensional Analysis, Quantum Probability, and Related Topics*, **04**, 2001.

[Mur03] Naofumi Muraki. "The Five independences as natural products." *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, **6**(3):337–371, 2003.

[Nel15a] Brent Nelson. "Free monotone transport without a trace." *Communications in Mathematical Physics*, **334**(3):1245–1298, 2015.

[Nel15b] Brent Nelson. "Free transport for finite depth subfactor planar algebras." *Journal of Functional Analysis*, **268**(9):2586–2620, 2015.

[Nic09] Alexandru Nica. "Multi-variable subordination distributions for free additive convolution." *Journal of Functional Analysis*, **257**(2):428 – 463, 2009.

[OV00] Felix Otto and Cédric Villani. "Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality." *Journal of Functional Analysis*, **173**(2):361–400, 2000.

[Oza04] Narutaka Ozawa. "There is no separable universal II_1 factor." *Proceedings of the American Mathematical Society*, **132**:487–490, 2004.

[Pas73] William L. Paschke. "Inner product modules over B^*-algebras." *Transactions of the American Mathematical Society*, **182**:443–468, 1973.

[Pel06] Vladimir V. Peller. "Multiple operator integrals and higher operator derivatives." *Journal of Functional Analysis*, **233**:515–544, 04 2006.

[Pop83] Sorin Popa. "Maximal injective subalgebras in factors associated with free groups." *Advances in Mathematics*, **50**:27–48, 1983.

[Pop08] Mihai Popa. "A Combinatorial Approach to Monotonic Independence over a C^*-Algebra." *Pacific Journal of Mathematics*, **237**:299–325, 2008.

[Pop09] Mihai Popa. "A new proof for the multiplicative property of the boolean cumulants with applications to the operator-valued case." *Colloquium Mathematicum*, **117**(1):81–93, 2009.

[Pre71] András Prékopa. "Logarithmic concave measures with application to stochastic programming." *Acta Universitatis Szegediensis. Acta Scientiarum Mathematicarum*, **32**:301–315, 1971.

[Pre73] András Prékopa. "On logarithmic concave measures and functions." *Acta Universitatis Szegediensis. Acta Scientiarum Mathematicarum*, **34**:335–343, 1973.

[Pro76] Claudio Procesi. "The invariant theory of $n \times n$ matrices." *Advances in Mathematics*, **19**:306–381, 1976.

[PV13] Mihai Popa and Victor Vinnikov. "Non-commutative functions and the non-commutative Lévy-Hinčin formula." *Advances in Mathematics*, **236**:131–157, 2013.

[PX03] Gilles Pisier and Quanhua Xu. "Non-commutative L^p-spaces." In Williams B. Johnson and Joram Lindenstrauss, editors, *Handbook of the geometry of Banach spaces*, volume 2, pp. 1459–1517. Elsevier, 2003.

[Rai97] E. M. Rains. "Combinatorial Properties of Brownian Motion on the Compact Classical Groups." *Journal of Theoretical Probability*, **10**(3):659–679, 1997.

[Raz74] Yu P. Razmyslov. "Trace Identities of Full Matrix Algebras over a Field of Characteristic Zero." *Mathematics of the USSR-Izvestiya*, **8**(4):727, 1974.

[Raz87] Yu P. Razmyslov. "Trace Identities and Central Polynomials in the Matrix Superalgebras $M_{n,k}$." *Mathematics of the USSR-Sbornik*, **56**(1):187, 1987.

[Rie74] Marc A. Rieffel. "Induced representations of C^*-algebras." *Advances in Mathematics*, **13**:176–257, 1974.

[Sch17] Sebastian Schleißinger. "The Chordal Loewner Equation and Monotone Probability Theory." *Infinite Dimensional Analysis, Quantum Probability, and Related Topics*, **20**(3), 2017.

[Sen08] A. Sengupta. "Traces in two-dimensional QCD: the large-N limit." In *Traces in number theory, geometry and quantumfields*, volume 38 of *Aspects of Mathematics*, pp. 193–212. Vieweg, 2008.

[Sim79] Barry Simon. *Functional Integration and Quantum Physics.* Academic Press, 1979.

[Sim05] Barry Simon. *Trace Ideals and Their Applications.* Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2 edition, 2005.

[Spe94] Roland Speicher. "Multiplicative functions on the lattice of non-crossing partitions and free convolution." *Mathematische Annalen*, **298**:611–628, 1994.

[Spe97] Roland Speicher. "On universal products." In Dan-Virgil Voiculescu, editor, *Free Probability Theory*, volume 12 of *Fields Institute Communications*, pp. 257–266. Amer. Math. Soc., 1997.

[Spe98] Roland Speicher. "Combinatorial theory of the free product with amalgamation and operator-valued free probability theory." *Memoirs of the American Mathematical Society*, **132**(627), 1998.

[Sti55] W. Forrest Stinespring. "Positive functions on C^*-algebras." *Proceedings of the American Mathematical Society*, **6**:211–216, 1955.

[SW97] Roland Speicher and Reza Woroudi. "Boolean convolution." In Dan-Virgil Voiculescu, editor, *Free Probability Theory*, volume 12 of *Fields Institute Communications*, pp. 267–279. Amer. Math. Soc., 1997.

[Tao12] Terence Tao. *An Introduction to Random Matrix Theory*, volume 132 of *Graduate Texts in Mathematics*. American Mathematical Society, 2012.

[Tay72] Joseph L. Taylor. "A general framework for a multi-operator functional calculus." *Advances in Mathematics*, **9**:183–252, 1972.

[Tay73] Joseph L. Taylor. "Functions of several noncommuting variables." *Bulletin of the American Mathematical Society*, **79**:1–34, 1973.

[Tro59] H. F. Trotter. "On the Product of semi-groups of operators." *Proceedings of the American Mathematical Society*, **10**:545–551, 1959.

[VDN92] Dan-Virgil Voiculescu, Kenneth J. Dykema, and Alexandru Nica. *Free Random Variables*, volume 1 of *CRM Monograph Series*. American Mathematical Society, Providence, RI, 1992.

[Voi85] Dan-Virgil Voiculescu. "Symmetries of some reduced free product C^*-algebras." In Huzihiro Araki, Calvin C. Moore, Şerban-Valentin Stratila, and Dan Voiculescu, editors, *Operator Algebras and their Connections with Topology and Ergodic Theory*, pp. 556–588. Springer, Berlin, Heidelberg, 1985.

[Voi86] Dan-Virgil Voiculescu. "Addition of certain non-commuting random variables." *Journal of Functional Analysis*, **66**(3):323–346, 1986.

[Voi93] Dan-Virgil Voiculescu. "The analogues of entropy and Fisher's information in free probability, I." *Communications in Mathematical Physics*, **155**(1):71–92, 1993.

[Voi94] Dan-Virgil Voiculescu. "The analogues of entropy and of Fisher's information in free probability, II." *Inventiones Mathematicae*, **118**:411–440, 1994.

[Voi95] Dan-Virgil Voiculescu. "Operations on certain non-commutative operator-valued random variables." In *Recent Advances in Operator Algebras*, number 232 in Astérisque, pp. 243–275. Societe mathematique de France, 1995.

[Voi96] Dan-Virgil Voiculescu. "The analogues of entropy and of Fisher's information measure in free probability, III: Absence of Cartan subalgebras." *Geometric and Functional Analysis*, **6**:172–199, 1996.

[Voi98a] Dan-Virgil Voiculescu. "The analogues of entropy and of Fisher's information in free probability V." *Inventiones Mathematicae*, **132**:189–227, 1998.

[Voi98b] Dan-Virgil Voiculescu. "A strengthened asymptotic freeness result for random matrices with applications to free entropy." *International Mathematics Research Notices*, **1998**(1):41–63, 1998.

[Voi00] Dan-Virgil Voiculescu. "The coalgebra of the difference quotient and free probability." *International Mathematics Research Notices*, **2000**(2):79–106, 2000.

[Voi02a] Dan-Virgil Voiculescu. "Analytic subordination consequences of free Markovianity." *Indiana University Mathematics Journal*, **51**:1161–1166, 2002.

[Voi02b] Dan-Virgil Voiculescu. "Free Entropy." *Bulletin of the London Mathematical Society*, **34**:257–278, 2002.

[Voi04] Dan-Virgil Voiculescu. "Free analysis questions I: duality transform for the coalgebra of $\partial_{X:B}$." *International Mathematics Research Notices*, **2004**(16):793–822, 2004.

[Voi10] Dan-Virgil Voiculescu. "Free Analysis Questions II: The Grassmannian Completion and the Series Expansions at the Origin." *Journal für die reine und angewandte Mathematik (Crelles Journal)*, **645**:155–236, 2010.

[Wil17] John D. Williams. "Analytic function theory for operator-valued free probability." *Journal für die reine und angewandte Mathematik (Crelles Journal)*, **2017**:119–149, 2017.