DAVID JEKEL

**AN ELEMENTARY APPROACH TO
FREE ENTROPY THEORY FOR CONVEX POTENTIALS**

msp

AN ELEMENTARY APPROACH TO
FREE ENTROPY THEORY FOR CONVEX POTENTIALS

DAVID JEKEL

We present an alternative approach to the theory of free Gibbs states with convex potentials. Instead of solving SDEs, we combine PDE techniques with a notion of asymptotic approximability by trace polynomials for a sequence of functions on $M_N(\mathbb{C})_{\mathrm{sa}}^m$ to prove the following. Suppose μ_N is a probability measure on $M_N(\mathbb{C})_{\mathrm{sa}}^m$ given by uniformly convex and semiconcave potentials V_N, and suppose that the sequence DV_N is asymptotically approximable by trace polynomials. Then the moments of μ_N converge to a noncommutative law λ. Moreover, the free entropies $\chi(\lambda)$, $\underline{\chi}(\lambda)$, and $\chi^*(\lambda)$ agree and equal the limit of the normalized classical entropies of μ_N.

1. Introduction

1A. *Motivation and main ideas.* Since Voiculescu [1993; 1994; 1998] introduced the free entropy of a noncommutative law, a number of open problems have prevented a satisfying unification of the theory (as explained in [Voiculescu 2002]). The free entropy χ was defined by taking the lim sup as $N \to \infty$ of the normalized log volume of the space of microstates, where the microstates are certain tuples of $N \times N$ self-adjoint matrices having approximately the correct distribution. It is unclear whether using the lim inf instead of the lim sup would yield the same quantity. Voiculescu also defined a nonmicrostates free entropy χ^* by integrating the free Fisher information of $X + t^{1/2}S$, where S is a free semicircular family free from X, and conjectured that $\chi = \chi^*$.

Biane, Capitaine, and Guionnet [Biane et al. 2003] showed that $\chi \leq \chi^*$ as a consequence of their large deviation principle for the GUE (see also [Cabanal Duvillard and Guionnet 2001]). The proof relied

on stochastic differential equations relative to Hermitian Brownian motion and analyzed exponential functionals of Brownian motion. Recent work of Dabrowski [2016] combined these ideas with stochastic control theory and ultraproduct analysis in order to show that $\chi = \chi^*$ for free Gibbs states defined by a convex and sufficiently regular potential. This resolves this part of the unification problem for a significant class of noncommutative laws.

This paper will prove a result similar to Dabrowski's using deterministic rather than stochastic methods. We want to argue as directly as possible that the classical entropy and Fisher's information of a sequence of random matrix models converge to their free counterparts. Let us motivate and sketch the main ideas, beginning with the heuristics behind Voiculescu's nonmicrostates entropy χ^*.

Consider a noncommutative law λ of an m-tuple and suppose λ is the limit of a sequence of random $N \times N$ matrix distributions μ_N given by convex, semiconcave potentials $V_N : M_N(\mathbb{C})^m_{sa} \to \mathbb{R}$. Let $\sigma_{t,N}$ be the distribution of m independent GUE matrices which each have normalized variance t, and let σ_t be the noncommutative law of m free semicircular variables which each have variance t. Let $V_{N,t}$ be the potential corresponding to the convolution $\mu_N * \sigma_{t,N}$. The classical Fisher information $\mathcal{I}$ satisfies

$$\frac{d}{dt} \frac{1}{N^2} h(\mu_N * \sigma_{t,N}) = \frac{1}{N^3} \mathcal{I}(\mu_N * \sigma_{t,N}) = \int \|DV_{N,t}(x)\|_2^2 \, d(\mu_N * \sigma_{t,N})(x),$$

and from this we deduce that

$$\frac{1}{N^2} h(\mu_N) + \frac{m}{2} \log N = \frac{1}{2} \int \left(\frac{m}{1+t} - \frac{1}{N^3} \mathcal{I}(\mu_N * \sigma_{t,N}) \right) dt + \frac{m}{2} \log 2\pi e.$$

As $N \to \infty$, we expect the left-hand side to converge to the microstates free entropy $\chi(\lambda)$ because the distribution μ_N should be concentrated on the microstate spaces of the law λ. On the other hand, we expect the right-hand side to converge to the Voiculescu's nonmicrostates free entropy $\chi^*(\lambda)$ defined by

$$\chi^*(\lambda) = \frac{1}{2} \int \left(\frac{m}{1+t} - \Phi^*(\lambda \boxplus \sigma_t) \right) dt + \frac{m}{2} \log 2\pi e,$$

where Φ^* is the free Fisher information and $\boxplus$ denotes the free convolution [Voiculescu 1998].

Under suitable assumptions on V_N, the microstates free entropy $\chi(\lambda)$ is the lim sup of normalized classical entropies of μ_N. On the right-hand side, we want to show that $N^{-3} \mathcal{I}(\mu_N * \sigma_{t,N}) \to \Phi^*(\lambda \boxplus \sigma_t)$ for all $t \geq 0$. Since the Fisher information is the $L^2(\mu_N)$ norm squared of the score function or (classical) conjugate variable $DV_{N,t}(x)$, we want to prove that the classical conjugate variables $DV_{N,t}(x)$ behave asymptotically like the free conjugate variables for $\lambda \boxplus \sigma_t$ for all t.

This would not be surprising because classical objects associated to invariant random matrix ensembles often behave asymptotically like their free counterparts. For instance, Biane [1997] showed that the entrywise Segal–Bargmann transform of noncommutative functions evaluated on $N \times N$ matrices can be approximated by the free Segal–Bargmann transform computed through analytic functional calculus. Similarly, Guionnet and Shlyakhtenko [2014, Theorem 4.7] showed that classical monotone transport maps for certain random matrix models approximate the free monotone transport. Moreover, Dabrowski's approach [2016] to proving $\chi = \chi^*$ involved constructing solutions to free SDEs as ultraproducts of the solutions to classical SDEs.

In Section 3D, we make precise the idea that a sequence of functions on $M_N(\mathbb{C})^m_{sa}$ has a "well-defined, noncommutative asymptotic behavior" by defining *asymptotic approximability by trace polynomials* (Definition 3.24). We assume that DV_N at time zero has the approximation property and must show that the same is true for $DV_{N,t}$ for all t.

First, we show that this property is preserved under several operations on sequences, including composition and convolution with the Gaussian law $\sigma_{N,t}$ (see Section 3D). Then in Section 6 we analyze the PDE that describes the evolution of $V_{N,t}$. We show that for all t the solution $V_{N,t}$ can be approximated in a dimension-independent way by applying a sequence of simpler operations, each of which preserves asymptotic approximability by trace polynomials. In other words, if the initial data DV_N is asymptotically approximable by trace polynomials, then so is $DV_{N,t}$, and hence we obtain convergence of the classical Fisher information to the free Fisher information.

This proves the equality $\chi(\lambda) = \chi^*(\lambda)$ whenever a sequence of log-concave random matrix models μ_N converges to λ in an appropriate sense (Theorem 7.1). Another result (Theorem 4.1), proved by similar techniques, establishes sufficient conditions for a sequence of log-concave random matrix models μ_N to converge in moments to a noncommutative law λ, so that Theorem 7.1 can be applied. As a consequence, we show that $\chi = \chi^*$ for a class of free Gibbs states.

1B. *Main results.* To fix notation, let $M_N(\mathbb{C})^m_{sa}$ be space of m-tuples $x = (x_1, \ldots, x_m)$ of self-adjoint $N \times N$ matrices and let $\|x\|_2 = \left(\sum_j \tau_N(x_j^2)\right)^{1/2}$, where $\tau_N = (1/N)\operatorname{Tr}$. We denote by $\|x\|_\infty$ the maximum of the operator norms $\|x_j\|$. Recall that a trace polynomial $f(x_1, \ldots, x_m)$ is a linear combination of terms of the form

$$p(x) \prod_{j=1}^n \tau(p_j(x)),$$

where p and p_j are noncommutative polynomials in $x_1, \ldots, x_m$ (see Section 3A).

Consider a sequence of potentials $V_N : M_N(\mathbb{C})^m_{sa} \to \mathbb{R}$ such that $V_N(x) - (c/2)\|x\|_2^2$ is convex and $V_N(x) - (C/2)\|x\|_2^2$ is concave for some $0 < c < C$. Define the associated probability measure μ_N by

$$d\mu_N(x) = \frac{1}{Z_N} e^{-N^2 V_N(x)}\, dx, \quad Z_N = \int_{M_N(\mathbb{C})^m_{sa}} e^{-N^2 V_N(x)}\, dx.$$

Assume that the sequence of normalized gradients $DV_N(x) = N\nabla V_N(x)$ is asymptotically approximable by trace polynomials in the sense that for every $\epsilon > 0$ and $R > 0$ there exists a trace polynomial $f(x)$ such that

$$\limsup_{N \to \infty}\ \sup_{\|x\|_\infty \leq R} \|DV_N(x) - f(x)\|_2 \leq \epsilon,$$

where $\|x\|_\infty$ denotes the maximum of the operator norms of the x_j. Also, assume that $\int (x - \tau_N(x))\, d\mu_N(x)$ is bounded in operator norm as $N \to \infty$ (it will be zero if μ_N is unitarily invariant or has expectation zero). In this case, we have the following:

(1) There exists a constant R_0 such that $\mu_N(\|x\|_\infty \geq R_0 + \delta) \leq me^{-cN\delta^2/2}$ for $\delta > 0$.

(2) There exists a noncommutative law λ such that

$$\lim_{N\to\infty} \int \tau_N(p(x))\, d\mu_N(x) = \lambda(p)$$

for every noncommutative polynomial p.

(3) The measures μ_N exhibit exponential concentration around λ in the sense that

$$\lim_{N\to\infty} \frac{1}{N^2} \log \mu_N(\|x\|_\infty \leq R, |\tau_N(p(x)) - \lambda(p)| \geq \delta) < 0$$

for every $R > 0$ and every noncommutative polynomial p.

(4) The law λ has finite free entropy and we have

$$\chi(\lambda) = \underline{\chi}(\lambda) = \chi^*(\lambda) = \lim_{N\to\infty} \frac{1}{N^2}\Big(h(\mu_N) + \frac{m}{2}\log N\Big),$$

where χ and $\underline{\chi}$ are respectively the $\limsup$ and $\liminf$ versions of microstates free entropy, χ^* is the nonmicrostates free entropy, and h is the classical entropy.

(5) The same holds for $\mu_N * \sigma_{t,N}$ and $\lambda \boxplus \sigma_t$, where $\sigma_{t,N}$ is the law of m independent GUE matrices with variance t and σ_t is the law of m free semicircular variables with variance t.

(6) The law λ has finite free Fisher information. If $\mathcal{I}$ is the classical Fisher information and Φ^* is the free Fisher information, then

$$\lim_{N\to\infty} \frac{1}{N^3}\mathcal{I}(\mu_N * \sigma_{t,N}) = \Phi^*(\lambda \boxplus \sigma_t).$$

(7) The functions $t \mapsto (1/N^3)\mathcal{I}(\mu_N * \sigma_{t,N})$ and $t \mapsto \Phi^*(\lambda \boxplus \sigma_t)$ are decreasing and Lipschitz in t with the absolute value of the derivative bounded by $C^2 m(1 + Ct)^{-2}$.

Claims (1) and (3) are standard concentration estimates (see Section 2E), which we do not prove in this paper, but we include them in the statement to clarify the big picture. Claim (2) is proved in Theorem 4.1, which is similar to the earlier results [Guionnet and Shlyakhtenko 2009, Theorem 4.4; Dabrowski et al. 2016, Proposition 50 and Theorem 51; Dabrowski 2016, Theorem 4.4]. Claims (4) through (7) come from Theorem 7.1, which is similar to [Dabrowski 2016, Theorem A].

In particular, we recover [Dabrowski 2016, Theorem A] that $\chi(\lambda) = \underline{\chi}(\lambda) = \chi^*(\lambda)$ when the law λ is a free Gibbs state given by a sufficiently regular convex noncommutative potential $V(X)$, because taking $V_N = V$ will define a sequence of random matrix models μ_N which concentrate around the noncommutative law λ.

Unlike Dabrowski, we do not provide an explicit formula for $(d/dt)\Phi(\lambda \boxplus \sigma_t)$. However, we are able to prove that $\Phi(\lambda \boxplus \sigma_t)$ is Lipschitz in t rather than merely having a derivative in $L^2(dt)$ (and hence being $\frac{1}{2}$-Hölder continuous) as shown by Dabrowski. Our results also allow slightly more flexibility in the choice of random matrix models, so that we do not have to assume that V_N is given by exactly the same formula for every N or that V_N is exactly unitarily invariant.

1C. *Organization of paper.* Section 2 establishes notation and reviews basic facts from noncommutative probability and random matrix theory.

Section 3 defines the algebra of trace polynomials and describes how they behave under differentiation and convolution with Gaussians. We then introduce the notion that a sequence $\{\phi_N\}$ of functions $M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})_{\mathrm{sa}}^m$ of $\mathbb{C}$ is asymptotically approximable by trace polynomials. We show that this approximation property is preserved under several operations including composition and Gaussian convolution.

Section 4 proves Theorem 4.1 concerning the convergence of moments for the measure μ_N (claims (1)–(3) of Section 1B). We evaluate $\int u \, d\mu_N$ for a Lipschitz function u as $\lim_{t\to\infty} T_t^{V_N} u$, where $T_t^{V_N}$ is the semigroup such that $u_t = T_t^{V_N} u$ solves the equation $\partial_t u_t = (2N)^{-1}\Delta u_t - DV \cdot \nabla u_t$. We approximate $T_t^{V_N}$ by iterating simpler operations in order to show that if $N\nabla V_n$ and u_N are asymptotically approximable by trace polynomials, then so is $T_t^{V_N} u_N$, and hence that $\lim_{N\to\infty} \int u_N \, d\mu_N$ exists.

Section 5 reviews the definitions of free entropy and Fisher's information. We also show that the microstates free entropies $\chi(\lambda)$ and $\underline{\chi}(\lambda)$ are the lim sup and lim inf of normalized classical entropies of μ_N, provided that μ_N concentrates around λ and satisfies some mild operator norm tail bounds, and that $\{V_N\}$ is asymptotically approximable by trace polynomials. Similarly, if $\{DV_N\}$ is asymptotically approximable by trace polynomials, then the normalized classical Fisher information converges to the free Fisher information.

Section 6 considers the evolution of the potential $V_N(x, t)$ corresponding to $\mu_N * \sigma_{t,N}$, where $\sigma_{t,N}$ is the law of m independent GUE of variance t. Our goal is to show that if $DV_N(x, 0)$ is asymptotically approximable by trace polynomials, then so is $N\nabla V_N(x, t)$ for all $t > 0$, so that we can apply our previous result that the classical Fisher information converges to the free Fisher information. As in Section 4, we construct the semigroup R_t which solves the PDE as a limit of iterates of simpler operations which are known to preserve asymptotic approximation by trace polynomials.

In Section 7 we conclude the proof of our main theorem on free entropy and Fisher's information (Theorem 7.1), which establishes claims (4)–(7) of Section 1B, assuming a weakened version of the hypothesis and conclusion of Theorem 4.1.

In Section 8, we characterize the limiting noncommutative laws λ which arise in Theorem 4.1 as the free Gibbs states for a certain class of potentials. In particular, we apply Theorem 7.1 to show that $\chi = \chi^*$ for several types of free Gibbs states considered in previous literature.

2. Preliminaries

Here we fix notation and discuss background results that will be used throughout the paper.

2A. *Notation for matrix algebras.* Let $M_N(\mathbb{C})$ denote the $N \times N$ matrices over $\mathbb{C}$ and let $M_N(\mathbb{C})_{\mathrm{sa}}$ be the self-adjoint elements. Note that $M_N(\mathbb{C})_{\mathrm{sa}}^m$ is a real inner product space with the inner product $\langle x, y \rangle_{\mathrm{Tr}} := \sum_{j=1}^m \mathrm{Tr}(x_j y_j)$ for $x = (x_1, \dots, x_m)$ and $y = (y_1, \dots, y_m)$. Moreover, $M_N(\mathbb{C})^m$ can be canonically identified with the complexification $\mathbb{C} \otimes_{\mathbb{R}} M_N(\mathbb{C})_{\mathrm{sa}}^m$ by decomposing each matrix into its self-adjoint and anti-self-adjoint parts.

Being a real inner product space, $M_N(\mathbb{C})_{\mathrm{sa}}$ is isomorphic to $\mathbb{R}^{mN^2}$. An explicit choice of coordinates can be made using the following orthonormal basis for $M_N(\mathbb{C})_{\mathrm{sa}}$:

$$\mathcal{B}_N = \{E_{k,k}\}_{k=1}^N \cup \left\{\frac{1}{\sqrt{2}}E_{k,\ell} + \frac{1}{\sqrt{2}}E_{\ell,k}\right\}_{k<\ell} \cup \left\{\frac{i}{\sqrt{2}}E_{k,\ell} - \frac{i}{\sqrt{2}}E_{\ell,k}\right\}_{k<\ell}. \tag{2-1}$$

This basis has the property that for all $x, y, z \in M_N(\mathbb{C})$, we have

$$\sum_{b\in\mathcal{B}_N} xbybz = xz\,\mathrm{Tr}(y), \tag{2-2}$$

which follows from an elementary computation.

We denote the norm corresponding to Tr by $|\cdot|$ (essentially the Euclidean norm). We denote the normalized trace by $\tau_N = (1/N)\,\mathrm{Tr}$. We denote the corresponding inner product by $\langle x, y\rangle_2 = \sum_{j=1}^m \tau_N(x_j y_j)$ and the norm by $\|\cdot\|_2$. For $x \in M_N(\mathbb{C})$, we denote the operator norm by $\|x\|$. Similarly, if $x = (x_1, \dots, x_m) \in M_N(\mathbb{C})^m$, we write $\|x\|_\infty = \max_j \|x_j\|$.

The symbols ∇ and Δ will represent the gradient and Laplacian operators with respect to the coordinates of $M_N(\mathbb{C})_{\mathrm{sa}}$ in the nonnormalized inner product $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathrm{Tr}}$. The symbols D and L_N will denote the normalized versions $N\nabla$ and $(1/N)\Delta$ respectively, as well as the corresponding linear transformations on the algebra of trace polynomials. This normalization and notation will be explained and justified in Section 3B.

2B. *Noncommutative probability spaces and laws.* The following are standard definitions and facts in noncommutative probability. For further background, see [Voiculescu et al. 1992; Nica and Speicher 2006; Anderson et al. 2010, §5].

Definition 2.1. A *von Neumann algebra* is a unital $\mathbb{C}$-algebra $\mathcal{M}$ of bounded operators on a Hilbert space $\mathcal{H}$ which is closed under adjoints and closed in the weak operator topology.

Definition 2.2. A *tracial von Neumann algebra* or *noncommutative probability space* is a von Neumann algebra $\mathcal{M}$ together with a bounded linear map $\tau : \mathcal{M} \to \mathbb{C}$ which is continuous in the weak operator topology and satisfies $\tau(1) = 1$, $\tau(xy) = \tau(yx)$, and $\tau(x^*x) \geq 0$. The map τ is called a *trace*.

Definition 2.3. For $m \geq 1$, we denote by $\mathrm{NCP}_m = \mathbb{C}\langle X_1, \dots, X_m\rangle$ the algebra of noncommutative polynomials in $X_1, \dots, X_m$, equipped with conjugate-linear involution $*$ such that $X_j^* = X_j$ and $(pq)^* = q^*p^*$. A *noncommutative law* (for an m-tuple) is a map $\lambda : \mathrm{NCP}_m \to \mathbb{C}$ such that

(1) λ is linear,

(2) λ is unital (that is, $\lambda(1) = 1$),

(3) λ is positive, that is, for every $p(X) \in \mathbb{C}\langle X_1, \dots, X_m\rangle$, we have $\lambda(p(X)^*p(X)) \geq 0$,

(4) λ is tracial, that is, $\lambda(p(X)q(X)) = \lambda(q(X)p(X))$.

We denote by Σ_m the space of noncommutative laws equipped with the topology of pointwise convergence on $\mathbb{C}\langle X_1, \dots, X_m\rangle$, that is, convergence in noncommutative moments.

Definition 2.4. We say that a noncommutative law λ is *bounded by R* if we have

$$|\lambda(X_{i_1}, \dots, X_{i_n})| \leq R^n.$$

We denote the space of such laws by $\Sigma_{m,R}$.

Definition 2.5. Suppose that $x_1, \ldots x_m$ are bounded self-adjoint elements of a tracial von Neumann algebra $(\mathcal{M}, \tau)$. Then the *law of* $x = (x_1, \ldots, x_m)$ is the map

$$\lambda_x : \mathbb{C}\langle X_1, \ldots, X_n \rangle \to \mathbb{C} : p(X) \mapsto \tau(p(x)).$$

Definition 2.6. Let $M_N(\mathbb{C})$ be the algebra of $N \times N$ matrices over $\mathbb{C}$. Let $\tau_N = (1/N) \operatorname{Tr}$ be the normalized trace. Then $(M_N(\mathbb{C}), \tau_N)$ is a tracial von Neumann algebra, and hence, for every m-tuple of self-adjoint matrices $x = (x_1, \ldots, x_m)$, the law λ_x is defined by Definition 2.5.

Proposition 2.7. *The space $\Sigma_{m,R}$ is compact, separable, and metrizable. Moreover, every $\mu \in \Sigma_{m,R}$ can be realized as λ_x for some tuple $x = (x_1, \ldots, x_m)$ of self-adjoint elements of a tracial von Neumann algebra $(\mathcal{M}, \tau)$ with $\|x\|_\infty \le R$.*

For the proof of the claim that every noncommutative law can be realized by operators, see [Anderson et al. 2010, Proposition 5.2.14].

2C. *Noncommutative L^α-norms.* On several occasions, we will need to use the noncommutative L^α norms for $\alpha \in [1, +\infty]$. (Here we use α rather than p since the letter p will often be used for a polynomial.) If y is any element of a tracial von Neumann algebra $(\mathcal{M}, \tau)$, then we define $|y| = (y^* y)^{1/2}$ using continuous functional calculus. For $\alpha \in [0, +\infty)$, we define $\|y\|_\alpha = \tau(|y|^\alpha)^{1/\alpha}$. We also define $\|y\|_\infty$ to be the operator norm.

Proposition 2.8. *If $(\mathcal{M}, \tau)$ is a tracial von Neumann algebra and $\alpha \in [1, +\infty]$, then $\|\cdot\|_\alpha$ defines a norm. Moreover, we have the noncommutative Hölder's inequality*

$$\|x_1 \cdots x_n\|_\alpha \le \|x_1\|_{\alpha_1} \cdots \|x_n\|_{\alpha_n}$$

whenever

$$\alpha, \alpha_1, \ldots, \alpha_n \in [1, +\infty], \quad \frac{1}{\alpha_1} + \cdots + \frac{1}{\alpha_n} = \frac{1}{\alpha}.$$

Moreover, we have $|\tau(y)| \le \|y\|_1$.

A standard proof of the Hölder inequality uses polar decomposition, complex interpolation, and the three lines lemma. We will in fact only need this inequality for the trace τ_N on $M_N(\mathbb{C})$. Modulo renormalization of the trace, the inequality for matrices follows from the treatment of trace-class operators in [Simon 2005]; see especially Theorems 1.15 and 2.8, as well as the references cited on p. 31. For the setting of von Neumann algebras, a convenient proof can be found in [Correa da Silva 2018, Theorems 2.4–2.6]; for an overview and further history see [Pisier and Xu 2003, §2].

Remark 2.9. One can define the noncommutative L^α norm for a tuple $(y_1, \ldots, y_m)$ as

$$\|(y_1, \ldots, y_m)\|_\alpha = \begin{cases} \tau(|y_1|^\alpha + \cdots + |y_m|^\alpha)^{1/\alpha}, & \alpha \in [1, +\infty), \\ \max_j \|y_j\|, & \alpha = +\infty. \end{cases}$$

However, for tuples, we will only need to use the 2 and ∞ norms.

2D. *Free independence, semicircular law, and GUE.* We will use the following standard definitions and facts from free probability. For further background, refer to [Voiculescu 1986; 1991; Voiculescu et al. 1992; Nica and Speicher 2006; Anderson et al. 2010].

Let $(\mathcal{M}, \tau)$ be a tracial von Neumann algebra, and let $\mathcal{A}_1, \ldots, \mathcal{A}_n$ be unital $*$-subalgebras of $\mathcal{M}$. Then we say that $\mathcal{A}_1, \ldots, \mathcal{A}_m$ are *freely independent* if given $a_1, \ldots, a_k$ with $a_j \in \mathcal{A}_{i_j}$ and $i_j \neq i_{j+1}$ and $\tau(a_j) = 0$ for each j, we have also $\tau(a_1 \cdots a_k) = 0$.

In particular, if $S_1, \ldots, S_n$ are *subsets* of $\mathcal{M}$, then we say that they are freely independent if the unital $*$-subalgebras they generate are freely independent. Thus, for instance, self-adjoint elements $x_1, \ldots, x_m$ of $\mathcal{M}$ are freely independent if given polynomials $f_1, \ldots, f_k$ and indices $i_1, \ldots, i_k$ with $i_j \neq i_{j+1}$ such that $\tau(f_j(X_{i_j})) = 0$, we have also $\tau(f_1(X_{i_1}) \cdots f_k(X_{i_k})) = 0$.

The *free convolution* of two noncommutative laws μ and ν (of self-adjoint m-tuples) is defined as the noncommutative law of $(x_1 + y_1, \ldots, x_m + y_m)$, given that $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_m\}$ are freely independent and the noncommutative law of $(x_1, \ldots, x_m)$ is μ, and the noncommutative law of $(y_1, \ldots, y_m)$ is ν. Then $\boxplus$ is well-defined, independent of the particular choice of operators that realize the laws μ and ν. Moreover, $\boxplus$ is commutative and associative.

If $X_1, \ldots, X_m$ are freely independent, then their joint law is determined by the individual laws of the X_j, each of which is represented by a compactly supported probability measure on $\mathbb{R}$. The *semicircle law* (of mean zero and variance 1) is the probability measure given by density

$$\frac{1}{2\pi} \sqrt{4 - x^2}\, \mathbf{1}_{[-2,2]}(x)\, dx.$$

We denote by σ_t the noncommutative law of m freely independent semicircular random variables which each have mean zero and variance t (that is, $\sigma_t(X_j) = 0$ and $\sigma_t(X_j^2) = t$).

These free semicircular families play the role of multivariable Gaussians in free probability. Moreover, the noncommutative laws $\{\sigma_t\}_{t \geq 0}$ form a semigroup under free convolution, that is, $\sigma_s \boxplus \sigma_t = \sigma_{s+t}$ for $s, t \geq 0$.

We denote by $\sigma_{t,N}$ the probability distribution on $M_N(\mathbb{C})_{\mathrm{sa}}^m$ for m independent GUE matrices of normalized variance t, that is,

$$d\sigma_{t,N}(x) = \frac{1}{Z_{N,t}} \exp\left(-N \sum_{j=1}^m \frac{\mathrm{Tr}(x_j^2)}{2t}\right) dx,$$

where $Z_{N,t}$ is chosen so that $\sigma_{t,N}$ is a probability measure. It is well known that the independent GUE matrices behave in the large-N limit like freely independent semicircular random variables; in Section 3, we shall directly state and prove the specific results we will use.

2E. *Concentration and operator norm tail bounds.* The following is a standard concentration estimate for uniformly log-concave random matrix models. The best known proof goes through the log-Sobolev inequality and Herbst's argument (see [Anderson et al. 2010, §4.4.2]), although it can also be proved by directly using the heat semigroup associated to V as in [Ledoux 1992]. We state the theorem here with free probabilistic normalizations.

Theorem 2.10. *Suppose that* $V : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$ *is a potential such that* $V(x) - (c/2)\|x\|_2^2$ *is convex. Define*

$$d\mu(x) = \frac{1}{Z} \exp(-N^2 V(x))\, dx, \quad Z = \int \exp(-N^2 V(x))\, dx.$$

Suppose that $f : M_N(\mathbb{C})^m_{\mathrm{sa}} \to \mathbb{R}$ *is K-Lipschitz with respect to* $\|\cdot\|_2$. *Then*

$$\mu\left(x : f(x) - \int f \, d\mu \geq \delta\right) \leq e^{-cN^2\delta^2/2K^2},$$

and since the same estimate can be applied to $-f$, *we have also*

$$\mu\left(x : \left| f(x) - \int f \, d\mu \right| \geq \delta\right) \leq 2e^{-cN^2\delta^2/2K^2}.$$

   In particular, this concentration estimate applies to the GUE law $\sigma_{t,N}$ with $c = 1/t$. In addition to the concentration estimate, we will also use the fact that such uniformly convex random matrix models have subgaussian moments and therefore have good tail bounds on the probability of large operator norm. The following theorem is a special case of [Hargé 2004, Theorem 1.1] and the application to random matrix models is taken from the proof of [Guionnet and Maurel-Segala 2006, Theorem 3.4].

Theorem 2.11. *Let V and μ be as in Theorem 2.10, and suppose that $f : M_N(\mathbb{C})^m_{\mathrm{sa}} \to \mathbb{R}$ is convex. Let $a = \int x \, d\mu(x)$. Then*

$$\int f(x - a) \, d\mu(x) \leq \int f(y) \, d\sigma_{c^{-1},N}(y).$$

In particular, if $\|x\|_\alpha$ denotes the L^α norm from Section 2C, then for every $\alpha \in [1, +\infty]$ and $\beta \in [1, +\infty)$ we have

$$\int \|x_j - a_j\|^\beta_\alpha \, d\mu(x) \leq \int \|y_j\|^\beta_\alpha \, d\sigma_{c^{-1},N}(y).$$

Proof. The convexity assumption on V means that μ has a log-concave density with respect to the Gaussian measure $\sigma_{c^{-1},N}(y)$. Therefore, the first claim follows from [Hargé 2004, Theorem 1.1]. The second claim follows because norms on vector spaces are convex functions, and the function $t \mapsto t^\beta$ on $[0, +\infty)$ is convex for $\beta \geq 1$. □

Corollary 2.12. *Let $V_N : M_N(\mathbb{C})_{\mathrm{sa}} \to \mathbb{R}$ be a function such that $V_N(x) - (c/2)\|x\|^2_2$ is convex and let μ_N be the corresponding measure. Let $a_{N,j} = \int x_j \, d\mu_N(x)$. Then*

$$\limsup_{N \to \infty} \int \|x_j - a_{N,j}\| \, d\mu_N(x) \leq 2c^{-1/2},$$

and

$$\mu_N\left(x : \|x_j\| \geq \int \|y_j\| \, d\mu_N(y_j) + \delta\right) \leq e^{-c\delta^2 N/2}.$$

Proof. In light of Theorem 2.11, for the first claim of the corollary, it suffices to check the special case $\sigma_{c^{-1},N}$. This special case is a standard result in random matrix theory; see for instance the proof of [Anderson et al. 2010, Theorem 2.1.22]. The second claim follows from Theorem 2.10 after we observe that the function on $M_N(\mathbb{C})^N_{\mathrm{sa}}$ given by $x \mapsto \|x_j\|_\infty$ is $N^{1/2}$-Lipschitz with respect to $\|\cdot\|_2$. □

2F. Semiconvex and semiconcave functions. We recall the following terminology and facts about semiconvex and semiconcave functions. These results are typically applied to functions from $\mathbb{R}^n \to \mathbb{R}$, but of course they hold equally well if $\mathbb{R}^n$ is replaced by a finite-dimensional real inner product space. In particular, we focus on the case of $M_N(\mathbb{C})_{\mathrm{sa}}^m$.

A function $u : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$ is *semiconvex* if there exists some $c \in \mathbb{R}$ such that $u(x) - (c/2)\|x\|_2^2$ is convex. If this holds for some $c > 0$, then u is said to be *uniformly convex*. Similarly, $u : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$ is said to be *semiconcave* if there exists $C \in \mathbb{R}$ such that $u(x) - (C/2)\|x\|_2^2$ is concave, and it is *uniformly concave* if this holds for some $C < 0$.

Fix m and N. Let $c \leq C$ be real numbers. Then we define

$$\mathcal{E}_{m,N}(c, C) = \left\{ u : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R} : u(x) - \frac{c}{2}\|x\|_2^2 \text{ is convex and } u(x) - \frac{C}{2}\|x\|_2^2 \text{ is concave} \right\}.$$

We will often suppress m and N in the notation and simply write $\mathcal{E}(c, C)$. Throughout the paper, we rely on the following basic properties of functions in $\mathcal{E}(c, C)$.

Proposition 2.13. (1) *The space $\mathcal{E}(c, C)$ is closed under translation, averaging with respect to probability measures, and pointwise limits.*

(2) *A function u is in $\mathcal{E}(c, C)$ if and only if for every point $x_0 \in M_N(\mathbb{C})_{\mathrm{sa}}^m$ there exists some $p \in M_N(\mathbb{C})_{\mathrm{sa}}^m$ such that*

$$u(x_0) + \langle p, x - x_0 \rangle_2 + \tfrac{1}{2}c\|x - x_0\|_2^2 \leq u(x) \leq u(x_0) + \langle p, x - x_0 \rangle_2 + \tfrac{1}{2}C\|x - x_0\|_2^2.$$

(3) *In particular, if $u \in \mathcal{E}(c, C)$, then u is differentiable everywhere.*

(4) *If $u \in \mathcal{E}(c, C)$, then the gradient Du is $\max(|c|, |C|)$-Lipschitz with respect to $\|\cdot\|_2$.*

(5) *If $u \in \mathcal{E}(c, C)$, then*

$$c\|x - y\|_2^2 \leq \langle Du(x) - Du(y), x - y \rangle_2 \leq C\|x - y\|_2^2.$$

(6) *If $u \in \mathcal{E}(c, C)$ for some $c > 0$, then u is bounded below and achieves a global minimum at its unique critical point.*

Sketch of proof. (1) This follows from elementary computation and the fact that the same holds for the class of convex functions.

(2), (3) Suppose that $u \in \mathcal{E}(c, C)$. The convex functions $u(x) - (c/2)\|x\|_2^2$ and $(C/2)\|x\|_2^2 - u(x)$ must have supporting hyperplanes at x_0. This yields one vector p which satisfies the left inequality of (2) and another vector p' satisfying the right inequality. Then one checks that p must equal p' and this implies that u is differentiable at x_0. The converse direction of (2) follows again from the characterization of convexity using supporting hyperplanes.

(4), (5) For smooth functions in $\mathcal{E}(c, C)$, one can check these properties directly using calculus. Now consider a general $u \in \mathcal{E}(c, C)$. Let $u_n = u * \rho_n$, where ρ_n is a smooth probability density supported in the ball of radius $1/n$ around 0. Then u_n is smooth and $u_n \to u$ locally uniformly. Also, $u_n \in \mathcal{E}(c, C)$ by (1); hence Du_n is $\max(|c|, |C|)$-Lipschitz. By the Arzelà–Ascoli theorem, after passing to a subsequence,

we may assume that Du_n converges locally uniformly to some F. It follows from this local uniform convergence that $F = Du$. Moreover, since (4) and (5) hold for Du_n, they also hold for Du.

(6) This is left as an exercise. $\qquad\square$

3. Trace polynomials

In this section, we consider the algebra of trace polynomials in noncommutative variables $X_1, \ldots, X_m$, first defined in [Razmyslov 1974; 1985]. As in [Rains 1997; Cébron 2013; Driver et al. 2013], we describe how trace polynomials behave under differentiation (Section 3B) and convolution with Gaussian (Section 3C). Finally, in Section 3D, we define the property of asymptotic approximability by trace polynomials for a sequence of functions on $M_N(\mathbb{C})_{\mathrm{sa}}^m$, which is one of the key technical tools in our proof.

3A. *Definitions.*

Definition 3.1. We define the *-*algebra of scalar-valued trace polynomials*, or TrP_m^0, as follows. Let $\mathcal{V}$ be the vector space $\mathrm{NCP}_m / \mathrm{Span}(pq - qp : p, q \in \mathrm{NCP}_m)$. We define the vector space

$$\mathrm{TrP}_m^0 = \bigoplus_{n=0}^{\infty} \mathcal{V}^{\odot n}, \tag{3-1}$$

where $\odot$ is the symmetric tensor power over $\mathbb{C}$. Then TrP_m^0 forms a commutative algebra with the tensor operator $\odot$ as the multiplication. We denote the element $p_1 \odot \cdots \odot p_n$ by $\tau(p_1) \cdots \tau(p_n)$, where τ is a formal symbol.

To state the definition more suggestively, an element of TrP_m^0 is a linear combination of terms of the form $\tau(p_1(X)) \cdots \tau(p_n(X))$, where $p_1, \ldots, p_n$ are noncommutative polynomials in $X_1, \ldots, X_m$ and τ is a formal symbol thought of as the trace. By forming a quotient vector space, we identify $\tau(pq)$ with $\tau(qp)$. The trace polynomials form a commutative *-algebra TrP_m^0 over $\mathbb{C}$ where the *-operation is

$$(\tau(p_1(X)) \cdots \tau(p_n(X)))^* = \tau(p_1(X)^*) \cdots \tau(p_n(X)^*) \tag{3-2}$$

and the multiplication operation is the one suggested by the notation.

We define TrP_m^k to be the vector space

$$\mathrm{TrP}_m^k := \mathrm{TrP}_m^0 \otimes \mathbb{C}\langle X_1, \ldots, X_m \rangle^{\otimes k}.$$

We call the elements of TrP_m^1 *operator-valued trace polynomials*. We use the term *trace polynomials* more generally to describe elements of TrP_m^k or tuples of elements from TrP_m^k. Note that TrP_m^1 forms a *-algebra because it is the tensor product of two *-algebras.

Definition 3.2. Suppose that $\mathcal{M}$ is a von Neumann algebra with trace σ. Given $f \in \mathrm{TrP}_m^1$ and a self-adjoint tuple $x = (x_1, \ldots, x_m)$ of elements of $\mathcal{M}$, we define $f(x)$ to be the element of $\mathcal{M}$ given by replacing the formal symbols X_j and τ in f by the operator x_j and the trace σ on $\mathcal{M}$. For instance, if $f(X) = p_0(X) \otimes \tau(p_1(X)) \cdots \tau(p_n(X))$ in TrP_m^1, then

$$f(x) = p_0(x)\sigma(p_1(x)) \cdots \sigma(p_n(x)).$$

In particular, we define $f(x)$ when x is an m-tuple of self-adjoint $N \times N$ matrices by setting $\tau = \tau_N$.

Definition 3.3. If $f \in \mathrm{TrP}_m^0$ and λ is a noncommutative law, we define the evaluation $\lambda(f)$ to be the number obtained by replacing the symbol τ with λ everywhere in f. For example, if $f(X_1, X_2, X_3) = \tau(X_1)\tau(X_2 X_3) + \tau(X_2^2)$, then we define

$$\lambda(f) = \lambda(f(X_1, \ldots, X_m)) = \lambda(X_1)\lambda(X_2 X_3) + \lambda(X_2^2).$$

Definition 3.4. We define the *degree* for elements of NCP_m and TrP_m^k as follows. If $p \in \mathrm{NCP}_m$ is a monomial $p(X_1, \ldots, X_m) = X_{i_1} \cdots X_{i_d}$, then we define $\deg'(p) = d$. If $p_1, \ldots, p_\ell$ and $q_1, \ldots, q_k$ are noncommutative monomials, then consider the element $\tau(p_1) \cdots \tau(p_\ell) q_1 \otimes \cdots \otimes q_k \in \mathrm{TrP}_m^k$, and define

$$\deg'(\tau(p_1) \cdots \tau(p_\ell) q_1 \otimes \cdots \otimes q_k) = \deg'(p_1) \cdots \deg'(p_\ell) \deg'(q_1) \cdots \deg'(q_k).$$

For general $f \in \mathrm{TrP}_m^k$, we define the *degree*, $\deg(f)$, as the infimum of $\max(\deg'(f_1), \ldots, \deg'(f_\ell))$, where $f = f_1 + \cdots + f_\ell$ and each f_j is a product of noncommutative monomials and traces of noncommutative monomials as above. Similarly, for general $f \in \mathrm{NCP}_m$, we define $\deg(f)$ as the infimum of $\max(\deg'(f_1), \ldots, \deg'(f_\ell))$, where $f = f_1 + \cdots + f_\ell$ and each f_j is a noncommutative monomial.

Remark 3.5. One can check that if f is a product of monomials as above, then $\deg(f) = \deg'(f)$. Moreover, the degree makes TrP_m^0 and TrP_m^1 into graded algebras. Finally, we observe that if $f \in \mathrm{TrP}_m^0$ or TrP_m^1, then the function on $M_N(\mathbb{C})_{\mathrm{sa}}^m$ defined by $x \mapsto f(x)$ is a polynomial in the entries of $x_1, \ldots, x_m$, and the degree of $x \mapsto f(x)$ with respect to the entries is bounded above by the degree of f in TrP_m^0 or TrP_m^1. None of these facts will be used in what follows, so we omit the proofs.

We also observe that there is a composition operation $(\mathrm{TrP}_m^1)^m \times (\mathrm{TrP}_m^1)^m \to (\mathrm{TrP}_m^1)^m$ defined just as one would expect from manipulations in $M_N(\mathbb{C})$. If $f, g \in (\mathrm{TrP}_m^1)^m$, we define $f(g(x))$ by substituting $g_j(x)$ as the j-th argument of f. Then we multiply elements out by treating the terms of the form $\tau(p)$ like scalars. For instance, if $f(Y_1, Y_2) = (\tau(Y_1 Y_2) Y_2, Y_1 + \tau(Y_1^2) Y_2)$ and $g(X_1, X_2) = (\tau(X_1) X_2 + X_1, X_1)$, then $f \circ g(X_1, X_2) = (Z_1, Z_2)$, where

$$Z_1 = \tau([\tau(X_1) X_2 + X_1] X_1) X_1 = \tau(X_1) \tau(X_2 X_1) X_1 + \tau(X_1^2) X_1$$

and

$$
\begin{aligned}
Z_2 &= \tau(X_1) X_2 + X_1 + X_1 \tau[(\tau(X_1) X_2 + X_1)^2] \\
&= \tau(X_1) X_2 + X_1 + \tau[\tau(X_1)^2 X_2^2 + \tau(X_1) X_2 X_1 + \tau(X_1) X_1 X_2 + X_1^2] X_1 \\
&= \tau(X_1) X_2 + X_1 + [\tau(X_1)^2 \tau(X_2^2) + \tau(X_1) \tau(X_2 X_1) + \tau(X_1) \tau(X_1 X_2) + \tau(X_1^2)] X_1 \\
&= \tau(X_1) X_2 + X_1 + \tau(X_1)^2 \tau(X_2^2) X_1 + 2\tau(X_1) \tau(X_2 X_1) X_1 + \tau(X_1^2) X_1.
\end{aligned}
$$

One can check that composition on $(\mathrm{TrP}_m^1)^m$ is well-defined and associative. Moreover, if f and g are self-adjoint elements of $(\mathrm{TrP}_m^1)^m$, then they define functions $M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})_{\mathrm{sa}}^m$, and the element $f \circ g \in (\mathrm{TrP}_m^1)^m$ defined abstractly will produce a function $M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})_{\mathrm{sa}}^m$ which is the composition of the corresponding functions for f and g.

3B. *Differentiation of trace polynomials.* In this section, we give explicit formulas for the gradient and Laplacian of trace polynomials and in particular show that these operations have a well-defined limit as $N \to \infty$ (see [Rains 1997; Cébron 2013; Driver et al. 2013, §3]). We first recall the free difference quotients of [Voiculescu 1998].

Definition 3.6. We define the *free difference quotient* (or simply *noncommutative derivative*) $\mathcal{D}_j : \mathrm{NCP}_m \to \mathrm{NCP}_m \otimes \mathrm{NCP}_m$ by

$$\mathcal{D}_j[X_{i_1} \cdots X_{i_n}] = \sum_{k:i_k=j} X_{i_1} \cdots X_{i_{k-1}} \otimes X_{i_{k+1}} \cdots X_{i_n}.$$

We also define $\mathcal{D}_j : \mathrm{NCP}_m^{\otimes n} \to \mathrm{NCP}_m^{\otimes n+1}$ by

$$\mathcal{D}_j[p_1 \otimes \cdots \otimes p_n] = \sum_{k=1}^n p_1 \otimes \cdots \otimes p_{k-1} \otimes \mathcal{D}_j p_k \otimes p_{k+1} \otimes \cdots \otimes p_n.$$

Then of course $\mathcal{D}_j^k$ is a well-defined map $\mathrm{NCP}_m^{\otimes n} \to \mathrm{NCP}_m^{\otimes n+k}$.

Remark 3.7. We caution the reader that the notation used in Voiculescu's papers is ∂_j rather than $\mathcal{D}_j$. Moreover, the normalization for $\mathcal{D}_j^n f$ here differs from that of [Voiculescu 1998] by a factor of $n!$.

Definition 3.8. We define the *cyclic derivative* $\mathcal{D}_j^\circ : \mathrm{NCP}_m \to \mathrm{NCP}_m$ as the linear map given by

$$\mathcal{D}_j^\circ[X_{i_1} \cdots X_{i_n}] = \sum_{k:i_k=j} X_{i_{k+1}} \cdots X_{i_n} X_{i_1} \cdots X_{i_{k-1}}.$$

Definition 3.9. Given an algebra $\mathcal{A}$ (e.g., NCP_m), we define the *n-th hash operation* as the multilinear map $\mathcal{A}^{\otimes(n+1)} \times \mathcal{A}^{\otimes n} \to \mathcal{A}$ given by

$$(a_0 \otimes \cdots \otimes a_n) \# (b_1 \otimes \cdots \otimes b_n) = a_0 b_1 a_1 \cdots b_n a_n.$$

Example 3.10. Let $X = (X_1, X_2, X_3)$ and define $f(X) = X_1 X_2 X_1^2 X_3 X_2$. Then

$$\mathcal{D}_1 f(X) = 1 \otimes X_2 X_1^2 X_3 X_2 + X_1 X_2 \otimes X_1 X_3 X_2 + X_1 X_2 X_1 \otimes X_3 X_2,$$
$$\mathcal{D}_1^\circ f(X) = X_2 X_1^2 X_3 X_2 + X_1 X_3 X_2 X_1 X_2 + X_3 X_2 X_1 X_2 X_1,$$
$$\mathcal{D}_1 f(X) \# Y = Y X_2 X_1^2 X_3 X_2 + X_1 X_2 Y X_1 X_3 X_2 + X_1 X_2 X_1 Y X_3 X_2.$$

To compute $\mathcal{D}_1^2 f(X) = \mathcal{D}_1[\mathcal{D}_1 f(X)]$, we would add together the three terms

$$\mathcal{D}_1[1 \otimes X_2 X_1^2 X_3 X_2] = 1 \otimes X_2 \otimes X_1 X_3 X_2 + 1 \otimes X_2 X_1 \otimes X_3 X_2,$$
$$\mathcal{D}_1[X_1 X_2 \otimes X_1 X_3 X_2] = 1 \otimes X_2 \otimes X_1 X_3 X_2 + X_1 X_2 \otimes 1 \otimes X_3 X_2,$$
$$\mathcal{D}_1[X_1 X_2 X_1 \otimes X_3 X_2] = 1 \otimes X_2 X_1 \otimes X_3 X_2 + X_1 X_2 \otimes 1 \otimes X_3 X_2.$$

Now we will define several "derivative" operators on the spaces of scalar-valued and noncommutative trace polynomials which will correspond to differentiation with respect to the standard coordinates on $M_N(\mathbb{C})_{\mathrm{sa}}^m$. We begin with the gradient.

To fix notation, recall that in Section 2A we gave a canonical orthonormal basis for $M_N(\mathbb{C})_{\mathrm{sa}}$ with respect to the inner product $\langle x, y \rangle = \mathrm{Tr}(x^* y)$. Using these coordinates, we may identify $M_N(\mathbb{C})_{\mathrm{sa}}$ with $\mathbb{R}^{N^2}$

and hence identify $M_N(\mathbb{C})_{\text{sa}}^m$ with $\mathbb{R}^{mN^2}$. Similarly, we identify the complexification $\mathbb{C} \otimes M_N(\mathbb{C})_{\text{sa}}^m$ with $M_N(\mathbb{C})^m$ and with $\mathbb{C}^{mN^2}$. For $f : M_N(\mathbb{C})_{\text{sa}}^m \to \mathbb{C}$ and $x = (x_1, \ldots, x_m) \in M_N(\mathbb{C})_{\text{sa}}^m$, we denote by $\nabla f(x) \in M_N(\mathbb{C})^m$ the gradient computed in these coordinates; similarly, we denote by $\nabla_j f(x) \in M_N(\mathbb{C})$ the gradient with respect to x_j computed in these coordinates.

Definition 3.11. Define the *j-th gradient operator* $\text{TrP}_m^0 \to \text{TrP}_m^1$ by

$$D_j \left[\prod_{k=1}^n \tau(p_k) \right] = \sum_{k=1}^n \mathcal{D}_j^\circ p_k \prod_{\ell \neq k} \tau(p_\ell). \tag{3-3}$$

Note that D_j is defined so as to obey the Leibniz rule (that is, it is a derivation).

Lemma 3.12. *If $f \in \text{TrP}_m^0$ is viewed as a function $M_N(\mathbb{C})_{\text{sa}}^m \to \mathbb{C}$, then we have*

$$\nabla_j[f(x)] = \frac{1}{N}[D_j f](x). \tag{3-4}$$

Similarly, for $F : M_N(\mathbb{C})_{\text{sa}}^m \to M_N(\mathbb{C})^m$, let $J_j F$ denote the Jacobian linear transformation (a.k.a. Fréchet derivative) with respect to x_j. Then for a noncommutative polynomial p, we have

$$[J_j p(x)](y) = [\mathcal{D}_j p](x) \# y, \tag{3-5}$$

and hence by the product rule for $p \in \text{NCP}_m$ and $f \in \text{TrP}_m^0$, we have

$$[J_j(pf)(x)](y) = ([\mathcal{D}_j p](x) \# y) f(x) + p(x) \tau_N([D_j f](x) y). \tag{3-6}$$

Proof. By standard computations, for a noncommutative polynomial p and $y \in M_N(\mathbb{C})_{\text{sa}}$, we have

$$[J_j p(x)](y) = [\mathcal{D}_j p](x) \# y,$$

$$\nabla_j[\tau_N(p)](x) = \frac{1}{N}[\mathcal{D}_j^\circ p](x).$$

The claims (3-4) and (3-6) now follow from the product rule. $\qquad \square$

Next, we can define the algebraic Laplacian operators on TrP_m^0 and TrP_m^1, which correspond to computing the Laplacian on scalar-valued or vector-valued functions on $M_N(\mathbb{C})_{\text{sa}}^m$, still using the coordinates given in Section 2A.

For $f : M_N(\mathbb{C})_{\text{sa}}^m \to \mathbb{C}$, let $\Delta_j f$ be the Laplacian with respect to the coordinates of the j-th matrix x_j. Note that $\Delta f = \sum_{j=1}^m \Delta_j f$. Similarly, if $f : M_N(\mathbb{C})_{\text{sa}}^m \to M_N(\mathbb{C})$ is an operator-valued function, we define $\Delta_j f$ and Δf by applying Δ_j and Δ entrywise (as is standard notation for the Laplacian of a vector-valued function).

Motivated by (2-2) and the computation in Lemma 3.18 below, we define the map $\eta : \text{NCP}_m^{\otimes 3} \to \text{TrP}_m^1$ by

$$\eta(p_1 \otimes p_2 \otimes p_3) = p_1 p_3 \tau(p_2).$$

Definition 3.13. We define L_j and $L_{N,j} : \text{TrP}_m^0 \to \text{TrP}_m^0$ to be the unique linear operators such that

$$L_j[\tau(p)] = L_{N,j}[\tau(p)] = \tau \circ \eta[\mathcal{D}_j^2 p] \quad \text{for } p \in \text{NCP}_m \tag{3-7}$$

and such that the following product rule is satisfied:

$$L_j[f \cdot g] = L_j[f] \cdot g + f \cdot L_j[g], \tag{3-8}$$

$$L_{N,j}[f \cdot g] = L_{N,j}[f] \cdot g + f \cdot L_{N,j}[g] + \frac{2}{N^2} \tau(D_j f \cdot D_j g). \tag{3-9}$$

Then we define $L = \sum_{j=1}^m L_j$ and $L_N = \sum_{j=1}^m L_{N,j}$.

Remark 3.14. To show the existence of operators $L_{N,j}$ and L_j satisfying (3-7) and the product rule, one can define $L_{N,j}$ more explicitly as the linear operator $\mathrm{TrP}_m^0 \to \mathrm{TrP}_m^0$ given by

$$L_{N,j}[\tau(p_1) \cdots \tau(p_n)] = \sum_{k=1}^n \tau \circ \eta[\mathcal{D}^2 p_k] \cdot \prod_{i \neq k} \tau(p_i) + \frac{1}{N^2} \sum_{k=1}^n \sum_{\ell \neq k} \tau(\mathcal{D}_j^\circ p_k \cdot \mathcal{D}_j^\circ p_\ell) \prod_{i \neq k, \ell} \tau(p_i),$$

and check that this operator is well-defined and satisfies the product rule. Moreover, the uniqueness of the operator $L_{N,j}$ satisfying (3-7) and the product rule follows from the fact that TrP_m^0 is spanned by products of terms of the form $\tau(p)$ for $p \in \mathrm{NCP}_m$. The argument for the existence and uniqueness of L_j is the same.

Example 3.15. Let $X = (X_1, X_2)$. Consider $f(X) = \tau(f_1(X))\tau(f_2(X))$, where $f_1(X) = X_1 X_2 X_1 X_3$ and $f_2(X) = X_2^2 X_1$. Then

$$D_1[\tau(f_1)] = \mathcal{D}_1^\circ f_1 = X_2 X_1 X_3 + X_3 X_1 X_2,$$
$$D_1[\tau(f_2)] = \mathcal{D}_2^\circ f_2 = X_2^2,$$

and

$$L_1[\tau(f_1)] = L_{N,1}[\tau(f_1)] = \tau \circ \eta[\mathcal{D}_1^2 f_1] = \tau[\eta[1 \otimes X_2 \otimes X_3]] = \tau[1 \cdot X_3] \cdot \tau[X_2],$$
$$L_1[\tau(f_2)] = L_{N,1}[\tau(f_2)] = 0.$$

Therefore, we have

$$L_1[f] = L_1[\tau(f_1)]\tau(f_2) + \tau(f_1)L_1[\tau(f_2)] = \tau(X_3)\tau(X_2)\tau(X_2^2 X_1) + 0,$$

$$L_{N,1}[f] = L_{N,1}[\tau(f_1)]\tau(f_2) + \tau(f_1)L_{N,1}[\tau(f_2)] + \frac{2}{N^2} \tau[\mathcal{D}_1^\circ f_1 \mathcal{D}_1^\circ f_2]$$
$$= \tau(X_3)\tau(X_2)\tau(X_2^2 X_1) + \frac{2}{N^2} \tau[(X_2 X_1 X_3 + X_3 X_1 X_2)X_2^2].$$

One can carry out a similar computation for $L_2[f]$ and $L_{N,2}[f]$ and thus find $L[f]$ and $L_{N,2}[f]$.

Since we will also deal with the Laplacians of matrix-valued functions on matrices, we also need to define the algebraic Laplacian on *operator-valued* trace polynomials.

Definition 3.16. We also define L_j and $L_{N,j} : \mathrm{TrP}_m^1 \to \mathrm{TrP}_m^1$ to be the unique linear operators on the space of operator-valued trace polynomials such that

$$L_j[p] = L_{N,j}[p] = \eta[\mathcal{D}_j^2 p] \quad \text{for } p \in \mathrm{NCP}_m \tag{3-10}$$

and the following product rule is satisfied for $p \in \mathrm{NCP}_m$ and $f \in \mathrm{TrP}_m^0$:

$$L_j[p \cdot f] = L_j[p] \cdot f + p \cdot L_j[f], \tag{3-11}$$

$$L_{N,j}[p \cdot f] = L_{N,j}[p] \cdot f + p \cdot L_{N,j}[f] + \frac{2}{N^2} \mathcal{D}_j p \,\#\, D_j f, \tag{3-12}$$

where $L_j[f]$ and $L_{N,j}[f]$ are given by Definition 3.13. Then we define $L = \sum_{j=1}^m L_j$ and $L_N = \sum_{j=1}^m L_{N,j}$.

Remark 3.17. The argument for the existence and uniqueness of the operators L_j and $L_{N,j}$ on TrP_m^1 is similar to the argument for TrP_m^0, only it relies on the previous scalar-valued case since the scalar-valued case was used in the product rule.

Lemma 3.18. *Let $f \in \mathrm{TrP}_m^0$. Viewing f as a function $M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{C}$, we have*

$$\Delta_j f(x) = N[L_{N,j} f](x), \quad \Delta f(x) = N[L_N f](x). \tag{3-13}$$

The same formula holds if $f \in \mathrm{TrP}_m^1$ and f is viewed as a function $M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})$.

Proof. We begin with the special case of computing the Laplacian of $p \in \mathrm{NCP}_m$ (as a *matrix-valued* function). To differentiate, we use the basis $\mathcal{B}_N$ given by (2-1). Note that

$$
\begin{aligned}
\Delta_j p(x) &= \sum_{b \in \mathcal{B}_N} \frac{d^2}{dt^2}\Big|_{t=0} f(x_1, \ldots, x_{j-1}, x_j + tb, x_{j+1}, \ldots, x_m) \\
&= \sum_{b \in \mathcal{B}_N} \mathcal{D}_j^2 p(x) \,\#\, (b \otimes b) = N[\eta(\mathcal{D}_j^2 p)](x) = [L_{N,j} p](x),
\end{aligned}
$$

where the second-to-last equality follows from (2-2).

Next, we consider the case of computing the Laplacian of $\tau_N(p)$ (as a *scalar-valued* function) for $p \in \mathrm{NCP}_m$. Since τ_N is a linear map $M_N(\mathbb{C}) \to \mathbb{C}$, we have

$$\Delta_j[\tau_N(p(x))] = \tau_N(\Delta_j p(x)),$$

where the Laplacian Δ_j on the left-hand side is applied to a scalar-valued function and on the right-hand side it is applied to a matrix-valued function. Therefore, it follows from the previous computation that

$$\Delta_j[\tau_N(p(x))] = N\tau_N([\eta(\mathcal{D}_j^2 p)](x)) = [L_{N,j}[\tau(p)]](x).$$

For the general case of scalar-valued trace polynomials, recall that the vector space of trace polynomials is spanned by elements of the form $f = \tau(p_1) \cdots \tau(p_N)$, where $p_j \in \mathrm{NCP}_m^0$. Let $f_j = \tau(p_j) \in \mathrm{TrP}_m^0$. The Laplacian Δ_j of a product of functions can be computed using the product rule of differentiation as

$$\Delta_j f(x) = \sum_{j=1}^n N[L_{N,j} f_k](x) \prod_{i \neq k} \tau_N(f_i(x)) + \sum_{k=1}^n \sum_{\ell \neq k} \mathrm{Tr}(\nabla_j f_k(x) \nabla_j f_\ell(x)) \prod_{i \neq k, \ell} f_i(x).$$

The special case proved above shows that $\Delta_j[f_k(x)] = N[L_{N,j} f](x)$. Moreover, by (3-4), we have $\nabla_j[f_k(x)] = (1/N)[D_j f_k](x)$. Thus, we have

$$\Delta_j f(x) = \sum_{k=1}^n N[L_{N,j} f_k](x) \prod_{i \neq k} f_i(x) + \frac{1}{N} \sum_{k=1}^n \sum_{\ell \neq k} \tau_N([D_j f_k](x)[D_j f_\ell](x)) \prod_{i \neq k, \ell} f_i(x).$$

Because of the product rule in the definition of $L_{N,j}$, the right-hand side equals $N[L_{N,j}f](x)$. This completes the proof of (3-13) in the scalar-valued case. The proof for the operator-valued case is similar, using the cases proved above, as well as (3-4) and (3-6). $\square$

Corollary 3.19. *Let* $f \in \mathrm{TrP}_m^0$ *or* TrP_m^1. *If we view* f *as a function on* $M_N(\mathbb{C})_{\mathrm{sa}}^m$, *then* $(1/N)\Delta f$ *is a trace polynomial of lower degree than* f, *and we have coefficientwise*

$$\lim_{N \to \infty} \frac{1}{N}\Delta f(x) = \lim_{N \to \infty} L_N f(x) = Lf(x).$$

Remark 3.20. We have shown that if f is a scalar-valued trace polynomial, then viewed as a map $M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{C}$, we have

$$Du = N\nabla f, \quad L_N f = \frac{1}{N}\Delta f.$$

Therefore, in the rest of the paper, we will freely write Df and $L_N f$ for $N\nabla f$ and $(1/N)\Delta f$ for general functions $f : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{C}$. The same considerations apply to the Laplacian for operator-valued trace polynomials, viewed as maps $M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})$.

3C. *Convolution of trace polynomials and gaussians.* Let $f \in \mathrm{TrP}_m^0$ or $f \in \mathrm{TrP}_m^1$. Then viewing f as a function defined on $M_N(\mathbb{C})_{\mathrm{sa}}^m$, we may define the convolution of f with the probability measure $\sigma_{t,N}$ (the law of an m-tuple of independent GUE). This is equivalent to the classical convolution of f with the function $M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$ giving the density of the measure $\sigma_{t,N}$. Moreover, $f_t = f * \sigma_{t,N}$ is the solution to the heat equation with initial condition f, or more precisely

$$\partial_t f_t = \frac{1}{2N}\Delta f_t.$$

(The integral formula for the solution to the heat equation with the Laplacian Δ is well known [Evans 2010, §2.3], and to solve the equation with $(2N)^{-1}\Delta$ one renormalizes time by a factor of $(2N)^{-1}$, and this corresponds precisely to our normalizations in the definition of $\sigma_{N,t}$. We leave this computation to the reader.)

We showed in the last subsection that $L_N = (1/N)\Delta$ on trace polynomials is given by a purely algebraic computation. Moreover, examining the construction of L_N, one can see that it maps trace polynomials of degree $\leq d$ to trace polynomials of degree $\leq d$. We can view L_N and L as linear transformations on the finite-dimensional vector space of trace polynomials of degree $\leq d$ and define $\exp(tL_N/2)$ and $\exp(tL/2)$ by the matrix exponential.

Because this holds for any d, we know that $\exp(tL_N/2)$ and $\exp(tL/2)$ define linear transformations $\mathrm{TrP}_m^0 \to \mathrm{TrP}_m^0$ and $\mathrm{TrP}_m^1 \to \mathrm{TrP}_m^1$. Moreover, a standard computation shows that $f_t = \exp(tL_N/2)f$ satisfies the heat equation $\partial_t f_t = L_N f_t/2$. These observations, together with Corollary 3.19 yield the following.

Lemma 3.21. *Let* f *be a trace polynomial in* TrP_m^0 *or* TrP_m^1. *Then we have*

$$\sigma_{t,N} * f(x) = \left[\exp\left(\frac{tL_N}{2}\right)f\right](x), \tag{3-14}$$

with $\deg(\exp(tL_N/2)f) \leq \deg(f)$, *and we have*

$$\lim_{N \to \infty} \exp\left(\frac{tL_N}{2}\right)f = \exp\left(\frac{tL}{2}\right)f \quad \textit{coefficientwise}. \tag{3-15}$$

Example 3.22. Let $X = (X_1, \ldots, X_m)$ and define $f(X) = \sum_{j=1}^m X_j^2$. Note that $\mathcal{D}_j^2[f(X)] = 2(1 \otimes 1 \otimes 1)$ for each j, and hence $L[\tau(f)] = 2m = L_N[\tau(f)]$. We also have $\mathcal{D}_j^\circ f = 2X_j$. Hence,

$$L[\tau(f)^2] = 2L[\tau(f)]\tau(f) = 4m\tau(f),$$

$$L_N[\tau(f)^2] = 2L[\tau(f)]\tau(f) + 2\sum_{j=1}^m \tau(\mathcal{D}_j^\circ f \cdot \mathcal{D}_j^\circ f) = 4m\tau(f) + \frac{8m}{N^2}\tau(f).$$

Therefore, $(L/2)[\tau(f)^2] = 2m\tau(f)$ and $(L/2)[\tau(f)] = m$. Thus, the span of $\tau(f)^2$, $\tau(f)$, and 1 is invariant under the operator $(L/2)$, and $(L/2)$ is given by a nilpotent matrix on this subspace. Direct computation then shows that

$$e^{-tL/2}[\tau(f)^2] = \tau(f)^2 + 2mt\tau(f) + m^2 t^2.$$

A similar computation shows that

$$e^{-tL_N/2}[\tau(f)^2] = \tau(f)^2 + 2m\left(1 + \frac{2}{N^2}\right)t\tau(f) + m^2\left(1 + \frac{2}{N^2}\right)\frac{t^2}{2}.$$

Thus, as $N \to +\infty$, we have $e^{-tL_N/2}[\tau(f)^2] \to e^{-tL/2}[\tau(f)^2]$.

The probabilistic interpretation of $f * \sigma_{t,N} = \exp(tL_N/2)f$, which follows from a standard computation, is that $\sigma_{t,N} * f(x)$ is the expectation of $f(x + t^{1/2}Y)$, where Y is an m-tuple of independent GUE of variance 1. Moreover, for every probability measure μ on $M_N(\mathbb{C})_{\mathrm{sa}}^m$ with finite moments, we have

$$\int f(x)\, d(\mu * \sigma_{t,N})(x) = \int (\sigma_{t,N} * f)(x)\, d\mu(x) = \int \left[\exp\left(\frac{tL_N}{2}\right)f\right](x)\, d\mu(x). \qquad (3\text{-}16)$$

In the free setting, the operator $\exp(tL/2)$ has a similar relationship with the free convolution with σ_t. This fact is standard in free probability, but because we need it for Lemmas 3.28 and 7.4 below, we include a sketch of the proof here.

Lemma 3.23. *Let $\lambda \in \Sigma_{n,R}$ be a noncommutative law. Then for any trace polynomial $f \in \mathrm{TrP}_m^0$, we have*

$$\lambda \boxplus \sigma_t(f) = \lambda\left(\exp\left(\frac{tL}{2}\right)f\right). \qquad (3\text{-}17)$$

Proof. Because free convolution with σ_t forms a semigroup and $\exp(tL/2)$ is also a semigroup, it suffices to prove that

$$\frac{d}{dt}\bigg|_{t=0} \lambda \boxplus \sigma_t(f) = \frac{\lambda}{2}(Lf).$$

By the product rule, it suffices to handle the case of $f = \tau(p)$ for $p \in \mathrm{NCP}_m$ by showing that

$$\frac{d}{dt}\bigg|_{t=0} \lambda \boxplus \sigma_t(p) = \frac{\lambda}{2}(\eta(\mathcal{D}_j^2 p)).$$

Let $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_m)$ be freely independent m-tuples of self-adjoint elements of a tracial von Neumann algebra $(\mathcal{M}, \tau)$, such that the law of x is λ and the law of y is σ_1. We want to

compute $(d/dt)|_{t=0}\tau(p(x+t^{1/2}y))$. But note that

$$p(x+t^{1/2}y) = p(x) + t^{1/2}\sum_{j=1}^{m}\mathcal{D}_j p(x)\,\#\,y_j + \frac{t}{2}\sum_{j,k=1}^{m}\mathcal{D}_j\mathcal{D}_k p(x)\,\#\,(y_j\otimes y_k) + O(t^{3/2}).$$

A moment computation with free independence shows that the terms of order $t^{1/2}$ have expectation zero, and so do the terms of order t with $j\neq k$. We are left with

$$\frac{d}{dt}\Big|_{t=0}\tau(p(x+t^{1/2}y)) = \frac{1}{2}\sum_{j=1}^{n}\tau(\mathcal{D}_j^2 p(x)\,\#\,(y_j\otimes y_j)),$$

which using freeness evaluates to $\frac{1}{2}\sum_{j=1}^{n}\tau(\eta(\mathcal{D}_j^2 p(x))) = \tau(Lp(x)/2)$. $\square$

3D. *Asymptotic approximation by trace polynomials.*
Now we are ready to define the approximation property which captures the asymptotic behavior of functions on $M_N(\mathbb{C})_{\mathrm{sa}}^m$.

Definition 3.24. A sequence of functions $\phi_N : M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})^m$ is said to be *asymptotically approximable by trace polynomials* if for every $\epsilon > 0$ and $R > 0$ there exists some $f \in (\mathrm{TrP}_m^1)^m$ (an m-tuple of operator-valued trace polynomials) such that

$$\limsup_{N\to\infty}\ \sup_{\|x\|_\infty\le R}\ \|\phi_N(x) - f(x)\|_2 \le \epsilon.$$

In this case, we call f an (ϵ, R)-*approximation of* $\{\phi_N\}$. We make the same definitions for functions $\phi_N : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{C}$, except that we use scalar-valued trace polynomials (elements of TrP_m^0) and apply the absolute value rather than the 2-norm.

Observation 3.25. If $f \in (\mathrm{TrP}_m^1)^m$ and if f_N denotes the map $M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})^m$ given by $x \mapsto f(x)$, then f_N is asymptotically approximable by trace polynomials. Also, asymptotically approximable sequences form a vector space over $\mathbb{C}$.

Observation 3.26. Let $\{\phi_N^{(\ell)}\}_{N,\ell\in\mathbb{N}}$ be a sequence of functions where $\phi_N^{(\ell)} : M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})^m$. Suppose that $\{\phi_N\}$ is another sequence such that for every $R > 0$

$$\lim_{\ell\to\infty}\limsup_{N\to\infty}\ \sup_{\|x\|_\infty\le R}\ \|\phi_N^{(\ell)}(x) - \phi_N(x)\|_2 = 0.$$

If $\{\phi_N^{(\ell)}\}_{N\in\mathbb{N}}$ is asymptotically approximable by trace polynomials for each ℓ, then so is $\{\phi_N\}_{N\in\mathbb{N}}$. The same holds in the case of scalar-valued functions and scalar-valued trace polynomials.

Lemma 3.27. *Let* $\phi_N, \psi_N : M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})_{\mathrm{sa}}^m$. *Suppose that* $\{\phi_N\}$ *and* $\{\psi_N\}$ *are both asymptotically approximable by trace polynomials, and furthermore suppose that* $\{\phi_N\}_{N\in\mathbb{N}}$ *is uniformly Lipschitz in* $\|\cdot\|_2$, *that is, for some* $K > 0$,

$$\|\phi_N(x) - \phi_N(y)\|_2 \le K\|x - y\|_2 \quad \textit{for all } x, y, \textit{ for all } N.$$

Then $\{\phi_N \circ \psi_N\}$ *is asymptotically approximable by trace polynomials.*

Proof. It is straightforward to check that if $\{\phi_N\}$ is a sequence of functions that map self-adjoint tuples to *self-adjoint* tuples and if $(f_1, \ldots, f_m)$ is an (ϵ, R)-approximation of $\{\phi_N\}$, then so is $\frac{1}{2}(f_1 + f_1^*, \ldots, f_m + f_m^*)$. Thus, we may assume without loss of generality that the operator-valued trace polynomials used in our approximations for $\{\phi_N\}$ and $\{\psi_N\}$ are self-adjoint, so it makes sense to compose them with ϕ_N or ψ_N.

Choose $\epsilon > 0$ and $R > 0$. Choose an m-tuple of self-adjoint trace polynomials g which is an $(\epsilon/(2K), R)$-approximation of $\{\psi_N\}$. Since g is a trace polynomial, there exists some $R' > 0$ such that for any tuple x of self-adjoint matrices of any size, we have

$$\|x\|_\infty \leq R \quad \implies \quad \|g(x)\|_\infty \leq R'.$$

Now because ϕ_N is asymptotically approximable by trace polynomials, we can choose a polynomial f which is an $(\epsilon/2, R')$-approximation of $\{\phi_N\}$. Now we observe that when $\|x\|_\infty \leq R$ (hence $\|g(x)\|_\infty \leq R'$), we have

$$\|\phi_N \circ \psi_N(x) - f \circ g(x)\|_2 \leq \|\phi_N \circ \psi_N(x) - \phi_N \circ g(x)\|_2 + \|\phi_N \circ g(x) - f \circ g(x)\|_2$$
$$\leq K \sup_{\|x\|_\infty \leq R} \|\psi_N(x) - g(x)\|_2 + \sup_{\|y\|_\infty \leq R'} \|\phi_N(y) - f(y)\|_2.$$

Therefore,

$$\limsup_{N \to \infty} \sup_{\|x\|_\infty \leq R} \|\phi_N \circ \psi_N(x) - f \circ g(x)\|_2 \leq K \cdot \frac{\epsilon}{2K} + \frac{\epsilon}{2} = \epsilon. \qquad \square$$

Lemma 3.28. *Suppose that $\phi_N : M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})_{\mathrm{sa}}^m$ is asymptotically approximable by trace polynomials and that*

$$\|\phi_N(x)\|_2 \leq A\left(1 + \sum_j \tau_N(x_j^{2n})\right) \tag{3-18}$$

for some $A > 0$ and some integer $n \geq 0$. If $\{\phi_N\}$ is asymptotically approximable by trace polynomials, then so is $\{\phi_N * \sigma_{t,N}\}$.

Proof. Fix $R > 0$ and $\epsilon > 0$. Choose a trace polynomial f which is an $(\epsilon, R + 3t^{1/2})$ approximation for $\{\phi_N\}$. Now for x with $\|x\|_\infty \leq R$, we estimate

$$\|\sigma_{t,N} * \phi_N(x) - \sigma_{t,N} * f(x)\|_2 \leq \int \|\phi_N(x + y) - f(x + y)\|_2 \, d\sigma_{t,N}(y).$$

We break this integral into two pieces: The integral over the region where $\|y\|_\infty \leq 3t^{1/2}$ is bounded by ϵ as $N \to \infty$ by our choice of f. Furthermore, we claim that the integral over the region where $\|y\|_\infty > 3t^{1/2}$ vanishes as $N \to \infty$. Using assumption (3-18) and the fact that f is a trace polynomial, we see that there exists a $C > 0$ and integer $d > 0$, depending only on R, A, n, and f, such that

$$\sup_{\|x\|_\infty \leq R} [\|\phi_N(x + y)\|_2 + \|f(x + y)\|_2] \leq C\left(1 + \sum_j \tau_N(y_j^{2d})\right).$$

Therefore, we have

$$\int_{\|y\|_\infty \geq 3t^{1/2}} \|\phi_N(x + y) - f(x + y)\|_2 \, d\sigma_{t,N}(y) \leq C \int_{\|y\|_\infty \geq 3t^{1/2}} \left(1 + \sum_j \tau_N(y_j^{2d})\right) d\sigma_{t,N}(y).$$

This vanishes as $N \to \infty$ by Corollary 2.12 applied to the GUE. Therefore, we have

$$\limsup_{N \to \infty} \sup_{\|x\|_\infty \leq R} \|\sigma_{t,N} * \phi_N(x) - \sigma_{t,N} * f(x)\|_2 \leq \epsilon.$$

On the other hand, by Lemma 3.21, we have $\sigma_{t,N} * f = \exp(t L_N / 2) f \to \exp(t L / 2) f$ coefficientwise, and therefore,

$$\limsup_{N \to \infty} \sup_{\|x\|_\infty \leq R} \|\sigma_{t,N} * f(x) - \left[\exp\left(\frac{tL}{2}\right) f\right](x)\|_2 = 0,$$

so that

$$\limsup_{N \to \infty} \sup_{\|x\|_\infty \leq R} \|\sigma_{t,N} * \phi_N(x) - \left[\exp\left(\frac{tL}{2}\right) f\right](x)\|_2 \leq \epsilon. \qquad \square$$

Lemma 3.29. *Suppose that $\phi_N : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{C}$ and suppose that $\{D\phi_N\} = \{N \nabla \phi_N\}$ is asymptotically approximable by trace polynomials and that $\phi_N(0) = 0$. Then $\{\phi_N\}$ is asymptotically approximable by trace polynomials.*

Proof. Given a trace polynomial $F \in (\mathrm{TrP}_m^1)^m$, we can define

$$f(X) = \int_0^1 \tau(F(tX)X) \, dt$$

in TrP_m^0. Then we have

$$\sup_{\|x\|_\infty \leq R} |\phi_N(x) - f(x)| = \sup_{\|x\|_\infty \leq R} \left| \int_0^1 \langle D\phi_N(tx) - F(tx), x \rangle_2 \, dt \right|$$

$$\leq R \sup_{\|y\|_\infty \leq R} \|N \nabla \phi_N(y) - F(y)\|_2. \qquad \square$$

4. Convergence of moments

Our goal in this section is prove the following theorem. The convergence of moments is related to [Guionnet and Shlyakhtenko 2009, Theorem 4.4; Dabrowski et al. 2016, Proposition 50 and Theorem 51; Dabrowski 2016, Theorem 4.4], and we include versions of standard concentration estimates (see Section 2E) in the statement.

Theorem 4.1. *Let $V_N : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$ be a sequence of potentials such that $V_N(x) - (c/2)\|x\|_2^2$ is convex and $V_N(x) - (C/2)\|x\|_2^2$ is concave. Let μ_N be the associated measure. Suppose that the sequence $\{DV_N\}$ is asymptotically approximable by trace polynomials, and assume that*

$$M = \limsup_{N \to \infty} \max_j \left\| \int (x_j - \tau_N(x_j) 1) \, d\mu_N(x) \right\| < +\infty, \tag{4-1}$$

where 1 denotes the $N \times N$ identity matrix.

(1) *We have the following bounds on the operator norm: if $R_N = \max_j \int \|x_j\| \, d\mu_N(x)$, then*

$$\limsup_{N \to \infty} R_N \leq \frac{2}{c^{1/2}} + \frac{1}{c} \limsup_{N \to \infty} \max_j \left| \int \tau_N(x_j) \, d\mu_N(x) \right| + M$$

$$\leq \frac{2}{c^{1/2}} + \frac{1}{c} \limsup_{N \to \infty} \|DV_N(0)\|_2 + \frac{C - c}{2c^{3/2}} + M,$$

and as a consequence of concentration we have for each j that

$$\mu_N(\|x_j\| \geq R_N + \delta) \leq e^{-cN\delta^2/2}.$$

(2) *There exists a noncommutative law $\lambda \in \Sigma_{m,R_*}$, where $R_* = \limsup_{N\to\infty} R_N$, such that for every noncommutative polynomial p*

$$\lim_{N\to\infty} \int \tau_N(p(x)) \, d\mu_N(x) = \lambda(p).$$

(3) *The sequence $\{\mu_N\}$ exhibits exponential concentration around λ in the sense that, for every $R > 0$ and every neighborhood $\mathcal{U}$ of λ in Σ_m,*

$$\limsup_{N\to\infty} \frac{1}{N^2} \log \mu_N(x \in M_N(\mathbb{C})_{\mathrm{sa}}^m : \|x\|_\infty \leq R, \lambda_x \notin \mathcal{U}) < 0.$$

Remark 4.2. The rather artificial hypothesis that $\limsup_{N\to\infty} \max_j \left\| \int (x_j - \tau_N(x_j)) \, d\mu_k(x) \right\| < +\infty$ is trivially satisfied if either μ_N has expectation zero or μ_N is invariant under unitary conjugation and hence $\int x_j \, d\mu_N(x)$ is equal to $\int \tau_N(x_j) \, d\mu_N(x)$ times the identity matrix.

We have already seen in Section 2E that concentration estimates and operator norm tail bounds are standard. To prove that the moments converge, something more is needed; indeed, the only assumption relating the measures μ_N for different values of N is the fact that DV_N is asymptotically approximable by trace polynomials. But even if DV_N is given by the same "trace analytic-function" for different values of N, it is not immediate that the measure would concentrate in the same regions for matrices of different sizes.

To prove convergence of moments, we want to express $\int u \, d\mu_N$ in terms of DV_N for a Lipschitz function u. One of the standard techniques is to show μ_N is the unique stationary distribution for a process X_t that satisfies the SDE

$$dX_t = dY_t - \frac{DV_N(X_t)}{2} \, dt, \tag{4-2}$$

where Y_t is a GUE Brownian motion. This machinery lies behind the log-Sobolev inequality and concentration results, as well as earlier theorems about convergence of moments for general convex potentials.

Specifically, Dabrowski, Guionnet, and Shylakhtenko [Dabrowski et al. 2016, Proposition 5] used the free version of this SDE to show that for a noncommutative potential V satisfying certain convexity assumptions, there exists a free Gibbs law for V which is the unique stationary distribution. As an application, they showed convergence of moments for random matrix models given by $V_N = V$ [Dabrowski et al. 2016, Proposition 50 and Theorem 51], essentially a special case of our Theorem 4.1.

Dabrowski [2016, Theorem 4.4] was able to show convergence of moments under weaker convexity assumptions by constructing the solution to the free SDE as an ultralimit of the finite-dimensional solutions. Our theorem has convexity assumptions similar to Dabrowski's, but we consider a more general sequence of potentials V_N. Like Dabrowski, we analyze the free case by taking the limit of finite-dimensional results, but we use deterministic rather than stochastic methods.

Instead of the solving the SDE, we study the associated semigroup $T_t^{V_N}$, acting on Lipschitz functions u, given by

$$T_t^{V_N} u(x) = E_x[u(X_t)],$$

where X_t is the process solving the SDE (4-2) with initial condition x. The semigroup provides the solution to a certain PDE; that is, if $u(x, t) = T_t u_0(x)$, then we have

$$\partial_t u = \frac{1}{2N} \Delta u - \frac{N}{2} \nabla V_N \cdot \nabla u = \frac{L_N u}{2} - \frac{\langle DV_N, Du \rangle_2}{2}.$$

The semigroup $T_t^{V_N}$ will decrease the Lipschitz norms of functions and thus, if u is Lipschitz, then $T_t^{V_N} u$ will converge to $\int u \, d\mu_N$ as $t \to \infty$.

Solving the differential equation and taking $t \to \infty$ provides a way to evaluate $\int u \, d\mu_N$ in terms of DV_N. We will describe a construction of the semigroup T_t^V through iterating simpler operations (Section 4A), and then we will show (Lemma 4.10) that the iteration procedure preserves approximability by trace polynomials and hence conclude that $\lim_{N \to \infty} \int u \, d\mu_N$ exists.

4A. *Iterative construction of the semigroup.* To simplify notation in this section, we fix N and fix a potential $V : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$ such that $V(x) - (c/2)\|x\|_2^2$ is convex and $V(x) - (C/2)\|x\|_2^2$ is concave for some $0 < c \leq C$. We also write T_t rather than T_t^V.

We will construct T_t by combining two simpler semigroups corresponding to the stochastic and deterministic terms of $dY_t - (DV/2)(X_t) \, dt$. Recall that the solution to the heat equation $\partial_t u = (2N)^{-1} \Delta u$ with initial data u_0 is given by the heat semigroup:

$$P_t u_0(x) = \int u_0(x + y) \, d\sigma_{t,N}(y).$$

Meanwhile, the solution to $\partial_t u = -\frac{1}{2} \langle DV, Du \rangle_2$ with initial data u_0 is given by

$$S_t u_0(x) = u_0(W(x, t)),$$

where $W(x, t)$ is the solution to the ODE

$$\partial_t W(x, t) = -\tfrac{1}{2} DV(W(x, t)), \quad W(x, 0) = x. \tag{4-3}$$

We want to define $T_t = \lim_{n \to \infty} (P_{t/n} S_{t/n})^n$. This is motivated by Trotter's product formula which asserts that $e^{t(A+B)} = \lim_{n \to \infty} (e^{tA/n} e^{tB/n})^n$ for nice enough self-adjoint operators A and B (see [Trotter 1959; Kato 1978; Simon 1979, pp. 4–6]). In our case, we must show that $(P_{t/n} S_{t/n})^n$ converges as $n \to \infty$ and derive dimension-independent error bounds.

We use the following basic properties of the semigroups P_t and S_t. Here if $u : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{C}$, then $\|u\|_{\mathrm{Lip}}$ denotes the Lipschitz norm with respect to the *normalized L^2 metric* $\| \cdot \|_2$ on $M_N(\mathbb{C})_{\mathrm{sa}}^m$ and $\|u\|_{L^\infty}$ denotes the standard L^∞ norm. We are only concerned with Lipschitz functions, so in the following estimates, the reader may always assume u is Lipschitz, but of course $\|u\|_{L^\infty}$ may be infinite for Lipschitz functions.

Lemma 4.3. (1) $\|P_t u\|_{L^\infty} \leq \|u\|_{L^\infty}$.

(2) $\|P_t u\|_{\mathrm{Lip}} \leq \|u\|_{\mathrm{Lip}}$.

(3) $\|P_t u - u\|_{L^\infty} \leq m^{1/2} t^{1/2} \|u\|_{\mathrm{Lip}}$.

Proof. (1) and (2) follow from the fact that $P_t u$ is u convolved with a probability measure. To prove (3), suppose $\|u\|_{\mathrm{Lip}} < +\infty$. Then

$$|P_t u(x) - u(x)| = \left| \int (u(x + y) - u(x))\, d\sigma_{t,N}(y) \right|$$

$$\leq \int |u(x + y) - u(x)|\, d\sigma_{t,N}(y)$$

$$\leq \|u\|_{\mathrm{Lip}} \int \|y\|_2\, d\sigma_{t,N}(y).$$

Meanwhile,

$$\int \|y\|_2\, d\sigma_{t,N}(y) \leq \left(\int 1\, d\sigma_{t,N}(y) \right) \left(\int \|y\|_2^2\, d\sigma_{t,N}(y) \right)^{1/2} = (mt)^{1/2},$$

since y is an m-tuple $(y_1, \ldots, y_m)$ and $\int \tau_N(y_j^2)\, d\sigma_{t,N}(y) = t$ for each j. $\qquad \square$

Lemma 4.4. (1) *The solution $W(x, t)$ to* (4-3) *exists for all t.*

(2) $\|W(x, t) - W(y, t)\|_2 \leq e^{-ct/2}\|x - y\|_2$.

(3) $\|W(x, t) - x\|_2 \leq (t/2)\|DV(x)\|_2$.

(4) $\|(W(x, t) - x) - (W(y, t) - y)\|_2 \leq (C/c)(1 - e^{-ct/2})\|x - y\|_2$.

(5) $\|S_t u\|_{\mathrm{Lip}} \leq e^{-ct/2}\|u\|_{\mathrm{Lip}}$.

(6) $\|S_t u\|_{L^\infty} \leq \|u\|_{L^\infty}$ *if u is continuous.*

Proof. (1) The convexity and semiconcavity assumptions on V imply that DV is C-Lipschitz (see Proposition 2.13(4)) and therefore global existence of the solution follows from the Picard–Lindelöf theorem.

(2) Let $\widetilde{V}(x) = V(x) - (c/2)\|x\|_2^2$. By Proposition 2.13(5),

$$\langle DV(x) - DV(y), x - y \rangle_2 \geq c\|x - y\|_2^2.$$

Now observe that

$$\frac{d}{dt}\|W(x, t) - W(y, t)\|_2^2 = -\langle DV(W(x, t)) - DV(W(y, t)), W(x, t) - W(y, t) \rangle_2$$

$$\leq -c\|W(x, t) - W(y, t)\|_2^2,$$

and hence by Grönwall's inequality, $\|W(x, t) - W(y, t)\|_2^2 \leq e^{-ct}\|W(x, 0) - W(y, 0)\|_2^2 = e^{-ct}\|x - y\|_2^2$.

(3) Note that

$$\frac{d}{dt}\|W(x, t) - x\|_2^2 = -\langle DV(W(x, t)), W(x, t) - x \rangle_2$$

$$= -\langle DV(W(x, t)) - DV(x), W(x, t) - x \rangle_2 - \langle DV(x), W(x, t) - x \rangle_2$$

$$\leq \|DV(x)\|_2 \|W(x, t) - x\|_2.$$

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, we have

$$\frac{d}{dt}\|W(x,t) - x\|_2 \leq \tfrac{1}{2}\|DV(x)\|_2,$$

which proves (3).

(4) We observe that

$$\|(W(x,t) - x) - (W(y,t) - y)\|_2 \leq \frac{1}{2}\int_0^t \|DV(W(x,s)) - DV(W(y,s))\|_2\,ds$$

$$\leq \frac{C}{2}\int_0^t \|W(x,s) - W(y,s)\|_2\,ds$$

$$\leq \frac{C}{2}\int_0^t e^{-cs/2}\|x - y\|_2\,ds = \frac{C}{c}(1 - e^{-ct/2})\|x - y\|_2.$$

(5) This follows from (2).

(6) This is immediate because $S_t u$ is u precomposed with another function. ∎

Now we combine P_t and S_t as in Trotter's formula, except that for technical convenience we define our approximations using dyadic time intervals rather than subdividing $[0, t]$ into intervals of size t/n. We set $\mathbb{N} = \{1, 2, 3, \dots\}$ and $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. We also denote by $\mathbb{Q}_2^+ = \bigcup_{\ell \geq 0} 2^{-\ell}\mathbb{N}_0$ the nonnegative dyadic rationals.

Lemma 4.5. *For $t \in 2^{-\ell}\mathbb{N}_0$, define*

$$T_{t,\ell}u = (P_{2^{-\ell}}S_{2^{-\ell}})^{2^{\ell}t}u.$$

For $t \in \mathbb{Q}_2^+$, the limit $T_t u := \lim_{\ell \to \infty} T_{t,\ell}u$ exists and we have

$$\|T_{t,\ell}u - T_t u\|_{L^\infty} \leq \frac{Cm^{1/2}}{c(2 - 2^{1/2})}2^{-\ell/2}\|u\|_{\mathrm{Lip}}.$$

We also have $\|T_t u\|_{\mathrm{Lip}} \leq e^{-ct/2}\|u\|_{\mathrm{Lip}}$.

Proof. We want to show that the sequence $\{T_{t,\ell}u\}_\ell$ is Cauchy by estimating the difference between consecutive terms. Suppose that $t \in 2^{-\ell}\mathbb{N}_0$ and write $t = n/2^\ell$ and $\delta = 2^{-\ell-1}$. Note the telescoping series identity

$$T_{t,\ell+1}u - T_{t,\ell}u = \sum_{j=0}^{n-1}(P_\delta S_\delta)^{2j}P_\delta(S_\delta P_\delta - P_\delta S_\delta)S_\delta(P_{2\delta}S_{2\delta})^{n-1-j}u. \tag{4-4}$$

Thus, we want to estimate $S_\delta P_\delta - P_\delta S_\delta$ and then control the propagation of the errors through the applications of the other operators. Note that for a Lipschitz function v, we have using Lemma 4.4(4) that

$$|S_\delta P_\delta v(x) - P_\delta S_\delta v(x)| \leq \int |v(W(x,\delta) + y) - v(W(x+y,\delta))|\,d\sigma_{\delta,N}(y)$$

$$\leq \|v\|_{\mathrm{Lip}}\int \|(W(x,\delta) - x) - (W(x+y,\delta) - (x+y))\|_2\,d\sigma_{\delta,N}(y)$$

$$\leq \|v\|_{\mathrm{Lip}}\frac{C}{c}(1 - e^{-c\delta/2})\int \|y\|_2\,d\sigma_{\delta,N}(y)$$

$$\leq \|v\|_{\mathrm{Lip}}\frac{C}{c}(1 - e^{-c\delta/2})(m\delta)^{1/2},$$

where the last inequality follows by the same reasoning as Lemma 4.3(3). Therefore,

$$\|S_\delta P_\delta v - P_\delta S_\delta v\|_{L^\infty} \le \frac{C}{c} m^{1/2} \delta^{1/2} (1 - e^{-c\delta/2}) \|v\|_{\mathrm{Lip}}. \tag{4-5}$$

Therefore, we can estimate a single term in the telescoping series identity (4-4) by

$$
\begin{aligned}
\|(P_\delta S_\delta)^{2j} P_\delta (S_\delta P_\delta - P_\delta S_\delta) S_\delta (P_{2\delta} S_{2\delta})^{n-1-j} u\|_{L^\infty} &\le \|(S_\delta P_\delta - P_\delta S_\delta) S_\delta (P_{2\delta} S_{2\delta})^{n-1-j} u\|_{L^\infty} \\
&\le \frac{C}{c} m^{1/2} \delta^{1/2} (1 - e^{-c\delta/2}) \|S_\delta (P_{2\delta} S_{2\delta})^{n-1-j} u\|_{\mathrm{Lip}} \\
&\le \frac{C}{c} m^{1/2} \delta^{1/2} (1 - e^{-c\delta/2}) e^{-c\delta/2} e^{-c\delta(n-j-1)/2} \|u\|_{\mathrm{Lip}}.
\end{aligned}
$$

Here we have first applied the fact that P_δ and S_δ are contractions with respect to the L^∞ norm from Lemmas 4.3(1) and 4.4(6); second, we used the estimate (4-5) for $S_\delta P_\delta - P_\delta S_\delta$; and third we used the estimates $\|P_\delta u\|_{\mathrm{Lip}} \le \|u\|_{\mathrm{Lip}}$ and $\|S_\delta u\|_{\mathrm{Lip}} \le e^{-c\delta/2} \|u\|_{\mathrm{Lip}}$ found in Lemmas 4.3(2) and 4.4(5). Now summing up the telescoping series, we get

$$
\begin{aligned}
\|T_{t,\ell+1} u - T_{t,\ell} u\|_{L^\infty} &\le \sum_{j=0}^{n-1} \frac{C}{c} m^{1/2} \delta^{1/2} (1 - e^{-c\delta/2}) e^{-c\delta/2} e^{-c\delta(n-j-1)/2} \|u\|_{\mathrm{Lip}} \\
&\le \frac{C}{c} m^{1/2} \delta^{1/2} (1 - e^{-c\delta/2}) e^{-c\delta/2} \frac{1}{1 - e^{-c\delta/2}} \|u\|_{\mathrm{Lip}} \\
&= \frac{C}{c} m^{1/2} \delta^{1/2} e^{-c\delta/2} \|u\|_{\mathrm{Lip}} \le \frac{C}{2c} m^{1/2} \delta^{1/2} \|u\|_{\mathrm{Lip}}.
\end{aligned}
$$

In other words, we have

$$\|T_{t,\ell+1} u - T_{t,\ell} u\|_{L^\infty} \le \frac{C m^{1/2}}{2c} 2^{-(\ell+1)/2} \|u\|_{\mathrm{Lip}}.$$

It follows that the sequence is Cauchy with respect to $\|\cdot\|_{L^\infty}$ and we have the desired estimate on $\|T_{t,\ell} u - T_t u\|_{L^\infty}$ from summing the geometric series.

The estimate $\|T_{t,\ell} u\|_{\mathrm{Lip}} \le e^{-ct/2} \|u\|_{\mathrm{Lip}}$ follows from Lemmas 4.3(2) and 4.4(5), and then by taking the limit as $\ell \to +\infty$, we obtain $\|T_t u\|_{\mathrm{Lip}} \le e^{-ct/2} \|u\|_{\mathrm{Lip}}$. □

Lemma 4.6. *The semigroup T_t defined above extends to a semigroup defined for positive t such that for $s \le t$*

$$|T_t u(x) - T_s u(x)| \le e^{-cs/2} \Big(\frac{C}{c} (3\sqrt{2} + 5)(t - s)^{1/2} + \|DV(x)\|_2 (t - s) \Big) \|u\|_{\mathrm{Lip}},$$

and $\|T_t u\|_{\mathrm{Lip}} \le e^{-ct/2} \|u\|_{\mathrm{Lip}}$.

Proof. We first prove the estimate on $|T_t u - T_s u|$ for dyadic values of s and t. First, consider the case where $t = 2^{-\ell}$ and $s = 0$. Note that

$$(T_t - 1)u = (T_t - P_t S_t)u + (P_t - 1)S_t u + (S_t - 1)u.$$

The first term can be estimated by Lemma 4.5 with $\ell = 1$, the second term can be estimated by Lemmas 4.3(3) and 4.4(5) as

$$\|(P_t - 1)S_t u\|_{L^\infty} \le m^{1/2} t^{1/2} \|S_t u\|_{\mathrm{Lip}} \le m^{1/2} t^{1/2} \|u\|_{\mathrm{Lip}},$$

and the third term can be estimated by [Lemma 4.4](3). Altogether, we obtain

$$|T_t u(x) - u(x)| \leq \left(\frac{Cm^{1/2}}{c(2 - 2^{1/2})} t^{1/2} + m^{1/2} t^{1/2} + \frac{t}{2} \|DV(x)\|_2 \right) \|u\|_{\text{Lip}}.$$

In the case of general dyadic s and t, suppose $t > s$ and write $t - s$ in a binary expansion to obtain

$$t = s + \sum_{j=n+1}^{\infty} a_j 2^{-j},$$

where $a_j \in \{0, 1\}$ and $a_{n+1} = 1$. Note that $2^{-n-1} \leq |s - t| \leq 2^{-n}$. Let $t_k = s + \sum_{j=n+1}^{k} a_j 2^{-j}$. Then

$$
\begin{aligned}
|T_t u(x) - T_s u(x)| &\leq \sum_{j=n+1}^{\infty} |T_{t_j} u(x) - T_{t_{j-1}} u(x)| \\
&\leq \sum_{j=n+1}^{\infty} \left(\frac{Cm^{1/2}}{c(2 - 2^{1/2})} 2^{-j/2} + m^{1/2} 2^{-j/2} + \frac{2^{-j}}{2} \|DV(x)\|_2 \right) \|T_{t_{j-1}} u\|_{\text{Lip}} \\
&\leq \left(\left(\frac{Cm^{1/2}}{c(2 - 2^{1/2})} + 1 \right) \frac{1}{1 - 2^{-1/2}} \cdot 2^{-(n+1)/2} + \|DV(x)\|_2 \cdot 2^{-(n+1)} \right) \|T_s u\|_{\text{Lip}} \\
&\leq \left(\left(\frac{Cm^{1/2}}{c(2 - 2^{1/2})} + 1 \right) \frac{1}{1 - 2^{-1/2}} (t - s)^{1/2} + \|DV(x)\|_2 (t - s) \right) e^{-cs/2} \|u\|_{\text{Lip}} \\
&\leq e^{-cs/2} \left(\frac{Cm^{1/2}}{c} (3\sqrt{2} + 5)(t - s)^{1/2} + \|DV(x)\|_2 (t - s) \right) \|u\|_{\text{Lip}},
\end{aligned}
$$

where we used the crude estimate that $1 \leq Cm^{1/2}/c$ to combine the first two terms. Because this continuity estimate holds for dyadic values of s and t, we can extend the definition of $T_t u$ to all positive t. Furthermore, because $\|T_t u\|_{\text{Lip}} \leq e^{-ct/2} \|u\|_{\text{Lip}}$ for dyadic t, the same must hold for real values of t.

Now let us verify that $T_s T_t = T_{s+t}$ for all real t. Choose dyadic $s_n \searrow s$ and $t_n \searrow t$ and let u be a Lipschitz function. We know that $T_{s_n} T_{t_n} u = T_{s_n + t_n} u$ and that $T_{s_n + t_n} u \to T_{s+t} u$ locally uniformly, so it suffices to show that $T_{s_n} T_{t_n} u \to T_s T_t u$. Observe that

$$|T_{s_n} T_{t_n} u - T_s T_t u| \leq |(T_{s_n} - T_s) T_{t_n} u| + |T_s (T_{t_n} - T_t) u|.$$

The first term can be estimated by

$$|(T_{s_n} - T_s) T_{t_n} u(x)| \leq e^{-s/2} \left(\frac{C}{c} (3\sqrt{2} + 5)(s_n - s)^{1/2} + \|DV(x)\|_2 (s_n - s) \right) \|T_{t_n} u\|_{\text{Lip}},$$

which goes to zero as $n \to \infty$. For the second term, we first note that

$$|(T_{t_n} - T_t) u(x)| \leq e^{-t/2} \left(\frac{C}{c} (3\sqrt{2} + 5)(t_n - t)^{1/2} + \|DV(x)\|_2 (t_n - t) \right) \|u\|_{\text{Lip}}.$$

Let $h_n(x)$ be the right-hand side. Note that $u \leq v$ implies that $T_s u \leq T_s v$ because this holds for the operators P_s and S_s (since P_s is given by convolution and S_s is given by composition). Therefore,

$$|T_s (T_{t_n} - T_t) u(x)| \leq T_s |(T_{t_n} - T_t) u|(x) \leq T_s h_n(x).$$

Because DV is C-Lipschitz, we know that h_n is an $e^{-t/2}(t_n - t)C\|u\|_{\mathrm{Lip}}$-Lipschitz function and hence

$$|T_s h_n(x)| \le h_n(x) + |(T_s - 1)h_n(x)|$$

$$\le h_n(x) + e^{-t/2}(t_n - t)C\|u\|_{\mathrm{Lip}}\Big(\frac{C}{c}(3\sqrt{2} + 5)s^{1/2} + \|DV(x)\|_2 s\Big),$$

which goes to zero as $n \to \infty$. □

Lemma 4.7. *Let $u(x)$ be Lipschitz. Then $T_t u$ is a weak solution of the equation*

$$\partial_t T_t u = \frac{1}{2N}\Delta(T_t u) - \frac{N}{2}\nabla V \cdot \nabla(T_t u)$$

in the sense that for $\phi \in C_c^\infty(M_N(\mathbb{C})_{\mathrm{sa}}^m)$, we have

$$\int_{M_N(\mathbb{C})_{\mathrm{sa}}^m}[(T_{t_1}u)\phi - (T_{t_0}u)\phi] = \int_{t_0}^{t_1}\int_{M_N(\mathbb{C})_{\mathrm{sa}}^m}\Big[-\frac{1}{2N}\nabla(T_s u)\cdot\nabla\phi - \frac{N}{2}(\nabla V \cdot \nabla(T_s u))\phi\Big]\,ds.$$

Proof. Recall that by Rademacher's theorem if u is Lipschitz, then ∇u exists almost everywhere and it is in $L^\infty(M_N(\mathbb{C})_{\mathrm{sa}}^m)$. In particular, because ∇V is Lipschitz, we also know that the second derivatives of V exist almost everywhere and are in $L^\infty(M_N(\mathbb{C})_{\mathrm{sa}}^m)$.

We begin by considering $\int(S_\delta P_\delta - 1)u \cdot \phi$ for a Lipschitz $u : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$ and a $\phi \in C_c^\infty(M_N(\mathbb{C})_{\mathrm{sa}}^m)$ and $\delta > 0$. Note that

$$(S_\delta P_\delta - 1)u = (S_\delta - 1)P_\delta u + (P_\delta - 1)u.$$

Now $P_\delta u$ is the convolution of u with the Gaussian and so $\nabla(P_\delta u) = P_\delta(\nabla u)$. Because the gradient of the Gaussian is $O(\delta^{-1/2})$, we see that the first derivatives of $P_\delta(\nabla u)$ are $O(\delta^{-1/2})$ in L^∞ (here our estimates may depend on N):

$$P_\delta u(y) - P_\delta u(x) = \nabla P_\delta u(x) \cdot (x - y) + O(\delta^{-1/2}\|x - y\|_2^2).$$

Now using (4-3) and Lemma 4.4(3), we have $W(x, \delta) - x = (N\delta/2)\nabla V(x) + O(\delta^2)$ uniformly on any compact set K. Therefore,

$$(S_\delta - 1)P_\delta u(x) = P_\delta u(W(x, \delta)) - P_\delta u(x) = -\frac{N\delta}{2}\nabla(P_\delta u)(x)\cdot\nabla V(x) + O(\delta^{3/2}).$$

Now we have

$$\int(S_\delta P_\delta - 1)u \cdot \phi = \int(S_\delta - 1)P_\delta u\,\phi + \int(P_\delta - 1)u\,\phi$$

$$= -\frac{N\delta}{2}\int[\nabla(P_\delta u)\cdot\nabla V]\phi + \int u\,(P_\delta - 1)\phi + O(\delta^{3/2})$$

$$= -\frac{N\delta}{2}\int P_\delta u[(\Delta V)\phi + \nabla V \cdot \nabla\phi] + \int u\frac{\delta}{2N}\Delta\phi + O(\delta^{3/2})$$

$$= -\frac{N\delta}{2}\int u\,P_\delta[(\Delta V)\phi + \nabla V \cdot \nabla\phi] + \frac{\delta}{2N}\int u\,\Delta\phi + O(\delta^{3/2}),$$

where the error estimates depend only on C, N, $\|u\|_{\mathrm{Lip}}$, the support of ϕ, and the L^∞ norms of its derivatives. We also know from (4-5) that $(S_\delta P_\delta - P_\delta S_\delta)u$ is bounded by $\|u\|_{\mathrm{Lip}}(Cm^{1/2}/c)(1 - e^{-c\delta})\delta^{1/2}$,

which is $O(\delta^{3/2})$. Therefore,

$$\int (P_\delta S_\delta - 1)u \cdot \phi = -\frac{N\delta}{2} \int u \, P_\delta[\Delta V \phi + \nabla V \cdot \nabla \phi] + \frac{\delta}{2N} \int u \, \Delta \phi + O(\delta^{3/2}).$$

Now suppose that t is a dyadic rational and write $t = n\delta$, where $\delta = 2^{-\ell}$ for some integer ℓ. Recall that $T_{t,\ell} = (P_\delta S_\delta)^n$. Then by a telescoping series argument

$$\int (T_{t,\ell} - 1)u \cdot \phi = \sum_{j=0}^{n-1} \left(-\frac{N\delta}{2} \int T_{j\delta,\ell}u \, P_\delta[(\Delta V)\phi + \nabla V \cdot \nabla \phi] + \frac{\delta}{2N} \int T_{j\delta,\ell}u \, \Delta \phi \right) + O(\delta^{1/2}).$$

We fix a dyadic t and take $\ell \to \infty$ (and hence $\delta \to 0$). The above sum over j may be viewed as a Riemann sum for an integral from 0 to t, where δ is the mesh size. Using Lemma 4.6, we know that $T_t u$ is Hölder continuous in t. Also, by Lebesgue differentiation theory,

$$P_\delta[(\Delta V)\phi + \nabla V \cdot \nabla \phi] \to (\Delta V)\phi + \nabla V \cdot \nabla \phi$$

in $L^1_{\text{loc}}(M_N(\mathbb{C})^m_{\text{sa}})$. There is no difficulty in taking the limit as $\delta \to 0$ inside the integral because ϕ has compact support and all the functions we are integrating are bounded on compact sets. Thus, we obtain

$$\int (T_t - 1)u \cdot \phi \, dx = \int_0^t \int \left(-\frac{N}{2}T_s u[(\Delta V)\phi + \nabla V \cdot \nabla \phi] + \frac{1}{2N}T_s u \, \Delta \phi \right) dx \, ds.$$

We may extend this equality from dyadic t to all positive t using Lemma 4.6. Finally, after another integration by parts (which is justified by approximation by smooth functions in the appropriate Sobolev spaces), we have

$$\int (T_t - 1)u \cdot \phi \, dx = \int_0^t \int \left(-\frac{N}{2}[\nabla (T_s u) \cdot \nabla V]\phi - \frac{1}{2N}\nabla (T_s u) \cdot \nabla \phi \right) dx \, ds.$$

The asserted formula then follows by replacing u with $T_{t_0}u$ and t with $t_1 - t_0$. □

Lemma 4.8. *If μ is the measure given by the potential V and if u is Lipschitz, then we have*

$$\int T_t u \, d\mu = \int u \, d\mu.$$

Proof. By applying Lemma 4.7 and approximating $(1/Z)\exp(-N^2 V(x))$ by compactly supported smooth functions, we see that

$$\int T_t u \, d\mu - \int u \, d\mu = \frac{1}{Z} \iint_0^t \left[-\frac{1}{2N}\nabla (T_s u) \cdot \nabla [e^{-N^2 V}] - \frac{N}{2}(\nabla V \cdot \nabla (T_s u))e^{-N^2 V} \right] ds \, dx = 0. \quad \square$$

Lemma 4.9. *We have $T_t u(x) \to \int u \, d\mu$ as $t \to \infty$ and more precisely*

$$\left| T_t u(x) - \int u \, d\mu \right| \leq e^{-ct/2}\left(\frac{4Cm^{1/2}}{c^2}(6 + 5\sqrt{2})t^{-1/2} + \frac{2}{c}\|V(x)\|_2 \right)\|u\|_{\text{Lip}}.$$

Proof. Fix t and fix $r \geq t$. Let n be an integer. Then using Lemma 4.6,

$$|T_{t+r}u(x) - T_t u(x)| \leq \sum_{j=0}^{n-1} |T_{t+r(j+1)/n}u(x) - T_{t+rj/n}u(x)|$$

$$\leq \sum_{j=0}^{n-1} e^{-ct/2} e^{-crj/2n} \left(\frac{Cm^{1/2}}{c}(3\sqrt{2}+5)\left(\frac{r}{n}\right)^{1/2} + \|V(x)\|_2\left(\frac{r}{n}\right) \right) \|u\|_{\mathrm{Lip}}$$

$$\leq e^{-ct/2} \frac{1}{1-e^{-cr/2n}} \left(\frac{Cm^{1/2}}{c}(3\sqrt{2}+5)\left(\frac{r}{n}\right)^{1/2} + \|V(x)\|_2\left(\frac{r}{n}\right) \right) \|u\|_{\mathrm{Lip}}$$

$$\leq e^{-ct/2} \frac{2n}{cr} \left(\frac{Cm^{1/2}}{c}(3\sqrt{2}+5)\left(\frac{r}{n}\right)^{1/2} + \|V(x)\|_2\left(\frac{r}{n}\right) \right) \|u\|_{\mathrm{Lip}}.$$

Since $r \geq t$, we can choose n such that $t/4 \leq r/n \leq t/2$. Then we have

$$|T_{t+r}u(x) - T_t u(x)| \leq e^{-ct/2}\left(\frac{4Cm^{1/2}}{c^2}(6+5\sqrt{2})t^{-1/2} + \frac{2}{c}\|V(x)\|_2 \right) \|u\|_{\mathrm{Lip}}.$$

Because this holds for all sufficiently large r, this shows that $\lim_{t\to\infty} T_t u(x)$ exists. Because $\|T_t u\|_{\mathrm{Lip}} \leq e^{-ct/2}\|u\|_{\mathrm{Lip}}$, the limit must be constant and therefore equals $\int u \, d\mu$. Moreover, we have the asserted rate of convergence by taking $r \to \infty$ in the above estimate. $\qquad\square$

4B. *Approximability and convergence of moments.* Now we are ready to show that the sequence of semigroups $T_t^{V_N}$ associated to a sequence of potentials V_N will preserve asymptotic approximability by trace polynomials and as a consequence we will show that the moments of the associated measures μ_N converge.

Lemma 4.10. *Let $V_N : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$ be a sequence of potentials such that $V_N(x) - (c/2)\|x\|_2^2$ is convex and $V_N(x) - (C/2)\|x\|_2^2$ is concave, where $0 < c \leq C$. For each N, let μ_N be the associated measure. Let $S_t^{V_N}$ and $T_t^{V_N}$ denote the semigroups defined in the previous section. Suppose that the sequence $\{DV_N\}$ is asymptotically approximable by trace polynomials. Suppose that $\{u_N\}$ is a sequence of scalar-valued K-Lipschitz functions which is asymptotically approximable by (scalar-valued) trace polynomials. Then:*

(1) *$\{S_t^{V_N} u_N\}$ is asymptotically approximable by trace polynomials for each $t \geq 0$.*

(2) *$\{T_t^{V_N} u_N\}$ is asymptotically approximable by trace polynomials for each $t \geq 0$.*

(3) *$\lim_{N\to\infty} \int u_N \, d\mu_N$ exists.*

Proof. (1) Recall that $S_t^{V_N} u_N = u_N(W_N(x, t))$, where W_N is the solution to (4-3). Thus, by Lemma 3.27, it suffices to show that $W_N(x, t)$ is asymptotically approximable by trace polynomials for each t. To this end, we write $W_N(x, t)$ as the limit as $\ell \to \infty$ of Picard iterates $W_{N,\ell}$ given by

$$W_{N,0}(x, t) = x, \quad W_{N,\ell+1}(x, t) = x - \frac{1}{2}\int_0^t DV_k(W_N(x, s)) \, ds.$$

Because DV_N is C-Lipschitz, the standard Picard–Lindelöf arguments show that

$$\|W_{N,\ell}(x, t) - W_N(x, t)\|_2 \leq \sum_{n=\ell+1}^{\infty} \frac{C^{n-1}t^n}{2^n n!} \|DV_N(x)\|_2.$$

Because DV_N is asymptotically approximable by trace polynomials, we know that $\|DV_N(x)\|_2$ is uniformly bounded on $\|x\| \leq R$ for any given $R > 0$, and therefore, for each T and $R > 0$, the convergence of $W_{N,\ell}$ to W_N as $\ell \to \infty$ is uniform for all $\|x\| \leq R$ and $t \leq T$ and $N \in \mathbb{N}$. Thus, by Observation 3.26, it suffices to show that each Picard iterate $\{W_{N,\ell}(x,t)\}_N$ is asymptotically approximable by trace polynomials.

Fix $T > 0$. We claim that, for every ℓ, for every $R > 0$ and $\epsilon > 0$, there exists a trace polynomial $f(X,t)$ with coefficients that are polynomial functions of t such that

$$\limsup_{N \to \infty} \sup_{t \in [0,T]} \sup_{\|x\|_\infty \leq R} \|W_{N,\ell}(x,t) - f(x,t)\|_2 \leq \epsilon.$$

We proceed by induction on ℓ, with the base case $\ell = 0$ being trivial. For the inductive step, fix ϵ and R, and choose a trace polynomial $f(X,t)$ which provides an $(\epsilon/(CT), R)$ approximation for $W_{N,\ell}$ for all $t \leq T$. Let

$$R' = \sup_{t \in [0,T]} \sup_N \sup_{x \in M_N(\mathbb{C})_{\mathrm{sa}}^m : \|x\|_\infty \leq R} \|f(x,t)\| < +\infty.$$

Choose another trace polynomial $g(X)$ which is an $(\epsilon/T, R')$ approximation for $\{DV_N\}$, and let $h(X,t) = X - \frac{1}{2} \int_0^t g(f(X,s))\,ds$. Then arguing as in Lemma 3.27, we have for $\|x\| \leq R$ and $t \in [0,T]$ that

$$\|W_{N,\ell+1}(x,t) - h(x,t)\| \leq \frac{1}{2} \int_0^t \|DV_N(W_{N,\ell}(x,s)) - g(f(x,s))\|_2\,ds$$

$$\leq \frac{t}{2} \sup_{\|y\| \leq R'} \|DV_N(y) - g(y)\|_2 + \frac{Ct}{2} \sup_{s \in [0,T]} \sup_{\|x\| \leq R} \|W_{N,\ell}(x,s) - f(x,s)\|_2.$$

Taking $N \to \infty$, we see that $h(x,t)$ is an (ϵ, R) approximation for $\{W_{N,\ell}(x,t)\}_N$ for all $t \leq T$.

(2) We have shown that $S_t^{V_k}$ preserves asymptotic approximability. Moreover, if the sequence $u_N : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{C}$ is asymptotically approximable by trace polynomials and u_N is K-Lipschitz, then the sequence $P_t u_N$ is also asymptotically approximable by trace polynomials by Lemma 3.28 (the hypothesis (3-18) is satisfied since $|u_N(x)| \leq |u_N(0)| + K\|x\|_2$ and $|u_N(0)|$ is bounded as $N \to +\infty$ because u_N is asymptotically approximable by trace polynomials). Therefore, the iterated operator $T_{t,\ell}^{V_N} = (P_{2^{-\ell}} S_{2^{-\ell}}^{V_N})^{2^\ell t}$ preserves asymptotic approximability for dyadic values of t. Taking $\ell \to \infty$, we see by Observation 3.26 and Lemma 4.5 that $T_t^{V_N}$ preserves asymptotic approximability for dyadic values of t. Finally, we extend the approximability property to $T_t^{V_N}$ for all real t using Observation 3.26 and Lemma 4.6.

(3) We know by Lemma 4.9 that $T_t^{V_N} u_N(x) \to \int u_N \, d\mu_N$ as $t \to \infty$ with estimates that are independent of N. It follows by Observation 3.26 that the sequence of constant functions $\left\{ \int u_N \, d\mu_N \right\}$ is asymptotically approximable by trace polynomials. But since these functions are constant, this simply means that the limit of $\int u_N \, d\mu_N$ as $N \to \infty$ exists. $\qquad \square$

Proof of Theorem 4.1. (1) Let $a_N = \int x \, d\mu_N(x)$ and $a_{N,j} = \int x_j \, d\mu_N(x)$. Note that

$$R_N \leq \max_j \int \|x_j - a_{N,j}\| \, d\mu_N(x) + \max_j \left| \int \tau_N(x_j) \, d\mu_N(x) \right| + \max_j \left\| \int (x_j - \tau_N(x_j)) \, d\mu(x) \right\|.$$

When we take the $\limsup$ as $N \to \infty$, the first term is bounded by $2/c^{1/2}$ by Corollary 2.12, while the last term is bounded by M. It remains to estimate $\int \tau_N(x_j) \, d\mu_N(x)$.

Using integration by parts, we see that

$$\int DV_N(x)\,d\mu_N(x) = -\frac{1}{Z}\int \frac{1}{N}\nabla[e^{-V_N}(x)]\,dx = 0.$$

On the other hand, we may estimate $\|DV_N(x) - (DV_N(0) + ((C+c)/2)x)\|_2$ as follows. We assumed that $V_N(x) - (c/2)\|x\|_2^2$ is convex and $V_N(x) - (C/2)\|x\|_2^2$ is concave. Let $\widetilde{V}_N(x) = V_N(x) - ((C+c)/2)\|x\|_2^2$. Then $\widetilde{V}_N(x) + ((C-c)/2)\|x\|_2^2$ is convex and $\widetilde{V}_N(x) - ((C-c)/2)\|x\|_2^2$ is concave. Therefore, $D\widetilde{V}_N$ is $(C-c)/2$-Lipschitz with respect to $\|\cdot\|_2$ by [Proposition 2.13](4). It follows that

$$\left\| DV_N(x) - \left(DV_N(0) + \frac{C+c}{2}x\right)\right\|_2 = \|D\widetilde{V}_N(x) - D\widetilde{V}_N(0)\|_2 \le \frac{C-c}{2}\|x\|_2.$$

Therefore,

$$\begin{aligned}
\left\| DV_N(0) + \frac{C+c}{2}a_N \right\|_2 &= \left\| -DV_N(x) + \left(DV_N(0) + \frac{C+c}{2}x\right)\right\|_2 \\
&\le \frac{C-c}{2}\int \|x\|_2\,d\mu_N(x) \\
&\le \frac{C-c}{2}\left(\|a_N\|_2 + \left(\int \|x - a_N\|_2^2\,d\mu(x)\right)^{1/2}\right) \\
&\le \frac{C-c}{2}(\|a_N\|_2 + c^{-1/2}),
\end{aligned}$$

where the last step follows from [Theorem 2.11]. Altogether,

$$\frac{C+c}{2}\|a_N\|_2 \le \frac{C-c}{2}\|a_N\|_2 + \|DV_N(0)\|_2 + \frac{C-c}{2c^{1/2}}.$$

Then we move $((C-c)/2)\|a_N\|_2$ to the left-hand side and divide the equation by c to obtain

$$\left| \int \tau_N(x_j)\,d\mu_N(x)\right| \le \|a_N\|_2 \le \frac{1}{c}\|DV_N(0)\|_2 + \frac{C-c}{2c^{3/2}},$$

which proves the asserted estimate on R_N. The tail estimate on $\mu_N(\|x_j\| \ge R_N + \delta)$ follows from [Corollary 2.12].

(2) Fix a noncommutative polynomial p. Let $R_* = \limsup_{N\in\mathbb{N}} R_N$, which we know is finite because of (1) and suppose that $R' > R_*$. Let $\psi \in C_c^\infty(\mathbb{R})$ be such that $\psi(t) = t$ for $|t| \le R'$, and define $\Psi(x_1,\dots,x_m) = (\psi(x_1),\dots,\psi(x_m))$, where $\psi(x_j)$ is defined through the continuous functional calculus for self-adjoint operators. Now $x \mapsto \psi(x)$ is Lipschitz in $\|\cdot\|_2$ for $x \in M_N(\mathbb{C})_{\mathrm{sa}}$ with constants independent of N (see for instance [Proposition 8.8] below). It follows that $p(\Psi(x))$ is globally Lipschitz in $\|\cdot\|_2$ and it equals $p(x)$ when $\|x\| \le R'$.

Furthermore, we claim that the sequence $\tau_N(p(\Psi(x)))$ is asymptotically approximable by trace polynomials. To see this, choose some radius r and $\delta > 0$. By the Weierstrass approximation theorem, there exists a polynomial $\hat\psi(t)$ such that $|\psi(t) - \hat\psi(t)| \le \delta$ for $t \in [-r, r]$. By the spectral mapping theorem, we have $\|\psi(y) - \hat\psi(y)\| \le \delta$ if $y \in M_N(\mathbb{C})_{\mathrm{sa}}$ and $\|y\| \le r$. In particular, if we let $\widehat{\Psi}(x) = (\hat\psi(x_1),\dots,\hat\psi(x_m))$ for $x \in M_N(\mathbb{C})_{\mathrm{sa}}^m$, then we have $\|\Psi(x) - \widehat{\Psi}(x)\| \le \delta$ when $\|x\|_\infty \le r$. Given $\epsilon > 0$, we may choose δ small

enough to guarantee that $|\tau_N(p(\Psi(x))) - \tau_N(p(\widehat{\Psi}(x)))| \leq \epsilon$ for $\|x\|_\infty \leq r$, and clearly $\tau_N(p(\widehat{\Psi}(x)))$ is a trace polynomial. Thus, $\tau_N(p(\Psi(x)))$ is asymptotically approximable by trace polynomials.

Therefore, by Lemma 4.10, the limit

$$\lambda(p) = \lim_{N\to\infty} \int \tau_N(p(\Psi(x))) \, d\mu_N(x)$$

exists. Clearly, λ satisfies all the conditions to be a noncommutative law (Definition 2.3). Furthermore, because of the operator norm bounds (1), we know that $\int_{\|x\|\geq R'} \tau_N(p(x)) \, d\mu_N(x)$ is finite and approaches zero as $N \to \infty$ and the same holds for the integral of $\tau_N(p(\Psi(x)))$. Therefore,

$$\lim_{N\to\infty} \int \tau_N(p(x)) \, d\mu_N(x) = \lim_{N\to\infty} \int \tau_N(p(\Psi(x))) \, d\mu_N(x) = \lambda(p).$$

Also, we have $\lambda(p) = \lim_{N\to\infty} \int_{\|x\|\leq R'} \tau_N(p(x)) \, d\mu_N(x)$ and hence $\lambda \in \Sigma_{m,R'}$. But since this holds for every $R' > R_*$, we have $\lambda \in \Sigma_{m,R_*}$.

(3) It suffices to prove the concentration claim (3) for sufficiently large R, say $R > 2R_*$. Because the topology of $\Sigma_{m,R}$ is generated by the functions $\lambda \mapsto \lambda(p)$ for noncommutative polynomials p, it suffices to consider the case where $\mathcal{U} = \{\lambda' : |\lambda'(p) - \lambda(p)| < \epsilon\}$ for some noncommutative polynomial p. Choose a function $\psi \in C_c^\infty(\mathbb{R})$ with $\psi(t) = t$ for $|t| \leq R$, and let Ψ be as above. Then by Theorem 2.10,

$$\mu_N\left(\left|\tau_N(p(\Psi(x))) - \int \tau_N(p \circ \Psi) \, d\mu_N\right| \geq \frac{\epsilon}{2}\right) \leq 2e^{-N^2\epsilon^2/8\|\tau_N(p\circ\Psi)\|_{\mathrm{Lip}}^2}.$$

But by the same reasoning as in part (2), we know that large enough N, we have

$$\left|\int \tau_N(p \circ \Psi) \, d\mu_N - \lambda(p)\right| \leq \frac{\epsilon}{2},$$

and hence

$$\limsup_{N\to\infty} \frac{1}{N^2} \log \mu_N(\|x\| \leq R, \ |\tau_N(p(x)) - \lambda(p)| \geq \epsilon) < 0. \qquad \square$$

5. Entropy and Fisher's information

5A. *Classical entropy.* In this section, we will state 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. Recall that the (classical, continuous) entropy of a measure $d\mu(x) = \rho(x) \, dx$ on $\mathbb{R}^n$ 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 5.1. *Assume that μ is a probability measure on $\mathbb{R}^n$ with density ρ and that $\int |x|^2 \, d\mu(x) < +\infty$.*

(1) *The positive part of $-\rho \log \rho$ has finite integral and hence $\int -\rho \log \rho$ is well-defined in $[-\infty, +)$.*

(2) *We have $h(\mu) \leq (n/2) \log 2\pi a e$, where $a = \int |x|^2 \, d\mu(x)/n$, and equality is achieved in the case of a centered Gaussian with covariance matrix aI.*

(3) *Suppose $\{\mu_k\}$ 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 $\limsup_{k\to\infty} h(\mu_k) \le h(\mu)$.*

(4) *If ν is a probability measure with finite second moments, then $h(\mu * \nu) \ge h(\mu)$.*

Proof. (1) Let $a = \int |x|^2 \, d\mu(x)/n$. Let $g(x) = (2\pi a)^{-n/2} e^{-|x|^2/(2a)}$ be the Gaussian of variance a, and let γ be the corresponding Gaussian measure. Let $\tilde{\rho} = \rho/f$ be the density of μ relative to the Gaussian. We write

$$-\rho(x) \log \rho(x) = -\tilde{\rho}(x) \log \tilde{\rho}(x) \cdot g(x) - \tilde{\rho}(x) \log g(x) \cdot g(x)$$

$$= -\tilde{\rho}(x) \log \tilde{\rho}(x) \cdot g(x) + \left(\frac{|x|^2}{2a} + \frac{n}{2} \log 2\pi a \right) \rho(x).$$

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.

(2) The function $-t \log t$ is concave and its tangent line at $t = 0$ is $1 - t$, and hence $-t \log t \le 1 - t$. Thus,

$$\int -\tilde{\rho} \log \tilde{\rho} \, d\gamma \le \int (1 - \tilde{\rho}) \, d\gamma = 0,$$

so

$$h(\mu) \le \int \left(\frac{|x|^2}{2a} + \frac{n}{2} \log 2\pi a \right) \rho(x) \, dx = \frac{n}{2} + \frac{n}{2} \log 2\pi a = \frac{n}{2} \log 2\pi e.$$

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 of covariance matrix I and g its density. Let $\tilde{\rho}_k = \rho_k/g$. As before,

$$h(\mu_k) = \int -\tilde{\rho}_k \log \tilde{\rho}_k \, d\gamma + \int \left(\frac{|x|^2}{2} + \frac{n}{2} \log 2\pi \right) d\mu_k.$$

By assumption, the second term converges to $\int (|x|^2/2 + (n/2) \log 2\pi) \, d\mu$ 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

$$-\tilde{\rho}(x) \log \tilde{\rho}(x) \ge \int -\rho(x - y) \log \rho(x - y) \, d\nu(y).$$

The right-hand side is

$$\iint -\rho(x - y) \log \rho(x - y) \, d\nu(y) \, dx = \iint -\rho(x - y) \log \rho(x - y) \, dx \, d\nu(y) = h(\mu),$$

where the exchange of order is justified because we know that $-\rho \log \rho$ is integrable since $h(\mu) > -\infty$. Therefore, $h(\mu * \nu) = \int -\tilde{\rho} \log \tilde{\rho} \ge h(\mu)$. $\square$

5B. *Microstates free entropy.* Because there is no integral formula known for free entropy of multiple noncommutative variables as in the classical case, Voiculescu [1993; 1994] defined the free analogue of entropy using Boltzmann's microstates viewpoint on entropy.

Definition 5.2. For $\mathcal{U} \subseteq \Sigma_m$, we define the *microstate space*

$$\Gamma_N(\mathcal{U}) = \{x \in M_N(\mathbb{C})_{\mathrm{sa}}^m : \lambda_x \in \mathcal{U}\},$$

$$\Gamma_{N,R}(\mathcal{U}) = \{x \in M_N(\mathbb{C})_{\mathrm{sa}}^m : \lambda_x \in \mathcal{U}, \|x\|_\infty \leq R\}.$$

The *microstates free entropy* of a noncommutative law λ is defined as

$$\chi_R(\lambda) = \inf_{\mathcal{U} \ni \lambda} \limsup_{N \to \infty} \left(\frac{1}{N^2} \log \mathrm{vol}\, \Gamma_{N,R}(\mathcal{U}) + \frac{m}{2} \log N \right),$$

$$\chi(\lambda) = \sup_{R > 0} \chi_R(\lambda).$$

Here vol denotes the Lebesgue measure with respect to the identification of $M_N(\mathbb{C})_{\mathrm{sa}}^m$ with $\mathbb{R}^{mN^2}$ as in Section 2A, and $\mathcal{U}$ ranges over all open neighborhoods of λ in Σ_m. Similarly, we denote

$$\underline{\chi}_R(\lambda) = \inf_{\mathcal{U} \ni \lambda} \liminf_{N \to \infty} \left(\frac{1}{N^2} \log \mathrm{vol}\, \Gamma_{N,R}(\mathcal{U}) + \frac{m}{2} \log N \right),$$

$$\underline{\chi}(\lambda) = \sup_{R > 0} \underline{\chi}_R(\lambda).$$

Remark 5.3. Note that $\mathcal{U} \subseteq \mathcal{V}$ implies that

$$\limsup_{N \to \infty} \left(\frac{1}{N^2} \log \mathrm{vol}\, \Gamma_{N,R}(\mathcal{U}) + \frac{m}{2} \log N \right) \leq \limsup_{N \to \infty} \left(\frac{1}{N^2} \log \mathrm{vol}\, \Gamma_{N,R}(\mathcal{V}) + \frac{m}{2} \log N \right).$$

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.

Definition 5.4. A sequence of probability measures μ_N on $M_N(\mathbb{C})_{\mathrm{sa}}^m$ is said to *concentrate around the noncommutative law* λ if $\lambda_x \to \lambda$ in probability when x is chosen according to μ_N, that is, for any neighborhood $\mathcal{U}$ of λ in Σ_m, we have

$$\lim_{k \to \infty} \mu_N(x \in \Gamma_N(\mathcal{U})) = 1.$$

Proposition 5.5. *Let* $V_N : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$ *be a potential with* $\int \exp(-N^2 V_N(x))\, dx < +\infty$ *and let* μ_N *be the associated measure. Assume*:

(A) *The sequence* $\{\mu_N\}$ *concentrates around a noncommutative law* λ.

(B) *The sequence* $\{V_N\}$ *is asymptotically approximable by scalar-valued trace polynomials.*

(C) *For some* $n \geq 1$ *and* $a, b > 0$ *we have* $|V_N| \leq a + b \sum_{j=1}^m \tau_N(x_j^{2n})$.

(D) *There exists* $R_0 > 0$ *such that*

$$\lim_{N \to \infty} \int_{\|x\|_\infty \geq R_0} \left(1 + \sum_{j=1}^m \tau_N(x_j^{2n}) \right) d\mu_N(x) = 0,$$

where n is the same number as in (C).

Then λ can be realized as the law of noncommutative random variables $X = (X_1, \ldots, X_m)$ in a von Neumann algebra $(\mathcal{M}, \tau)$ with $\|X_j\| \leq R_0$. Moreover, we have

$$\chi(\lambda) = \chi_{R_0}(\lambda) = \limsup_{N \to \infty}\left(\frac{1}{N^2}h(\mu_N) + \frac{m}{2}\log N\right), \tag{5-1}$$

$$\underline{\chi}(\lambda) = \underline{\chi}_{R_0}(\lambda) = \liminf_{N \to \infty}\left(\frac{1}{N^2}h(\mu_N) + \frac{m}{2}\log N\right). \tag{5-2}$$

Proof. It follows from assumptions (A) and (D) that for every noncommutative 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 $|\lambda(X_{i_1} \ldots X_{i_k})| \leq R_0^k$ for every $i_1, \ldots, i_k$, and hence $\lambda \in \Sigma_{m, R_0}$. The fact that λ can be realized by operators in a von Neumann algebra is standard ([Proposition 2.7]).

Now let us evaluate χ_R and $\underline{\chi}_R$ for $R \geq R_0$. Recall that

$$d\mu_N(x) = \frac{1}{Z_N}\exp(-N^2 V_N(x))\,dx, \quad Z_N = \int \exp(-N^2 V_N(x))\,dx,$$

and note that

$$h(\mu_N) = N^2 \int V_N(x)\,d\mu_N(x) + \log Z_N.$$

The assumptions (C) and (D) imply that

$$\lim_{N \to \infty} \int_{\|x\|_\infty \geq R} |V_N(x)|\,d\mu_N(x) = 0 \quad \text{and} \quad \lim_{N \to \infty} \mu_N(x : \|x\|_\infty \geq R) = 0.$$

Therefore, if we let

$$d\mu_{N,R}(x) = \frac{1}{Z_{N,R}}\mathbf{1}_{\|x\|_\infty \leq R}\exp(-N^2 V_N(x))\,dx, \quad Z_{N,R} = \int_{\|x\|_\infty \leq R}\exp(-N^2 V_N(x))\,dx,$$

then, as $N \to \infty$, we have

$$\int V_N\,d\mu_N - \int V_N\,d\mu_{N,R} \to 0, \quad \log Z_N - \log Z_{N,R} \to 0,$$

and hence

$$\frac{1}{N^2}h(\mu_N) - \frac{1}{N^2}h(\mu_{N,R}) \to 0.$$

Fix $\epsilon > 0$. By assumption (B), there is a 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 noncommutative moments, the set $\mathcal{U} = \{\lambda' : |\lambda'(f) - \lambda(f)| < \epsilon/2\}$ is open. Now suppose that $\mathcal{V} \subseteq \mathcal{U}$ is a neighborhood of λ. Note that

$$\lim_{N \to \infty} \mu_{N,R}(\Gamma_{N,R}(\mathcal{V})) = \lim_{N \to \infty} \frac{Z_N}{Z_{N,R}}\mu_N(\Gamma_N(\mathcal{V}) \cap \{x : \|x\|_\infty \leq R\}) = 1,$$

where we have used that $Z_N/Z_{N,R} \to 1$ 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 (D). Moreover, by our choice of f and $\mathcal{U}$, we have

$$x \in \Gamma_{N,R}(\mathcal{V}) \quad \Longrightarrow \quad |V_N(x) - \lambda(f)| \leq \epsilon.$$

Therefore,

$$Z_{N,R}\mu_{N,R}(\Gamma_{N,R}(\mathcal{V})) = \int_{\Gamma_{N,R}(\mathcal{V})} \exp(-N^2 V_N(x))\, dx$$

$$= \text{vol}\,\Gamma_{N,R}(\mathcal{V}) \exp(-N^2(\lambda(f) + O(\epsilon))).$$

Thus,

$$\log Z_{N,R} + \log \mu_{N,R}(\Gamma_{N,R}(\mathcal{V})) = \log \text{vol}\,\Gamma_{N,R}(\mathcal{V}) - N^2(\lambda(f) + O(\epsilon)).$$

Meanwhile, note that $|f(x)|$ is bounded by some constant K whenever $\|x\|_\infty \leq R$ (where K is independent of N). Therefore,

$$\int V_N\, d\mu_{N,R} = \int_{\Gamma_{N,R}(\mathcal{V})} V_N\, d\mu_{N,R} + \int_{\Gamma_{N,R}(\mathcal{V}^c)} V_N\, d\mu_{N,R}$$

$$= \int_{\Gamma_{N,R}(\mathcal{V})} \lambda[f]\, d\mu_{N,R} + \int_{\Gamma_{N,R}(\mathcal{V}^c)} \lambda_x[f]\, d\mu_{N,R} + O(\epsilon)$$

$$= \lambda(f)\mu_{N,R}(\Gamma_{N,R}(\mathcal{V})) + O(\epsilon) + O\big(K\,\mu_N(\Gamma_{N,R}(\mathcal{V}^c))\big).$$

Altogether,

$$\frac{1}{N^2} h(\mu_{N,R}) = \int V_N\, d\mu_{N,R} + \frac{1}{N^2} \log Z_{N,R}$$

$$= \lambda(f)(\mu_{N,R}(\Gamma_{N,R}(\mathcal{V})) - 1) + \frac{1}{N^2} \log \text{vol}\,\Gamma_{N,R}(\mathcal{V})$$

$$+ O(\epsilon) + O(K\,\mu_N(\Gamma_{N,R}(\mathcal{V}^c))) - \frac{1}{N^2} \log \mu_{N,R}(\Gamma_{N,R}(\mathcal{V})).$$

Now we apply the fact that $\mu_{N,R}(\Gamma_{N,R}(\mathcal{V})) \to 1$ to obtain

$$\limsup_{N \to \infty} \frac{1}{N^2} |h(\mu_{N,R}) - \log \text{vol}\,\Gamma_{N,R}(\mathcal{V})| = O(\epsilon).$$

In light of Remark 5.3, 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

$$\chi_R(\lambda) = \limsup_{N \to \infty} \left(\frac{1}{N^2} h(\mu_{N,R}) + \frac{m}{2} \log N \right) + O(\epsilon)$$

$$= \limsup_{N \to \infty} \left(\frac{1}{N^2} h(\mu_N) + \frac{m}{2} \log N \right) + O(\epsilon).$$

Next, we take $\epsilon \to 0$ and obtain $\chi_R(\lambda) = \limsup_{N \to \infty} (N^{-2} \log h(\mu_N) + (m/2) \log N)$ 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 (5-1) is proved. The proof of (5-2) is identical. □

5C. Classical Fisher information. The classical Fisher information of a probability measure μ on $\mathbb{R}^n$ describes how the entropy changes when μ is convolved with a Gaussian. Suppose μ is given by the smooth density $\rho > 0$ on $\mathbb{R}^n$, and let γ_t be the multivariable Gaussian measure on $\mathbb{R}^n$ with covariance matrix tI. Then the density ρ_t for $\mu_t = \mu * \gamma_t$ evolves according to the heat equation $\partial_t \rho_t = \Delta/2\rho_t$. Integration by parts shows that $\partial_t h(\mu_t) = \frac{1}{2} \int |\nabla \rho_t/\rho_t|^2\, d\mu_t$ (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 $\mathbb{R}^n$-valued random variable $\Xi = -\nabla \rho(X) / \rho(X)$ satisfies the integration-by-parts relation

$$E[\Xi \cdot f(X)] = -\int \frac{\nabla \rho(x)}{\rho(x)} f(x) \rho(x) \, dx = \int \rho(x) \nabla f(x) \, dx = E[\nabla f(X)] \quad \text{for } f \in C_c^\infty(\mathbb{R}^n), \tag{5-3}$$

or equivalently $E[\Xi_j f(X)] = E[\partial_j f(X)]$ for each j.

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. Following the terminology used by Voiculescu in the free case, if X is an $\mathbb{R}^n$-valued random variable on the probability space (Ω, P), we say that an $\mathbb{R}^n$-valued random variable $\Xi \in L^2(\Omega, P)$ is a (*classical*) *conjugate variable for* X if $E[\Xi \cdot f(X)] = E[\nabla f(X)]$ and if each Ξ_j is in the closure of $\{f(X) : f \in C_c^\infty(\mathbb{R}^n)\}$ 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 $f \in C_c^\infty(\mathbb{R}^n)$, it follows that the conjugate variable 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 conjugate variable to X, and $\mathcal{I}(\mu) = +\infty$ if no conjugate variable exists.

The probabilistic viewpoint enables us to produce conjugate variables and estimate Fisher information using conditional expectations. (See [Voiculescu 1998, Proposition 3.7] for the free case.)

Lemma 5.6. *Suppose that X and Y are independent $\mathbb{R}^n$-valued random variables with $X \sim \mu$ and $Y \sim \nu$. If Ξ is a conjugate variable for X, then $E[\Xi | X + Y]$ is a conjugate variable 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 $g \in C_c^\infty(\mathbb{R}^n \times \mathbb{R}^n)$ that $E[\Xi_j g(X, Y)] = E[\partial_{X_j} g(X, Y)]$. In particular, if $f \in C_c^\infty(\mathbb{R}^n)$, 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 $\{f(X + Y) : f \in C_c^\infty(\mathbb{R}^n)\}$ and hence

$$E[E[\Xi_j | X + Y] f(X + Y)] = E[\partial_j f(X + Y)].$$

So $\mathcal{I}(\mu * \nu) = E[|E[\Xi | X + Y]|^2] \leq E[|\Xi|^2] = \mathcal{I}(\mu)$. By symmetry, $\mathcal{I}(\mu * \nu) \leq \mathcal{I}(\nu)$. $\square$

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 nonmicrostates free entropy χ^*. For the reader's convenience, we include a statement and proof in the random matrix setting with free

probabilistic normalizations. See also [Barron 1986, Lemma 1; Voiculescu 1998, Proposition 7.6]. Recall that we identify $M_N(\mathbb{C})^m_{\text{sa}}$ with $\mathbb{R}^{mN^2}$ using the orthonormal basis given in Section 2A rather than entrywise coordinates (since some entries are real and some are complex).

Lemma 5.7. *Let μ be a probability measure on $M_N(\mathbb{C})^m_{\text{sa}}$ with finite variance and with density ρ, and let $\sigma_{t,N}$ be the law of m independent GUEs of normalized variance t. If $a = (1/m) \int \|x\|_2^2 \, d\mu(x) = (1/(mN)) \int |x|^2 \, d\mu(x)$, then we have for $t \geq 0$ that*

$$\frac{m}{a+t} \leq \frac{1}{N^3}\mathcal{I}(\mu * \sigma_{t,N}) \leq \min\left(\frac{m}{t}, \frac{1}{N^3}\mathcal{I}(\mu)\right). \tag{5-4}$$

Moreover,

$$\frac{1}{N^2}h(\mu * \sigma_{t,N}) - \frac{1}{N^2}h(\mu) = \frac{1}{2}\int_0^t \frac{1}{N^3}\mathcal{I}(\mu * \sigma_{s,N}) \, ds \tag{5-5}$$

and

$$\frac{1}{N^2}h(\mu) + \frac{m}{2}\log N = \frac{1}{2}\int_0^\infty \left(\frac{m}{1+s} - \frac{1}{N^3}\mathcal{I}(\mu * \sigma_{s,N})\right) ds + \frac{m}{2}\log 2\pi e. \tag{5-6}$$

Proof. To prove (5-4), suppose $t \geq 0$ and let X and Y be random variables with the laws μ and $\sigma_{t,N}$ respectively. The lower bound is trivial if $\mathcal{I}(\mu * \sigma_{t,N}) = +\infty$, so suppose that $X + Y$ has a conjugate variable Ξ. Then after some computation, the integration-by-parts relation shows that $E\langle \Xi, X + Y\rangle_{\text{Tr}} = mN^2$. Thus,

$$E[|\Xi|^2] \geq \frac{|E\langle \Xi, X+Y\rangle_{\text{Tr}}|^2}{E|X+Y|^2} = \frac{(mN^2)^2}{N(ma+mt)} = \frac{N^3}{a+t}$$

since the variance of Y with respect to the nonnormalized inner product is Nmt and the variance of X is Na. The upper bound is trivial in the case where $t = 0$. If $t > 0$, then by the previous lemma $\mathcal{I}(\mu * \sigma_{t,N}) \leq \min(\mathcal{I}(\mu), \mathcal{I}(\sigma_{t,N}))$. Moreover, a direct computation shows that if $Y \sim \sigma_{t,N}$, then the conjugate variable is $(N/t)Y$ and the Fisher information is mN^3/t.

Next, to prove (5-5), let $\mu_t := \mu * \sigma_{t,N}$. By basic properties of convolving positive functions with the Gaussian, μ_t has a smooth density ρ_t. We claim that if $0 < \delta < t$, then

$$h(\mu_t) - h(\mu_\delta) = \frac{1}{2N}\int_\delta^t \mathcal{I}(\mu_s) \, ds = \frac{1}{2N}\int_\delta^t \int_{M_N(\mathbb{C})^m_{\text{sa}}} \frac{|\nabla \rho_s(x)|^2}{\rho_s(x)} \, dx \, ds. \tag{5-7}$$

This will follow from integration by parts, but to give a complete justification, we first introduce a smooth compactly supported "cutoff" function $\psi_R : M_N(\mathbb{C})^m_{\text{sa}} \to \mathbb{R}$ 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 rescaling by R of some fixed function, we can arrange that $\|\nabla\psi_R(x)\|_2 \leq C/R$ for some constant C. Because $\partial_s \rho_s = (2N)^{-1}\Delta\rho_s$, we have

$$\frac{d}{dt}\left[-\int \psi_R \rho_s \log \rho_s\right] = -\frac{1}{2N}\int \psi_R \cdot (\Delta\rho_s \log \rho_s + \Delta\rho_s)$$

$$= \frac{1}{2N}\int \psi_R \frac{|\nabla\rho_s|^2}{\rho_s} + \frac{1}{2N}\int \nabla\psi_R \cdot \nabla\rho_s \cdot (1 + \log \rho_s),$$

where all the integrals are taken over $M_N(\mathbb{C})_{\mathrm{sa}}^m$ with respect to dx. This implies

$$- \int \psi_R \rho_t \log \rho_t + \int \psi_R \rho_\delta \log \rho_\delta$$
$$= \frac{1}{2N} \int_\delta^t \int \psi_R \left| \frac{\nabla \rho_s}{\rho_s} \right|^2 d\mu_s - \frac{1}{2N} \int_\delta^t \int \left(\nabla \psi_R \cdot \frac{\nabla \rho_s}{\rho_s} \right) (1 + \log \rho_s)\, d\mu_s. \quad (5\text{-}8)$$

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

$$\lim_{R \to +\infty} \int_\delta^t \int \psi_R \left| \frac{\nabla \rho_s}{\rho_s} \right|^2 d\mu_s\, ds = \int_\delta^t \mathcal{I}(\mu_s)\, ds.$$

The second term on the right-hand side of (5-8) can be estimated as follows. Note that $\mu_s = \mu * \sigma_{s,N}$ and that $\sigma_{s,N}$ has a density that is bounded uniformly for $s \in [\delta, t]$ and $x \in M_N(\mathbb{C})_{\mathrm{sa}}^m$. Therefore, ρ_s is uniformly bounded for $s \in [\delta, t]$ and $x \in M_N(\mathbb{C})_{\mathrm{sa}}^m$ 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

$$\mu(x : |x| \le K) \ge \tfrac{1}{2}.$$

Now if $x \in M_N(\mathbb{C})_{\mathrm{sa}}^m$ and $|y| \le K$, then $|x - y| \ge |x| - K$ and hence

$$|x - y|^2 \le |x|^2 - 2K|x| + K^2 \ge 2|x|^2 + 2K^2,$$

where the last inequality follows because $2K|x| \le |x|^2/2 + 2K^2$ by the arithmetic-geometric mean inequality. Therefore, letting Z be the normalizing constant for $\sigma_{t,N}$, we have

$$\rho_s(x) = \frac{1}{Z} \int e^{-(N/(2t))|x-y|^2}\, d\mu(y) \ge \frac{1}{Z} \int_{|y| \le K} e^{-(N/(2t))|x-y|^2}\, d\mu(y)$$
$$\ge \frac{1}{Z} \int_{|y| \le K} e^{-(N/t)(|x|^2 + K^2)}\, d\mu(y) \ge \frac{e^{-NK^2/t}}{2Z} e^{-(N/t)|x|^2},$$

so that $\log \rho_s \ge K' - |x|^2$ 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 $|1 + \log \rho_s| \le \alpha |x|^2$. Recall that $\nabla \psi_R(x)$ is supported when $R \le |x| \le 2R$ and bounded by C/R and thus $|\nabla \psi_R(x)| \le C/|x|$. Altogether we have $|\nabla \psi_R (1 + \log \rho_s)| \le \beta |x|$ for some constant β when $|x|$ is large enough. Thus, the second term on the right-hand side of (5-8) is bounded by

$$\int_\delta^t \int |(\nabla \psi_R \cdot \Xi_s)(1 + \log \rho_s)|\, d\mu_s\, ds \le \beta \int_\delta^t \int_{|x| \ge R} |x| \left| \frac{\nabla \rho_s(x)}{\rho_s(x)} \right| d\mu_s(x)\, ds$$
$$\le \frac{\beta}{2} \int_\delta^t \int_{|x| \ge R} \left(|x|^2 + \left| \frac{\nabla \rho_s(x)}{\rho_s(x)} \right|^2 \right) d\mu_s(x)\, ds.$$

The right-hand side is the tail of the convergent integral

$$\int_\delta^t \int \left(|x|^2 + \left| \frac{\nabla \rho_s(x)}{\rho_s(x)} \right|^2 \right) d\mu_s(x)\, ds = \int_\delta^t [(a + ms) + \mathcal{I}(\mu_s)]\, ds < +\infty,$$

and therefore it goes to zero as $R \to +\infty$ by the dominated convergence theorem.

As for the left-hand side of (5-8), 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 (5-8), we obtain (5-7).

To complete the proof of (5-5), we must take $\delta \searrow 0$ in (5-7). We can take the limit of the right-hand side of (5-7) by the monotone convergence theorem. For the left-hand side of (5-7), Lemma 5.1(3) implies that $\limsup_{\delta \searrow 0} h(\mu_\delta) \le h(\mu)$ because $\rho_\delta \to \rho$ almost everywhere by Lebesgue differentiation theory. On the other hand, $h(\mu_\delta) \ge h(\mu)$ by Lemma 5.1(4); hence $h(\mu_\delta) \to h(\mu)$, so (5-5) is proved.

To prove (5-6), we follow [Voiculescu 1998, Proposition 7.6]. First, suppose that $h(\mu) > -\infty$. Note that

$$h(\mu) = \frac{1}{2} \int_0^t \left(\frac{mN^2}{1+s} - \frac{1}{N} \mathcal{I}(\mu_s) \right) ds - \frac{mN^2}{2} \log(1+t) + h(\mu_t).$$

If $h(\mu) > -\infty$, then

$$\int_0^1 \left(\frac{mN^2}{1+s} - \frac{1}{N} \mathcal{I}(\mu_s) \right) ds$$

is finite. In light of (5-4), the integral from 1 to ∞ is also finite and by the dominated convergence theorem

$$\lim_{t \to \infty} \frac{1}{2} \int_0^t \left(\frac{mN^2}{1+s} - \frac{1}{N} \mathcal{I}(\mu_s) \right) ds = \frac{1}{2} \int_0^\infty \left(\frac{mN^2}{1+s} - \frac{1}{N} \mathcal{I}(\mu_s) \right) ds.$$

It remains to understand the behavior of $h(\mu_t) - (mN^2/2) \log(1+t)$. By Lemma 5.1(4) and (2),

$$h(\mu_t) \ge h(\sigma_{t,N}) = \frac{mN^2}{2} \log \frac{2\pi e t}{N} = \frac{mN^2}{2} \log \frac{2\pi e}{N} + \frac{mN^2}{2} \log t.$$

On the other hand, by Lemma 5.1(2), since $\int |x|^2 \, d\mu_t(x) = N(a+tm)$, we have

$$h(\mu_t) \le \frac{mN^2}{2} \log \frac{2\pi e(a+t)}{N} = \frac{mN^2}{2} \log \frac{2\pi e}{N} + \frac{mN^2}{2} \log(a+t).$$

As $t \to \infty$, we have $\log(1+t) - \log(a+t) \to 0$ and $\log(1+t) - \log t \to 0$ and therefore

$$h(\mu_t) - \frac{mN^2}{2} \log(1+t) \to \frac{mN^2}{2} \log \frac{2\pi e}{N} = \frac{mN^2}{2} \log 2\pi e - \frac{mN^2}{2} \log N.$$

Hence,

$$h(\mu) = \frac{1}{2} \int_0^\infty \left(\frac{mN^2}{1+s} - \frac{1}{N} \mathcal{I}(\mu_s) \right) ds + \frac{mN^2}{2} \log 2\pi e - \frac{mN^2}{2} \log N,$$

which is equivalent to the asserted formula (5-6). In the case where $h(\mu) = -\infty$, we also have

$$\int_0^1 \left(\frac{mN^2}{1+s} - \frac{1}{N} \mathcal{I}(\mu_s) \right) ds = -\infty$$

by (5-5), but the integral from 1 to ∞ is finite as shown above. So both sides of (5-6) are $-\infty$. $\qquad \square$

5D. *Free Fisher information.* The starting point for the definition of free Fisher information is the integration-by-parts formula (5-3). Indeed, if we formally apply this to a noncommutative polynomial p and renormalize, we obtain

$$\int \tau_N \left(\frac{1}{N} \Xi_j(x) p(x) \right) d\mu(x) = \int \tau_N \otimes \tau_N (\mathcal{D}_j p(x)) \, d\mu(x), \qquad (5\text{-}9)$$

(and this integration by parts is justified under sufficient assumptions of finite moments). Voiculescu therefore made the following definitions:

Definition 5.8 [Voiculescu 1998, §3]. Let $X = (X_1, \ldots, X_m)$ 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 $\xi = (\xi_1, \ldots, \xi_m) \in L^2(\mathcal{M}, \tau)^m$ is a *(free) conjugate variable* of X if

$$\tau(\xi_j p(X)) = \tau \otimes \tau(\mathcal{D}_j p(X)) \tag{5-10}$$

for every noncommutative polynomial p. The free conjugate variable, if it exists, is unique. If it exists, we say that X (or equivalently the law of X) has finite free Fisher information and define $\Phi^*(X) := \Phi^*(\lambda_X) := \sum_j \tau(\xi_j^2)$. We also denote the conjugate variable ξ by $J(X)$.

Definition 5.9 [Voiculescu 1998, Definition 7.1]. The *nonmicrostates free entropy* of a noncommutative law λ is

$$\chi^*(\lambda) := \frac{1}{2} \int_0^\infty \left(\frac{m}{1+t} - \Phi^*(\lambda \boxplus \sigma_t) \right) + \frac{1}{2} \log 2\pi e.$$

Now we are ready to state conditions under which the classical Fisher information of a sequence of measures μ_N converges to the free Fisher information of the law λ. First, to clarify the normalization, note that if $d\mu_N(x) = (1/Z_N) \exp(-N^2 V_N(x)) \, dx$, then the classical conjugate variable is given by $\Xi_N = N^2 \nabla V_N$. The normalized conjugate variable used in (5-9) is $(1/N)\Xi_N = N\nabla V_N = DV_N$. The corresponding normalized Fisher information is then

$$\int \|DV_N\|_2^2 \, d\mu_N = \int \frac{1}{N} \left| \frac{1}{N} \Xi_N \right|^2 d\mu = \frac{1}{N^3} \mathcal{I}(\mu_N),$$

which is the same normalization as in Lemma 5.7.

Proposition 5.10. *Let* $V_N : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$ *be a potential with* $\int \exp(-N^2 V_N(x)) \, dx < +\infty$ *and let* μ_N *be the associated measure. Assume*:

(A) *The sequence* μ_N *concentrates around a noncommutative law* λ.

(B) *The sequence* $\{DV_N\}$ *is asymptotically approximable by trace polynomials.*

(C) *For some* $n \geq 0$ *and* $a, b > 0$ *we have* $\|DV_N\|_2^2 \leq a + b \sum_{j=1}^m \tau_N(x_j^{2n})$.

(D) *There exists* $R_0 > 0$ *such that*

$$\lim_{N \to \infty} \int_{\|x\|_\infty \geq R_0} \left(1 + \sum_{j=1}^m \tau_N(x_j^{2n}) \right) d\mu_N(x) = 0.$$

Then:

(1) *The law* λ *can be realized by self-adjoint random variables* $X = (X_1, \ldots, X_m)$ *in a tracial von Neumann algebra* $(\mathcal{M}, \tau)$ *with* $\|X_j\| \leq R_0$.

(2) *There exists a sequence of trace polynomials* $f^{(k)} \in (\mathrm{TrP}_m^1)^m$ *such that*

$$\lim_{k \to \infty} \limsup_{N \to \infty} \sup_{\|x\|_\infty \leq R_0} \|DV_N(x) - f^{(k)}(x)\|_2 = 0.$$

(3) *If $\{f^{(k)}\}$ is any sequence as in (2), then $\{f_k(X)\}$ converges in $L^2(\mathcal{M}, \tau)$ and the limit is the conjugate variable $J(X)$.*

(4) *The law λ has finite free Fisher information and $N^{-3}\mathcal{I}(\mu_N) \to \Phi^*(\lambda)$ as $N \to \infty$.*

Proof. (1) This follows from the same argument as Proposition 5.5.

(2) This follows from the definition of asymptotic approximability by trace polynomials.

(3) Let $\{f^{(k)}\}$ be a sequence as in (2). Because μ_N concentrates around λ and $\mu_N(\{x : \|x\|_\infty \le R_0\}) \to 1$ as $N \to +\infty$ by (D), we have

$$\lambda[(f^{(j)} - f^{(k)})^*(f^{(j)} - f^{(k)})] = \lim_{N \to \infty} \int_{\|x\|_\infty \le R_0} \tau_N[(f^{(j)} - f^{(k)})^*(f^{(j)} - f^{(k)})(x)] \, d\mu_N(x).$$

For every $\epsilon > 0$, if j and N are large enough, then $\sup_{\|x\|_\infty \le R_0} \|DV_N(x) - f^{(j)}(x)\|_2 < \epsilon$ by our assumption on $f^{(j)}$. In particular, if j and k are sufficiently large, then $\lambda[(f^{(j)} - f^{(k)})^*(f^{(j)} - f^{(k)})] < (2\epsilon)^2$. This shows that $\{f^{(k)}(X)\}$ is Cauchy in $L^2(M, \lambda)$ since X has the law λ.

Let $\xi = \lim_{k \to \infty} f^{(k)}(X)$. We must show that ξ is the conjugate variable for X. Let $\psi \in C_c^\infty(\mathbb{R})$ such that $\psi(y) = y$ when $|y| \le R_0$. For $x \in M_N(\mathbb{C})_{\text{sa}}^m$, let $\Psi(x) = (\psi(x_1), \dots, \psi(x_m))$. By (5-9), because $DV_N(x)$ is the classical conjugate variable for X, we have for every noncommutative polynomial p that

$$\int \tau_N[D_j V_N(x) \cdot p(\Psi(x))] \, d\mu_N(x) = \int D_j[\tau_N(p(\Psi(x)))] \, d\mu_N(x).$$

It follows from our assumptions (C) and (D) that

$$\lim_{N \to \infty} \int_{\|x\|_\infty \ge R_0} \|DV_N(x)\|_2^2 \, d\mu_N(x) = 0.$$

Because $p(\Psi(x))$ and $D_j[\tau_N(p(\Psi(x)))]$ are globally bounded in operator norm, the integral of these quantities over $\|x\|_\infty \ge R_0$ will vanish as $N \to \infty$ and therefore

$$\int_{\|x\|_\infty \le R_0} \tau_N[D_j V_N(x) p(\Psi(x))] \, d\mu_N(x) - \int_{\|x\|_\infty \le R_0} D_j[\tau(p(\Psi(x)))] \, d\mu_N(x) \to 0.$$

But since $p(\Psi(x)) = p(x)$ on this region, we have

$$\int_{\|x\|_\infty \le R_0} \tau_N[D_j V_N(x) p(x)] \, d\mu_N(x) - \int_{\|x\|_\infty < R_0} \tau_N \otimes \tau_N[\mathcal{D}_j p(x)] \, d\mu_N(x) \to 0.$$

Now the second term converges to $\lambda \otimes \lambda[\mathcal{D}_j p] = \tau \otimes \tau[\mathcal{D}_j p(X)]$ by our concentration assumption (A). For the first term, we can replace $D_j V_N(x)$ by $f_j^{(k)}(x)$ with an error bounded by

$$\sup_{\|x\|_\infty \le R_0} \|f^{(k)}(x) - DV_N(x)\|_2.$$

Then we apply concentration to conclude that

$$\int \tau_N[f_j^{(k)}(x)^* p(x)] \, d\mu_N(x) \to \lambda[(f_j^{(k)})^* p].$$

Overall,

$$\left|\lambda[(f_j^{(k)})^* p] - \lambda \otimes \lambda[\mathcal{D}_j p]\right| \le \limsup_{N \to \infty} \sup_{\|x\|_\infty \le R_0} \|f^{(k)}(x) - DV_N(x)\|_2.$$

Taking $k \to \infty$, we obtain $\tau[\xi_j p(X)] - \tau \otimes \tau[\mathcal{D}_j p(X)] = 0$ as desired.

(4) We know from (3) that λ has finite Fisher information. Assumptions (C) and (D) imply that

$$\frac{1}{N^3}\mathcal{I}(\mu_N) - \int_{\|x\|_\infty \le R_0} \|DV_N(x)\|_2^2 \, d\mu_N(x) \to 0.$$

By arguments similar to those before, we can approximate DV_N by $f^{(k)}$ on $\|x\|_\infty \le R_0$, approximate $\int_{\|x\|_\infty \le R_0} \|f^{(k)}\|_2^2 \, d\mu_N$ by $\lambda((f^{(k)})^* f^{(k)})$, and then approximate $\lambda((f^{(k)})^* f^{(k)})$ by $\tau(\xi^* \xi) = \Phi^*(\lambda)$, where the error terms vanish as $N \to \infty$ and then $k \to \infty$. This implies that $N^{-3}\mathcal{I}(\mu_N) \to \Phi^*(\lambda)$. $\square$

6. Evolution of the conjugate variables

6A. *Motivation and statement of the equation.* In the last section, we stated conditions under which the classical entropy and Fisher information of μ_N converge to their free counterparts for the limiting noncommutative law λ. In order to prove that $\chi(\lambda) = \chi^*(\lambda)$, we want to take the limit in the integral formula (5-6), and therefore, we want $N^{-3}\mathcal{I}(\mu_N * \sigma_{t,N}) \to \Phi^*(\lambda \boxplus \sigma_t)$ for all $t \ge 0$. In order to apply Proposition 5.10 to $\mu_N * \sigma_{t,N}$, we need to show that $\{DV_{N,t}\}_N$ is asymptotically approximable by trace polynomials, where $V_{N,t}$ is the potential corresponding to $\mu_N * \sigma_{t,N}$.

By adding a constant to each V_N if necessary, we may assume without loss of generality that $Z_N = 1$. We recall that $V_{N,t}(x)$ is given by

$$\exp(-N^2 V_{N,t}(x)) = \int \exp(-N^2 V_N(x+y)) \, d\sigma_{t,N}(y). \tag{6-1}$$

Then $\exp(-N^2 V_{N,t}(x))$ solves the normalized heat equation

$$\partial_t[\exp(-N^2 V_{N,t}(x))] = \frac{1}{2N}\Delta[\exp(-N^2 V_{N,t}(x))], \tag{6-2}$$

where $(1/N)\Delta = L_N$ is the normalized Laplacian. However, we do not know how to show that $DV_N(\,\cdot\,, t)$ is asymptotically approximable by trace polynomials from a direct analysis of the heat equation because of the dimension-dependent factor of N^2 in the exponent. What we want is a dimension-independent and "hands-on" way of producing $V_{N,t}$ from V_N.

As in Section 4, we will analyze the PDE which describes the evolution of the function $V_{N,t}$. First, let us derive the equation by rewriting (6-2) in terms of $V_{N,t}$ rather than $e^{-N^2 V_{N,t}}$. By the chain rule,

$$\partial_t[\exp(-N^2 V_{N,t})] = -N^2 \partial_t V_{N,t} \cdot \exp(-N^2 V_{N,t})$$

and

$$\Delta[\exp(-N^2 V_{N,t})] = [\Delta(-N^2 V_{N,t}) + |\nabla(-N^2 V_{N,t})|^2]\exp(-N^2 V_{N,t})$$
$$= (-N^2 \Delta V_{N,t} + N^4 |\nabla V_{N,t}|^2)\exp(-N^2 V_{N,t}),$$

where Δ and ∇ denote the classical (nonnormalized) Laplacian and gradient, where $M_N(\mathbb{C})_{\mathrm{sa}}^m$ has been identified with $\mathbb{R}^{mN^2}$ using the coordinates in Section 2A. Thus, our equation becomes

$$-N^2 \partial_t V_{N,t} = \frac{1}{2N}(-N^2 \Delta V_{N,t} + N^4 |\nabla V_{N,t}|^2),$$

$$\partial_t V_{N,t} = \frac{1}{2N} \Delta V_{N,t} - \frac{N}{2} |\nabla V_{N,t}|^2.$$

Recall that $(1/N)\Delta$ is the normalized Laplacian discussed in Section 3B. The normalized gradient is $DV_{N,t} = N\nabla V_{N,t}$, and the normalized Euclidean norm is

$$\|x\|_2^2 = \sum_{j=1}^m \tau_N(x_j^2) = \frac{1}{N}\sum_{j=1}^m \mathrm{Tr}(x_j^2) = \frac{1}{N}|x|^2.$$

Then

$$N|\nabla V_{N,t}|^2 = \frac{1}{N}|N\nabla V_{N,t}|^2 = \frac{1}{N}|DV_{N,t}|^2 = \|DV_{N,t}\|_2^2,$$

and therefore we obtain the following equation that is normalized in a dimension-independent way:

$$\partial_t V_{N,t} = \tfrac{1}{2} L_N V_{N,t} - \tfrac{1}{2}\|DV_{N,t}\|_2^2. \tag{6-3}$$

In the remainder of this section, we study a semigroup R_t acting on convex and semiconcave functions on $M_N(\mathbb{C})_{\mathrm{sa}}^m$ such that $V_{N,t} = R_t V_N$ (here R_t depends implicitly on N). In Sections 6B–6F, we construct R_t from scratch by iterating the heat semigroup and Hopf–Lax semigroup. Next, in Section 6G, we verify that $R_t V_N$ solves (6-3) in the *viscosity sense* (for background, see [Crandall et al. 1992]), and deduce that $R_t V_N$ must agree with the smooth solution $V_{N,t}$ defined by (6-1). Finally, in Section 6H, we show that if $\{DV_N\}$ is asymptotically approximable by trace polynomials, then so is $\{D(R_t V_N)\}$.

6B. *Strategy to approximate solutions.* To construct the semigroup R_t that solves (6-3), we view the equation as a hybrid between the heat equation $\partial_t u = (2N)^{-1}\Delta u$ and the Hamilton–Jacobi equation with quadratic potential $\partial_t u = -\frac{1}{2}\|Du\|_2^2$. The heat equation can be solved by the heat semigroup

$$P_t u(x) := \int u(x+y)\, d\sigma_{t,N}(y), \tag{6-4}$$

while the Hamilton–Jacobi equation can be solved using the inf-convolution semigroup

$$Q_t u(x) := \inf\left[u(x+y) + \frac{1}{2t}\|y\|_2^2 \right] \tag{6-5}$$

as a special case of the Hopf–Lax formula (see [Evans 2010, Chapter 3.3]).

In Dabrowski's approach, the solution to (6-3) was expressed through a formula of Boué, Dupuis, and Üstunel as the infimum of $E\left[u\left(x + B_t + \int_0^t Y_s\, ds\right) + \frac{1}{2}\int_0^t \|Y_s\|_2^2\, ds \right]$ over a certain class of stochastic processes Y_t adapted to a standard Brownian motion B_t (see [Dabrowski 2016, Theorem 3.1]). This formula, roughly speaking, combines the Gaussian convolution and inf-convolution operations by replacing the y in the definition of Q_t by a stochastic process and allowing it to evolve with B_t. Dabrowski [2016, Section 5] then identifies the minimizing process Y_t as a Brownian bridge and analyzes it using a forward-backward SDE. Through the Picard iteration for solving the SDE, he shows that the solution is well-approximated by noncommutative functions.

We instead give a deterministic proof following the same strategy as in Section 4 that is motivated by Trotter's formula, we define a semigroup $R_t u$ at dyadic times t by alternating between $P_{2^{-\ell}}$ and $Q_{2^{-\ell}}$ and then letting $\ell \to \infty$. We establish convergence through a telescoping series argument after showing that $P_t Q_t - Q_t P_t = o(t)$. Then we show that $R_t u$ depends continuously on t in order to extend its definition to all positive real t.

In contrast to Section 4, we must understand how the semigroups P_t, Q_t, and R_t affect Du as well as u, and we want $D(R_t u)$ to be Lipschitz for all t. We therefore view these operators as acting on spaces of the form

$$\mathcal{E}(c, C) = \{u : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}, \, u(x) - (c/2)\|x\|_2^2 \text{ is convex and } u(x) - (C/2)\|x\|_2^2 \text{ is concave}\},$$

where $0 \leq c \leq C < +\infty$, where we suppress the dependence on m and N in the notation. These spaces have the virtue that if $u \in \mathcal{E}(c, C)$, then $\|Du\|_{\mathrm{Lip}} \leq C$ automatically (see Proposition 2.13(4)).

At every step of the proof, we include estimates both for u and for Du. In addition, controlling the error propagation requires more work because Q_t and R_t are not contractions with respect to $\|Du\|_{L^\infty}$.

The following theorem summarizes the results of the construction. To clarify the notation, for a measurable function $u : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$, the notation $\|u\|_{L^\infty}$ is the standard L^∞ norm. If $F : M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})_{\mathrm{sa}}^m$ (for instance $F = Du$ for some $u : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$), then $\|F\|_{L^\infty} = \sup_{x \in M_N(\mathbb{C})_{\mathrm{sa}}^m} \|F(x)\|_2$; similarly, $\|F\|_{\mathrm{Lip}}$ is the Lipschitz norm of F when using $\|\cdot\|_2$ in both the domain and the target space. Note that $\|F\|_2$ does *not* denote the L^2 norm of F with respect to any measure, but rather $\left(\sum_{j=1}^m \tau(F_j^2)\right)^{1/2}$, which is a function of x. Recall that $\mathbb{Q}_2^+$ denotes the nonnegative dyadic rationals, and $\mathbb{N}_0$ denote the natural numbers with zero included. Moreover, we assume throughout the section that $0 \leq c \leq C < +\infty$.

Theorem 6.1. *There exists a semigroup of nonlinear operators* $R_t : \bigcup_{C>0} \mathcal{E}(0, C) \to \bigcup_{C>0} \mathcal{E}(0, C)$ *with the following properties*:

(1) *Change in convexity*: *If* $u \in \mathcal{E}(c, C)$ *where* $0 \leq c \leq C$, *then* $R_t u \in \mathcal{E}(c(1+ct)^{-1}, C(t+Ct)^{-1})$.

(2) *Approximation by iteration*: *For* $\ell \in \mathbb{Z}$ *and* $t \in 2^{-\ell}\mathbb{N}_0$, *write* $R_{t,\ell} u = (P_{2^{-\ell}} Q_{2^{-\ell}})^{2^\ell t} u$. *Fix such a value of* t *and fix* $u \in \mathcal{E}(0, C)$.

   (a) *If* $2^{-\ell-1}C \leq 1$, *then*

$$|R_t u - R_{t,\ell} u| \leq \left(\frac{3}{2}\frac{C^2 mt}{1+Ct} + \log(1+Ct)(m+Cm+\|Du\|_2^2)\right)2^{-\ell}.$$

   (b) $\|D(R_{t,\ell} u) - D(R_t u)\|_{L^\infty} \leq [t/2 + C(t/2)^2]C^2 m^{1/2}(2 \cdot 2^{-\ell/2} + 2^{-3\ell/2}C)$.

(3) *Continuity in time*: *Suppose* $s \leq t \in \mathbb{R}_+$ *and* $u \in \mathcal{E}(0, C)$.

   (a) $R_t u \leq R_s u + (m/2)[\log(1+Ct) - \log(1+Cs)]$.

   (b) $R_t u \geq R_s u - ((t-s)/2)(Cm + \|Du\|_2^2)$.

   (c) *If* $C(t-s) \leq 1$, *then* $\|D(R_t u) - D(R_s u)\|_2 \leq 5Cm^{1/2}2^{1/2}(t-s)^{1/2} + C(t-s)\|Du\|_2$.

(4) *Error estimates*: *Let* $t \in \mathbb{R}_+$ *and* $u, v \in \mathcal{E}(0, C)$. *Then*:

   (a) $\|D(R_t u) - D(R_t v)\|_{L^\infty} \leq (1+Ct)\|Du - Dv\|_{L^\infty}$.

(b) *If $u \leq v + a + b\|Dv\|_2^2$, where $a \in \mathbb{R}$ and $b \geq 0$, then*

$$R_t u \leq R_t v + a + b\frac{C^2 mt}{1 + Ct} + b\|D(R_t v)\|_2^2.$$

(c) *We have*

$$\|D(R_t u)\|_2^2 \leq \frac{C^2 mt}{1 + Ct} + \|Du\|_2^2.$$

Remark 6.2. Knowing that $\exp(-N^2(R_t u)) = P_t \exp(-N^2 u)$, one can deduce (1) from the Brascamp–Lieb and Hölder inequalities, as in [Brascamp and Lieb 1976, Theorem 4.3]. But the proof of (1) given here is independent of that work.

We also point out that semigroups and discrete-time approximation schemes have been employed to study Hamilton–Jacobi equations in Hilbert space (e.g., by [Barbu 1986]), another setting that requires dimension-independent estimates.

6C. *The Hopf–Lax semigroup, the heat semigroup, and convexity.* We remind the reader of our standing assumption that $0 \leq c \leq C$.

Lemma 6.3. *Suppose $u \in \mathcal{E}(c, C)$. Then*:

(1) $P_t u \in \mathcal{E}(c, C)$.

(2) $\|D(P_t u) - Du\|_{L^\infty} \leq Cm^{1/2}t^{1/2}$.

Proof. (1) This follows because $\mathcal{E}(c, C)$ is closed under translation and averaging, and hence convolution by a probability measure.

(2) We know that Du is C-Lipschitz and thus

$$\|D(P_t u)(x) - Du(x)\|_2 \leq \int \|Du(x + y) - Du(x)\|_2 \, d\sigma_{t,N}(y)$$

$$\leq \int C\|y\|_2 \, d\sigma_{t,N}(y) \leq Cm^{1/2}t^{1/2}. \qquad \square$$

The following lemma gives basic properties of Q_t from the PDE literature; see for instance [Ekeland and Lasry 1980, pp. 309–311; Lasry and Lions 1986; Crandall et al. 1992, Lemma A.5; Evans 2010, Section 3.3.2]. For completeness and convenience, we include a proof of all the facts we will use.

Lemma 6.4. (1) *If $u, v : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$ and $u \leq v$, then $P_t u \leq P_t v$ and $Q_t u \leq Q_t v$.*

(2) *Suppose that $v(x) = a + \frac{1}{2}\langle p, x \rangle_2 \langle Ax, x \rangle_2$, where $a \in \mathbb{R}$, $p \in M_N(\mathbb{C})_{\mathrm{sa}}^m$, and A is a positive semidefinite linear map $M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})_{\mathrm{sa}}^m$. Then*

$$P_t v(x) = a + \frac{t}{2N^2}\operatorname{Tr}(A) + \langle p, x \rangle + \frac{1}{2}\langle Ax, x \rangle_2,$$

$$Q_t v(x) = a - \frac{t}{2}\|p\|^2 + \langle p, x \rangle + \frac{1}{2}\langle A(1 + tA)^{-1}(x - tp), x - tp \rangle.$$

Remark 6.5. Here $\operatorname{Tr}(A)$ denotes the trace of A as a linear transformation of the vector space $M_N(\mathbb{C})_{\mathrm{sa}}^m$, which is well-defined because the trace of a matrix is similarity-invariant. In particular, we may compute

$\mathrm{Tr}(A)$ using an orthonormal basis of $M_N(\mathbb{C})_{\mathrm{sa}}^m$, and the choice of basis and of the normalization of the inner product does not matter. Note that the trace of the identity is mN^2, which makes the normalization in the above formula dimension-independent.

Proof of Lemma 6.4. (1) is immediate to check from the definition. We leave the first formula of (2) as an exercise. To prove the last formula, fix $t > 0$ and $x \in M_N(\mathbb{C})_{\mathrm{sa}}^m$ and note that $u(y) + (1/(2t))\|y - x\|_2^2$ is a uniformly convex function of y and therefore it has a unique minimizer. The minimizer y must be a critical point and hence

$$0 = Du(y) + \frac{1}{t}(y - x) = p + Ay + \frac{1}{t}(y - x).$$

Thus, $(1 + tA)y = x - tp$ and $y - x = -t(p + Ay)$, so that

$$Q_t u(x) = u(y) + \frac{1}{2t}\|y - x\|_2^2,$$

which reduces after several lines of manipulation to the claimed formula. $\square$

Lemma 6.6. *Let $u \in \mathcal{E}(c, C)$.*

(1) *The operators $\{Q_t\}_{t \geq 0}$ form a semigroup; that is, $Q_s Q_t u = Q_{s+t} u$ for $s, t \geq 0$.*

(2) *For each $x_0 \in M_N(\mathbb{C})_{\mathrm{sa}}^m$ and $t \geq 0$, the infimum $Q_t u(x_0) = \inf_y [u(y) + (t/2)\|y - x_0\|_2^2]$ is achieved at a unique point y_0 satisfying $y_0 = x_0 - t Du(y_0)$.*

(3) *If $x_0 \in M_N(\mathbb{C})_{\mathrm{sa}}^m$ and y_0 is the minimizer from (2), then $D(Q_t u)(x_0) = Du(y_0)$.*

(4) *We have $Q_t u \in \mathcal{E}(c(1 + ct)^{-1}, C(1 + Ct)^{-1})$.*

(5) *$\|D(Q_t u)(x_0)\|_2 = \|Du(y_0)\|_2 \leq (1 + ct)^{-1}\|Du(x_0)\|_2$.*

Proof. (1) By definition

$$Q_s Q_t u(x) = \inf_y \left[Q_t u(y) + \frac{1}{2s}\|x - y\|_2^2 \right] = \inf_y \inf_z \left[u(z) + \frac{1}{2t}\|y - z\|_2^2 + \frac{1}{2s}\|x - y\|_2^2 \right]$$

$$= \inf_z \left[u(z) + \inf_y \left[\frac{1}{2t}\|y - z\|_2^2 + \frac{1}{2s}\|x - y\|_2^2 \right] \right].$$

But note that

$$\inf_y \left[\frac{1}{2t}\|y - z\|_2^2 + \frac{1}{2s}\|x - y\|_2^2 \right]$$

is by definition $Q_s f(z)$, where $f(x) = (1/(2t))\|x - z\|_2^2$. If $g(x) = (1/(2t))\|x\|_2^2$, then by the previous lemma, we have

$$Q_s g(x) = \frac{1}{2} \frac{t^{-1}}{1 + t^{-1}s}\|x\|_2^2 = \frac{1}{2(s+t)}\|x\|_2^2.$$

Since Q_s is clearly translation-invariant, $Q_s f(x) = ((s + t)/2)\|x - z\|_2^2$. Therefore,

$$Q_s Q_t u(x) = \inf_z \left[u(z) + \frac{1}{2(s+t)}\|x - z\|_2^2 \right] = Q_{s+t} u(x).$$

(2) Fix x_0. Note that the function $y \mapsto [u(y) + (1/(2t))\|y - x_0\|_2^2]$ is in $\mathcal{E}(c + 1/(2t), C + 1/(2t))$ and hence it achieves a global minimum at the unique critical point. Thus, the infimum is achieved at the point y_0 satisfying $Du(y_0) = (1/t)(y_0 - x_0)$, or in other words $y_0 = x_0 - tDu(y_0)$.

(3), (4) Let x_0 and y_0 be as above. Let $p = Du(y_0)$. Because $u \in \mathcal{E}(c, C)$, we have for all x that

$$u(y_0) + \langle p, x - y_0 \rangle_2 + \frac{c}{2}\langle x - y_0 \rangle \leq u(x) \leq u(y_0) + \langle p, x - y_0 \rangle_2 + \frac{C}{2}\|x - y_0\|_2^2.$$

Let $\underline{v}(y)$ and $\bar{v}(y)$ be the functions on the left and right-hand sides. Then by Lemma 6.4 (1), we have $Q_t\underline{v} \leq Q_tu \leq Q_t\bar{v}$. To compute $Q_t\underline{v}$, we apply Lemma 6.4 (2) with $A = cI$ and with a change of coordinates to translate y_0 to the origin, and we obtain

$$Q_t\underline{v}(x) = u(y_0) - \frac{t}{2}\|p\|_2^2 + \langle p, x - y_0 \rangle + \frac{c}{2}(1 + ct)^{-1}\|x - y_0 - tp\|^2.$$

Since $y_0 + tp = x_0$ and $p = (y_0 - x_0)/t$, this becomes

$$Q_t\underline{v}(x) = u(y_0) - \frac{t}{2}\|p\|_2^2 + t\|p\|_2^2 + \langle p, x - x_0 \rangle + \frac{c}{2}(1 + ct)^{-1}\|x - x_0\|_2^2$$

$$= u(y_0) + \frac{1}{2t}\|y_0 - x_0\|_2^2 + \langle p, x - x_0 \rangle + \frac{c}{2}(1 + ct)^{-1}\|x - x_0\|_2^2$$

$$= Q_tu(x_0) + \langle p, x - x_0 \rangle + \frac{c}{2}(1 + ct)^{-1}\|x - x_0\|_2^2.$$

The analogous computation holds for $Q_t\bar{v}$ as well. Thus, we have

$$Q_tu(x_0) + \langle p, x - x_0 \rangle + \frac{c}{2}(1 + ct)^{-1}\|x - x_0\|_2^2 \leq Q_tu(x) \leq Q_tu(x_0) + \langle p, x - x_0 \rangle + \frac{C}{2}(1 + Ct)^{-1}\|x - x_0\|_2^2.$$

This inequality implies that $D(Q_tu)(x_0) = p = Du(y_0)$. Since the above inequality holds for every x_0, we see that $Q_tu \in \mathcal{E}(c(1 + ct)^{-1}, C(1 + Ct)^{-1})$ by Proposition 2.13(2).

(5) Let x_0, y_0, and p be as above. Then we have

$$\langle Du(y_0) - Du(x_0), y_0 - x_0 \rangle_2 \geq c\|y_0 - x_0\|_2^2.$$

But recall that $y_0 - x_0 = -tDu(y_0)$ and hence

$$-t\langle Du(y_0) - Du(x_0), Du(y_0) \rangle \geq ct^2\|Du(y_0)\|_2^2.$$

Rearranging produces

$$(1 + ct)\|Du(y_0)\|_2^2 \leq \langle Du(x_0), Du(y_0) \rangle_2 \leq \|Du(x_0)\|_2\|Du(y_0)\|_2,$$

and hence $(1 + ct)\|Du(y_0)\|_2 \leq \|Du(x_0)\|_2$ as desired. □

Corollary 6.7. *Let $u \in \mathcal{E}(c, C)$ and $s, t \geq 0$.*

(1) *For each x, the gradient $D(Q_tu)(x)$ is the unique vector p satisfying $p = Du(x - tp)$.*

(2) *We have $Q_tu(x) = u(x - tD(Q_tu)(x)) + (t/2)\|D(Q_tu)(x)\|_2^2$.*

(3) *$u(x) - (t/2)(1 + Ct)\|D(Q_tu)(x)\|^2 \leq Q_tu(x) \leq u(x) - (t/2)(1 + ct)\|D(Q_tu)(x)\|_2^2$.*

Proof. (1) and (2) follow from [Lemma 6.6](2) and (3).

To prove (3), fix x and let $y = x - tD(Q_t u)(x)$. By [Proposition 2.13](2),

$$u(y) + \langle Du(y), x - y \rangle_2 + \frac{c}{2}\|x - y\|_2^2 \le u(x) \le u(y) + \langle Du(y), x - y \rangle + \frac{C}{2}\|x - y\|_2^2.$$

Hence,

$$u(x) - \langle Du(y), x - y \rangle_2 - \frac{C}{2}\|x - y\|_2^2 \le u(y) \le u(x) - \langle Du(y), x - y \rangle_2 - \frac{c}{2}\|x - y\|_2^2.$$

But from the previous lemma, we know that $Du(y) = D(Q_t u)(x)$ and $x - y = tD(Q_t u)(x)$, so that

$$u(x) - t\|D(Q_t u)(x)\|_2^2 - \frac{C}{2}t^2\|D(Q_t u)(x)\|_2^2 \le u(y) \le u(x) - t\|D(Q_t u)(x)\|_2^2 - \frac{c}{2}\|D(Q_t u)(x)\|_2^2.$$

Finally, we substitute $Q_t u(x) = u(y) + (t/2)\|D(Q_t u)(x)\|_2^2$ and obtain (3). □

6D. *Estimates for error propagation.* To prepare for our iteration procedure, we first prove some estimates to control the propagation of errors.

Lemma 6.8. *If $u, v \in \mathcal{E}(c, C)$, then we have*:

(1) $\|D(P_t u) - D(P_t v)\|_{L^\infty} \le \|Du - Dv\|_{L^\infty}$.

(2) $\|D(Q_t u) - D(Q_t v)\|_{L^\infty} \le (1 + Ct)\|Du - Dv\|_{L^\infty}$.

Proof. The first inequality follows because $D(P_t u) - D(P_t v)$ is the convolution of $Du - Dv$ with the Gaussian density. To prove the second inequality, note that

$$\|D(Q_t u)(x) - D(Q_t v)(x)\|_2 = \|Du(x - tD(Q_t u)(x)) - Dv(x - tD(Q_t v)(x))\|_2$$
$$\le \|Du(x - tD(Q_t v)(x)) - Dv(x - tD(Q_t v)(x))\|_2$$
$$+ \|Du(x - tD(Q_t u)(x)) - Du(x - tD(Q_t v)(x))\|_2$$
$$\le \|Du - Dv\|_{L^\infty} + Ct\|D(Q_t u)(x) - D(Q_t v)(x)\|_2,$$

where the last inequality follows because Du is C-Lipschitz. This implies that for $t < 1/C$

$$\|D(Q_t u) - D(Q_t v)\|_{L^\infty} \le (1 - Ct)^{-1}\|Du - Dv\|_{L^\infty}.$$

Now we improve the estimate using the semigroup property of Q_t. Fix a positive integer k and for $j = 1, \ldots, k$, let $t_j = tj/k$, and let $C_j = C(1 + Ct_j)^{-1}$. Then $Q_{t_j} u$ and $Q_{t_j} v$ are in $\mathcal{E}(0, C_j)$. Thus, we have

$$\|D(Q_{t_{j+1}} u) - D(Q_{t_{j+1}} v)\|_{L^\infty} \le \left(1 - \frac{C_j t}{k}\right)^{-1}\|D(Q_{t_j} u) - D(Q_{t_j} v)\|_{L^\infty},$$

and hence

$$\|D(Q_t u) - D(Q_t v)\|_{L^\infty} \le \|Du - Dv\|_{L^\infty} \prod_{j=0}^{k-1} \frac{1}{1 - C_j t/k}.$$

Now

$$\log \prod_{j=0}^{k-1} \frac{1}{1 - C_j t/k} = \sum_{j=0}^{k-1} -\log\left(1 - \frac{C_j t}{k}\right) = \sum_{j=0}^{k-1}\left(\frac{C_j t}{k} + O\left(\frac{1}{k^2}\right)\right)$$

$$= \sum_{j=0}^{k-1} \frac{C}{1 + C t_j}(t_{j+1} - t_j) + O\left(\frac{1}{k}\right)$$

$$= \int_0^t \frac{C}{1 + Cs}\, ds + O\left(\frac{1}{k}\right) = \log(1 + Ct) + O\left(\frac{1}{k}\right).$$

Hence,

$$\|D(Q_t u) - D(Q_t v)\|_{L^\infty} \le \left(1 + Ct + O\left(\frac{1}{k}\right)\right)\|Du - Dv\|_{L^\infty},$$

and the proof is completed by taking $k \to \infty$. □

Lemma 6.9. *Suppose that $u : M_N(\mathbb{C})^m_{\mathrm{sa}} \to \mathbb{R}$ is convex and let $v \in \mathcal{E}(c, C)$ and $u \le v + a + b\|Dv\|_2^2$ for some $a \in \mathbb{R}$ and $b \ge 0$.*

(1) $P_t u \le P_t v + a + bC^2 mt + b\|D(P_t v)\|_2^2.$

(2) $Q_t u \le Q_t v + a + b\|D(Q_t v)\|_2^2.$

Proof. (1) Using monotonicity and linearity of P_t, we have

$$P_t u \le P_t v + a + b \int \|Dv(x + y)\|_2^2 \, d\sigma(y).$$

So it suffices to show that

$$\int \|Dv(x + y)\|_2^2 \, d\sigma_{t,N}(y) - \|D(P_t v)(x)\|_2^2 \le C^2 mt.$$

In probabilistic terms, the left-hand side is the variance of the random variable $Dv(x+Y)$, where $Y \sim \sigma_{t,N}$. Since the variance is translation-invariant, this is the same as the variance of $Dv(x + Y) - Dv(x)$, and this is bounded above by the second moment

$$E\|Dv(x + Y) - Dv(x)\|_2^2 \le C^2 \cdot E\|Y\|_2^2 = C^2 mt.$$

(2) Note that

$$Q_t u(x) = \inf_y \left[u(y) + \frac{1}{2t}\|y - x\|_2^2\right]$$

$$\le u(x - t D(Q_t v)(x)) + \frac{t}{2}\|D(Q_t v)(x)\|_2^2$$

$$\le v(x - t D(Q_t v)(x)) + \frac{t}{2}\|D(Q_t v)(x)\|_2^2 + a + b\|Dv(x - t D(Q_t v)(x))\|_2^2$$

$$= Q_t v(x) + a + b\|D(Q_t v)(x)\|_2^2,$$

where the last equality follows from Corollary 6.7(1) and (2). □

Lemma 6.10. *Let $u \in \mathcal{E}(0, C)$. Then:*

(1) $\|D(Q_t u)\|_2^2 \le \|Du\|_2^2.$

(2) $\|D(P_t u)\|_2^2 \le C^2 mt + \|Du\|_2^2.$

Proof. The first claim follows from Lemma 6.6(5). To prove the second claim, note that by Minkowski's inequality

$$\|D(P_t u)(x)\|_2^2 = \left\|\int Du(x+y)\,d\sigma_{t,N}(y)\right\|_2^2$$

$$\leq \int \|Du(x+y)\|_2^2\,d\sigma_{t,N}(y) \leq C^2 mt + \|Du(x)\|_2^2,$$

where the last inequality was shown in the proof of Lemma 6.9(1). □

Next, we iterate the previous inequalities to obtain our main lemma on error propagation.

Lemma 6.11. *Let $t_1, \ldots, t_n > 0$ and write*

$$t^* = t_1 + \cdots + t_n,$$

$$R = P_{t_n} Q_{t_n} \cdots P_{t_1} Q_{t_1}.$$

Let $u, v \in \mathcal{E}(c, C)$.

(1) $Ru, Rv \in \mathcal{E}(c(1+ct^*)^{-1}, C(1+Ct^*)^{-1})$.

(2) $\|D(Ru) - D(Rv)\|_{L^\infty} \leq (1+Ct^*)\|Du - Dv\|_{L^\infty}$.

(3) *If $u \leq v + a + b\|Dv\|_2^2$ with $a \in \mathbb{R}$ and $b \geq 0$, then we have*

$$Ru \leq Rv + a + b\frac{C^2 mt^*}{1+Ct^*} + b\|D(Rv)\|_2^2.$$

   In particular, $u \leq v$ implies $Ru \leq Rv$.

(4) *We have*

$$\|D(Ru)\|_2^2 \leq \frac{C^2 mt^*}{1+Ct^*} + \|Du\|_2^2 \leq Cm + \|Du\|_2^2.$$

Proof. (1) Let $u \in \mathcal{E}(c, C)$. Let $t_j^* = t_1 + \cdots + t_j$ and $u_j = P_{s_j} Q_{t_j} \cdots P_{s_1} Q_{t_1} u$. We show by induction that $u_j \in \mathcal{E}(c(1+ct_j^*)^{-1}, C(1+Ct_j^*)^{-1})$. The base case $j = 0$ is trivial. For the induction step, note that

$$\frac{c(1+ct_j^*)^{-1}}{1+[c(1+ct_j^*)^{-1}]t_{j+1}} = \frac{c}{(1+ct_j^*)+ct_{j+1}} = c(1+ct_{j+1}^*)^{-1}$$

and the same holds with c replaced by C. Hence, by Lemma 6.6(4), if $u_j \in \mathcal{E}(c(1+ct_j^*)^{-1}, C(1+Ct_j^*)^{-1})$, then $Q_{t_{j+1}} u_j \in \mathcal{E}(c(1+ct_{j+1}^*)^{-1}, C(1+Ct_{j+1}^*)^{-1})$. By Lemma 6.3, this implies that $u_{j+1} = P_{t_{j+1}} Q_{t_{j+1}} u_j \in \mathcal{E}(c(1+ct_{j+1}^*)^{-1}, C(1+Ct_{j+1}^*)^{-1})$. The same argument of course applies to v.

(2) Let t_j^* and u_j be as in the proof of (1) and define v_j similarly to u_j. We show by induction that $\|Du_j - Dv_j\|_{L^\infty} \leq (1+Ct_j^*)\|Du - Dv\|_{L^\infty}$. The base case $j = 0$ is trivial. For the induction step, recall that $u_j, v_j \in \mathcal{E}(c(1+ct_j^*)^{-1}, C(1+Ct_j^*)^{-1})$ and hence by Lemma 6.8 and the induction hypothesis

$$\|D(Q_{t_{j+1}} u_j) - D(Q_{t_{j+1}} v_j)\|_{L^\infty} \leq (1+C(1+Ct_j^*)^{-1}t_{j+1})\|Du_j - Dv_j\|_{L^\infty}$$

$$\leq (1+C(1+Ct_j^*)^{-1}t_{j+1})(1+Ct_j^*)\|Du - Dv\|_{L^\infty}$$

$$= (1+Ct_{j+1}^*)\|Du - Dv\|_{L^\infty}.$$

Then by Lemma 6.8 again, since $u_{j+1} = P_{t_{j+1}} Q_{t_{j+1}} u_j$ and $v_{j+1} = P_{t_{j+1}} Q_{t_{j+1}} v_j$, we have

$$\|Du_{j+1} - Dv_{j+1}\|_{L^\infty} \leq (1 + Ct_{j+1}^*) \|Du - Dv\|_{L^\infty}.$$

This completes the induction and the case $j = n$ is the claim (2).

(3) First, we show by induction on j that

$$u_j \leq v_j + a + b \sum_{i=1}^{j} \frac{C^2 m t_i}{(1 + Ct_i^*)^2} + b\|Dv_j\|_2^2.$$

The base case $j = 0$ is trivial. If the claim holds for u_j and v_j, then it also holds for $Q_{t_{j+1}} u_j$ and $Q_{t_{j+1}} v_j$ by Lemma 6.9(2). Then we apply Lemma 6.9(1) together with the fact that $Q_{t_{j+1}} u_j$ and $Q_{t_{j+1}} v_j$ are in $\mathcal{E}(c(1 + ct_{j+1}^*)^{-1}, C(t + Ct_{j+1}^*)^{-1})$ to conclude that

$$u_{j+1} \leq v_{j+1} + a + b \sum_{i=1}^{j+1} \frac{C^2 m t_i}{(1 + Ct_i^*)^2} + b\|Dv_{j+1}\|_2^2.$$

This completes the induction. Finally, we observe that $\sum_{i=1}^{n} C^2 m t_i/(1 + Ct_i^*)^2$ is the lower Riemann sum for the function $C^2 m/(1 + Ct)^2$ on the interval $[0, t^*]$ with respect to the partition $\{0, t_1^*, \ldots, t_n^*\}$. Thus,

$$\sum_{i=1}^{n} \frac{C^2 m t_i}{(1 + Ct_i^*)^2} \leq \int_0^{t^*} \frac{C^2 m}{(1 + Ct)^2}\, dt = Cm\left(1 - \frac{1}{1 + Ct^*}\right) = \frac{C^2 m t^*}{1 + Ct^*}.$$

This shows the main claim of (3), and the claim that $u \leq v$ implies $Ru \leq Rv$ is the special case when $a = 0$ and $b = 0$.

(4) By Lemma 6.10, we have $\|D(Q_{t_{j+1}} u_j)\|_2^2 \leq \|Du_j\|_2^2$ and

$$\|Du_{j+1}\|_2^2 \leq \frac{C^2 m t_{j+1}}{1 + Ct_{j+1}^*} + \|D(Q_{t_{j+1}} u_j)\|_2^2 \leq \frac{C^2 m t_{j+1}}{1 + Ct_{j+1}^*} + \|Du_j\|_2^2.$$

We sum from $j = 0, \ldots, n - 1$ and obtain the same lower Riemann sum as in the proof of (3). The final estimate $Cm + \|Du\|_2^2$ follows because $C^2 m t/(1 + Ct) \leq Cm$. $\qquad\square$

6E. *Iterative construction of R_t for dyadic t.* We are now ready to carry out the Trotter's formula strategy and construct the semigroup for dyadic values of t. The next step is to show that the operators P_t and Q_t almost commute when t is small.

Lemma 6.12. *Let $u \in \mathcal{E}(c, C)$ and $t > 0$.*

(1) $\|D(Q_t P_t u) - D(P_t Q_t u)\|_{L^\infty} \leq C^2 m^{1/2}(2 + Ct)t^{3/2}.$

(2) $P_t Q_t u \leq Q_t P_t u.$

(3) *If $Ct \leq 1$, then $Q_t P_t u \leq P_t Q_t u + 2C^2 m t^2 + 2Ct^2\|D(P_t Q_t u)\|_2^2.$*

Proof. (1) Applying Corollary 6.7(1) to $P_t u$ yields

$$D(Q_t P_t u)(x) = D(P_t u)(x - tD(Q_t P_t u)(x)) = \int Du(x + y - tD(Q_t P_t u)(x))\, d\sigma_{t,n}(y).$$

On the other hand,

$$D(P_t Q_t u)(x) = \int D(Q_t u)(x + y)\, d\sigma_{t,n}(y) = \int Du(x + y - tD(Q_t u)(x + y))\, d\sigma_{t,n}(y).$$

Because Du is C-Lipschitz, we have

$$\|D(Q_t P_t u)(x) - D(P_t Q_t u)(x)\|_2 \leq Ct \int \|D(Q_t u)(x + y) - D(Q_t P_t u)(x)\|_2\, d\sigma_{t,n}(y).$$

We can estimate the integrand by

$$\|D(Q_t u)(x + y) - D(Q_t u)(x)\|_2 + \|D(Q_t u)(x) - D(Q_t P_t u)(x)\|_2.$$

Integrating the first term and using the fact that $D(Q_t u)$ is C-Lipschitz (since $Q_t u \in \mathcal{E}(0, C)$ by Lemma 6.6(4)), we have

$$\int \|D(Q_t u)(x + y) - D(Q_t u)(x)\|_2\, d\sigma_{t,n}(y) \leq C \int \|y\|_2\, d\sigma_{t,n} \leq Cm^{1/2} t^{1/2}.$$

Meanwhile, the second term is independent of y and thus it is unchanged when we integrate it against the probability measure $\sigma_{t,N}$, and this quantity can be estimated using Lemmas 6.8(2) and 6.3(2) as

$$\|D(Q_t u)(x) - D(Q_t P_t u)(x)\|_2 \leq (1 + Ct)\|Du - D(P_t u)\|_{L^\infty} \leq (1 + Ct)Cm^{1/2} t^{1/2}.$$

Altogether, we obtain

$$\|D(Q_t P_t u)(x) - D(P_t Q_t u)(x)\|_2 \leq C^2 m^{1/2}(2 + Ct)t^{3/2}.$$

(2) The idea is that the average of the infimum is less than or equal to the infimum of the average. More precisely,

$$\begin{aligned}
P_t Q_t u(x) &= \int \inf_y \Big(u(y) + \frac{1}{2t}\|(x + z) - y\|_2^2\Big)\, d\sigma_{t,N}(z) \\
&= \int \inf_y \Big(u(y - z) + \frac{1}{2t}\|x - y\|_2^2\Big)\, d\sigma_{t,N}(z) \\
&\leq \inf_y \int \Big(u(y - z) + \frac{1}{2t}\|x - y\|_2^2\Big)\, d\sigma_{t,N}(z) \\
&= \inf_y \Big(P_t u(y) + \frac{1}{2t}\|x - y\|_2^2\Big) = Q_t P_t u(x).
\end{aligned}$$

(3) By Corollary 6.7(3),

$$Q_t P_t u \leq P_t u - \frac{t}{2}\|D(Q_t P_t u)\|_2^2. \tag{6-6}$$

Also by Corollary 6.7(3),

$$u \leq Q_t u + \frac{t}{2}(1 + Ct)\|D(Q_t u)\|_2^2.$$

Hence, by Lemma 6.9, since $Q_t u \in \mathcal{E}(c(1 + ct)^{-1}, C(1 + Ct)^{-1}) \subseteq \mathcal{E}(0, C)$, we have

$$P_t u \leq P_t Q_t u + \frac{C^2 m t^2}{2}(1 + Ct) + \frac{t}{2}(1 + Ct)\|D(P_t Q_t)\|_2^2. \tag{6-7}$$

Plugging (6-7) into (6-6), we obtain

$$Q_t P_t u \le P_t Q_t u + \frac{C^2 m t^2}{2}(1 + Ct) - \frac{t}{2}\|D(Q_t P_t u)\|_2^2 + \frac{t}{2}(1 + Ct)\|D(P_t Q_t)\|_2^2. \tag{6-8}$$

By using part (1), we have

$$
\begin{aligned}
\|D(Q_t P_t u)\|_2^2 &\ge [\|D(P_t Q_t)u\|_2 - C^2 m^{1/2} t^{3/2}(2 + Ct)]^2 \\
&\ge \|D(P_t Q_t u)\|_2^2 - 2C^2 m^{1/2} t^{3/2}(2 + Ct)\|D(P_t Q_t u)\|_2 \\
&\ge \|D(P_t Q_t u)\|_2^2 - (2 + Ct)[C^3 m t^2 + Ct\|D(P_t Q_t u)\|_2^2],
\end{aligned}
$$

where the last step follows from the arithmetic-geometric mean inequality

$$2Cm^{1/2}t^{1/2}\|D(P_t Q_t u)\|_2 \le C^2 m t + \|D(P_t Q_t u)\|_2^2.$$

So substituting our estimate for $\|D(Q_t P_t u)\|_2^2$ into (6-8), we see that $P_t Q_t u - Q_t P_t u$ is bounded by

$$\frac{C^2 m t^2}{2} + \frac{t}{2}(2 + Ct)[C^3 m t^2 + Ct\|D(P_t Q_t u)\|_2^2] - \frac{t}{2}\|D(P_t Q_t u)\|_2^2 + \frac{t}{2}(1 + Ct)\|D(P_t Q_t)\|_2^2.$$

Now we cancel the first-order terms $(t/2)\|D(P_t Q_t u)\|_2^2$ and we estimate $2 + Ct$ by 3 using our assumption that $Ct \le 1$. Thus, this is bounded by

$$\frac{C^2 m t^2}{2} + \frac{3t}{2}[C^3 m t^2 + Ct\|D(P_t Q_t u)\|_2^2] + \frac{Ct^2}{2}\|D(P_t Q_t u)\|_2^2 \le 2C^2 m t^2 + 2Ct^2\|D(P_t Q_t u)\|_2^2,$$

where we have again used our assumption $Ct \le 1$ to cancel a factor of Ct from the term $t \cdot C^3 m t^2$. $\quad\square$

Finally, we can construct the semigroup R_t for dyadic values of t. As in the statement of Theorem 6.1, we define $R_{t,\ell}u = (P_{2^{-\ell}} Q_{2^{-\ell}})^{2^\ell t}u$ whenever $\ell \in \mathbb{Z}$ and $t \in 2^{-\ell}\mathbb{N}_0$.

Lemma 6.13. *Let $C \ge 0$. For $t \in \mathbb{Q}_2^+$ and $u \in \mathcal{E}(0, C)$, the limit $R_t u = \lim_{\ell \to \infty} R_{t,\ell}u$ exists. Moreover, we have for $t \in 2^{-\ell}\mathbb{N}_0$ that*:

(1) $R_{t,\ell}u \le R_t u$.

(2) *If $C/2^{\ell+1} \le 1$, then*

$$R_t u \le R_{t,\ell}u + \left(\frac{3}{2}\frac{C^2 m t}{1 + Ct} + \log(1 + Ct)(m + Cm + \|Du\|_2^2)\right)2^{-\ell}.$$

(3) $\|D(R_{t,\ell}u) - D(R_t u)\|_{L^\infty} \le [t/2 + C(t/2)^2]C^2 m^{1/2}(2 \cdot 2^{-\ell/2} + 2^{-3\ell/2}C)$.

Proof. First, we prove some intermediate claims relating $R_{t,\ell}u$ and $R_{t,\ell+1}u$. To this end, we fix $\ell \in \mathbb{Z}$ and suppose $t = 2^{-\ell}n$ for some $n \in \mathbb{N}_0$. Let $\delta = 2^{-\ell-1}$. For $j = 0, \ldots, n$, define

$$u_j = (P_\delta Q_\delta)^{2(n-j)}(P_{2\delta} Q_{2\delta})^j u.$$

and note that

$$u_0 = R_{t,\ell+1}u, \quad u_n = R_{t,\ell}u.$$

Let

$$v_j = Q_\delta(P_{2\delta} Q_{2\delta})^j u.$$

Then for $j = 1, \ldots, n$, we have

$$u_{j-1} = [(P_\delta Q_\delta)^{2(n-j)} P_\delta](Q_\delta P_\delta v_{j-1}),$$
$$u_j = [(P_\delta Q_\delta)^{2(n-j)} P_\delta](P_\delta Q_\delta v_{j-1}).$$

We also define for $k = 1, \ldots, 2n$,

$$C_k = C(1 + Ck\delta)^{-1}, \quad c_k = c(1 + ck\delta)^{-1}.$$

Thus, by Lemmas 6.11(1) and 6.6(4), we have $v_{j-1} \in \mathcal{E}(c_{2j-1}, C_{2j-1})$.

First, we claim that

$$R_{t,\ell} u \leq R_{t,\ell+1} u. \tag{6-9}$$

Now by Lemma 6.12(2), we have

$$P_\delta Q_\delta v_{j-1} \leq Q_\delta P_\delta v_{j-1}.$$

Hence, by monotonicity of P_t and Q_t (Lemma 6.11(3)), we have $u_j \leq u_{j-1}$. Hence, $R_{t,\ell} u = u_n \leq u_0 = R_{t,\ell+1} u$, proving (6-9).

For an inequality in the other direction, we claim that

$$R_{t,\ell+1} u \leq R_{t,\ell} u + \left(\frac{3}{2} \frac{C^2 mt}{1 + Ct} + \log(1 + Ct)(m + Cm + \|Du\|_2^2) \right) 2^{-\ell-1}. \tag{6-10}$$

By Lemma 6.12(3), since $v_{j-1} \in \mathcal{E}(c_{2j-1}, C_{2j-1})$, we obtain

$$Q_\delta P_\delta v_{j-1} \leq P_\delta Q_\delta v_{j-1} + 2C_{2j-1}^2 m\delta^2 + 2C_{2j-1}\delta^2 \|D(P_\delta Q_\delta v_{j-1})\|_2^2.$$

Thus, by Lemma 6.9(1), since $Q_\delta P_\delta v_{j-1}$ and $P_\delta Q_\delta v_{j-1}$ are in $\mathcal{E}(c_{2j}, C_{2j})$, we have

$$P_\delta Q_\delta P_\delta v_{j-1} \leq P_{2\delta} Q_\delta v_{j-1} + 2C_{2j-1} m\delta^2 + 2C_{2j-1}\delta^2(C_{2j}^2 m\delta + \|D(P_{2\delta} Q_\delta v_{j-1})\|_2^2).$$

Recalling that u_{j-1} and u_j are obtained by applying $(P_\delta Q_\delta)^{2(n-j)}$ to $P_\delta Q_\delta P_\delta v_{j-1}$ and $P_{2\delta} Q_\delta v_{j-1}$, and that $P_\delta Q_\delta P_\delta v_{j-1}$ and $P_{2\delta} Q_\delta v_{j-1}$ are in $\mathcal{E}(c_{2j}, C_{2j})$, we may apply Lemma 6.11(3) and to conclude that

$$u_{j-1} \leq u_j + 2C_{2j-1} m\delta^2 + 2C_{2j-1}\delta^2\left(C_{2j}^2 m\delta + \frac{C_{2j}^2 m(n-j)\delta}{1 + 2C_{2j}(n-j)\delta} + \|Du_j\|_2^2\right).$$

By our assumption, $C_{2j}\delta \leq C\delta \leq 1$, and thus

$$C_{2j}^2 m\delta + \frac{C_{2j}^2 m(n-j)\delta}{1 + 2C_{2j}(n-j)\delta} \leq C_{2j} m + \frac{C_{2j} m}{2} = \frac{3C_{2j} m}{2} \leq \frac{3C_{2j-1} m}{2}.$$

Therefore,

$$u_{j-1} - u_j \leq 2C_{2j-1} m\delta^2 + 3C_{2j-1}^2 m\delta^2 + 2C_{2j-1}\delta^2 \|Du_j\|_2^2.$$

By Lemma 6.11(4), we have $\|Du_j\|_2 \leq Cm + \|Du\|_2^2$, and hence

$$u_{j-1} - u_j \leq 3C_{2j-1}^2 m\delta^2 + 2C_{2j-1}\delta^2(m + Cm + \|Du\|_2^2).$$

Therefore, summing from $j = 1, \ldots, n$, we have

$$R_{t,\ell+1}u - R_{t,\ell}u \leq 3m\delta^2 \sum_{j=1}^{n} C_{2j-1}^2 + 2\delta^2(m + Cm + \|Du\|_2^2)\Bigg(\sum_{j=1}^{n} C_{2j-1}\Bigg)$$

$$= \frac{3m\delta}{2}\Bigg(\sum_{j=1}^{n} C_{2j-1}^2(2\delta)\Bigg) + \delta(m + Cm + \|Du\|_2^2)\Bigg(\sum_{j=1}^{n} C_{2j-1}(2\delta)\Bigg).$$

Recalling the definition of C_{2j-1}, two times the first sum is $\sum_{j=1}^{n} C^2(2\delta)/(1 + C(2j-1)\delta)^2$, which is the Riemann sum for the function $\phi(s) = C^2/(1 + Cs)^2$ on the interval $[0, t] = [0, 2n\delta]$, where we use a partition into subintervals of length 2δ and evaluate ϕ at the midpoint of each interval. Because ϕ is convex, the value of ϕ at the midpoint is less than or equal to the average value over the subinterval and therefore

$$\sum_{j=1}^{n} \frac{C^2(2\delta)}{(1 + C(2j-1)\delta)^2} \leq \int_0^t \frac{C^2}{(1 + Cs)^2}\, ds = \frac{C^2 t}{1 + Ct}.$$

By similar reasoning,

$$\sum_{j=1}^{n} C_{2j-1}(2\delta) = \sum_{j=1}^{n} \frac{C\delta}{(1 + C(2j-1)\delta)} \leq \int_0^t \frac{C}{1 + Cs}\, ds = \log(1 + Ct).$$

Therefore,

$$R_{t,\ell+1}u - R_{t,\ell}u \leq \Bigg(\frac{3}{2}\frac{C^2 mt}{1 + Ct} + \log(1 + Ct)(m + Cm + \|Du\|_2^2)\Bigg)\delta,$$

which proves (6-10).

Together, (6-9) and (6-10) show that

$$|R_{t,\ell+1}u - R_{t,\ell}u| \leq \Bigg(\frac{3}{2}\frac{C^2 mt}{1 + Ct} + \log(1 + Ct)(m + Cm + \|Du\|_2^2)\Bigg)2^{-\ell-1}.$$

Because the right-hand side is summable in ℓ, we see that the sequence $\{R_{t,\ell}u(x)\}_{\ell \in \mathbb{N}}$ is Cauchy and hence converges. Thus, $\lim_{\ell \to \infty} R_{t,\ell}u$ exists. Also, by (6-9) the convergence is monotone and thus $R_{t,\ell}u \leq R_t u$, establishing (1). On the other hand, we obtain (2) by summing up the estimate (6-10) from ℓ to ∞ using the geometric series formula.

It remains to prove (3). We first claim that

$$\|D(R_{t,\ell+1}u) - D(R_{t,\ell}u)\|_{L^\infty} \leq \Bigg[\frac{t}{2} + C\Big(\frac{t}{2}\Big)^2\Bigg]C^2 m^{1/2}(2 + 2^{-(\ell+1)}C)2^{-(\ell+1)/2}. \tag{6-11}$$

By Lemma 6.11(1), we know $Q_\delta P_\delta v_{j-1}$ and $P_\delta Q_\delta v_{j-1}$ are in $\mathcal{E}(c(1 + 2cj\delta)^{-1}, C(1 + 2Cj\delta)^{-1})$, and hence in $\mathcal{E}(0, C)$. Therefore, by Lemmas 6.11(2) and 6.12(1), we have

$$\|Du_j - Du_{j-1}\|_{L^\infty} \leq [1 + 2C(n-j)\delta]\|D(Q_\delta P_\delta v_j) - D(P_\delta Q_\delta v_j)\|_{L^\infty}$$

$$\leq [1 + 2C(n-j)\delta]C^2 m^{1/2}(2 + C\delta)\delta^{3/2}.$$

Therefore,

$$\|D(R_{t,\ell+1}u) - D(R_{t,\ell}u)\|_{L^\infty} \leq \sum_{j=1}^{n}\|Du_j - Du_{j-1}\|_{L^\infty} \leq \sum_{j=1}^{n}[1 + 2C(n-j)\delta]C^2m^{1/2}(2+C\delta)\delta^{3/2}$$

$$= [n + Cn(n-1)\delta]C^2m^{1/2}(2+C\delta)\delta^{3/2}$$

$$\leq \left[\frac{t}{2} + C\left(\frac{t}{2}\right)^2\right]C^2m^{1/2}(2+C\delta)\delta^{1/2}$$

$$= \left[\frac{t}{2} + C\left(\frac{t}{2}\right)^2\right]C^2m^{1/2}(2+2^{-(\ell+1)}C)2^{-(\ell+1)/2}$$

since $2n\delta = t$. This proves (6-11).

Because $[t/2 + C(t/2)^2]C^2m^{1/2}(2+2^{-(\ell+1)}C)2^{-(\ell+1)/2}$ is summable with respect to ℓ, we see that $\{D(R_{t,\ell}u)\}_{\ell\in\mathbb{N}}$ is Cauchy with respect to the L^∞ norm (even though the individual functions may not be in L^∞) and hence converges uniformly to some function. We already know that $R_{t,\ell}u$ converges to $R_t u$, so the limit of $D(R_{t,\ell}u)$ must be $D(R_t u)$. We obtain the estimate (3) by summing (6-11) from ℓ to ∞ using the geometric series formula. $\square$

Corollary 6.14. *Let $0 \leq c \leq C$. Let $u, v \in \mathcal{E}(c, C)$ and let $t \geq 0$ be a dyadic rational.*

(1) $R_t u, R_t v \in \mathcal{E}(c(1+ct)^{-1}, C(1+Ct)^{-1})$.

(2) $\|D(R_t u) - D(R_t v)\|_{L^\infty} \leq (1+Ct)\|Du - Dv\|_{L^\infty}$.

(3) *If $u \leq v + a + b\|Dv\|_2^2$ for some $a \in \mathbb{R}$ and $b \geq 0$, then*

$$R_t u \leq R_t v + a + b\frac{C^2mt}{1+Ct} + b\|D(R_t v)\|_2^2.$$

(4) $\|D(R_t u)\|_2^2 \leq (C^2mt)/(1+Ct) + \|Du\|_2^2$.

Proof. We know that these properties hold for $R_{t,\ell}$ by Lemma 6.11. By Lemma 6.13, they also hold in the limit taking $\ell \to \infty$. (For (1), we rely on Proposition 2.13(1).) $\square$

6F. *Continuity and semigroup property.* In order to extend R_t to all real $t \geq 0$, we prove estimates that show that R_t depends continuously on t. We begin with some simple estimates for P_t and Q_t.

Lemma 6.15. *Let $\ell \in \mathbb{Z}$ and suppose that $t \in 2^{-\ell}\mathbb{N}_0$ and $u \in \mathcal{E}(0, C)$. Then:*

(1) $u \leq P_t u \leq u + (C/2)mt$.

(2) $u - (t/2)\|Du\|_2^2 \leq Q_t u \leq u$.

(3) $\|D(Q_t u) - Du\|_2 \leq Ct\|Du\|_2$.

Proof. (1) Because u is convex and $u(x) - (C/2)\|x\|_2^2$ is concave, we have

$$u(x) + \langle Du(x), y\rangle \leq u(x+y) \leq u(x) + \langle Du(x), y\rangle + \frac{C}{2}\|y\|_2^2.$$

Integrating with respect to $d\sigma_{t,N}(y)$ yields

$$u(x) \leq P_t u(x) \leq u(x) + \frac{Cmt}{2} \quad \text{for } u \in \mathcal{E}(0, C).$$

(2) As for the operator Q_t, it is immediate from the definition that $Q_t u \leq u$. On the other hand, using Corollary 6.7(2) and the convexity of u,

$$
\begin{aligned}
Q_t u(x) &= u(x - tD(Q_t u)(x)) + \frac{t}{2}\|D(Q_t u)(x)\|_2^2 \\
&\geq u(x) - t\langle D(Q_t u)(x), Du(x)\rangle_2 + \frac{t}{2}\|D(Q_t u)(x)\|_2^2 \\
&\geq u(x) - \frac{t}{2}\|Du(x)\|_2^2,
\end{aligned}
$$

where the last inequality follows because $\langle D(Q_t u)(x), Du(x)\rangle_2 \leq \frac{1}{2}\|D(Q_t u)(x)\|_2^2 + \frac{1}{2}\|Du(x)\|_2^2$.

(3) Using the fact that Du is C-Lipschitz, together with Corollary 6.7(1) and Lemma 6.6(5),

$$
\begin{aligned}
\|D(Q_t u)(x) - Du(x)\|_2 &= \|Du(x - tD(Q_t u)(x)) - Du(x)\|_2 \\
&\leq Ct\|D(Q_t u)(x)\|_2 \\
&\leq Ct\|Du(x)\|_2. \qquad \square
\end{aligned}
$$

Lemma 6.16. *Let $s \leq t$ be two numbers in $\mathbb{Q}_2^+$, and let $u \in \mathcal{E}(0, C)$.*

(1) $R_t u \leq R_s u + (m/2)[\log(1 + Ct) - \log(1 + Cs)]$.

(2) $R_t u \geq R_s u - ((t - s)/2)(Cm + \|Du\|_2^2)$.

(3) *If $C(t - s) \leq 1$, then $\|D(R_t u) - D(R_s u)\|_2 \leq 5Cm^{1/2}2^{1/2}(t - s)^{1/2} + C(t - s)\|Du\|_2$.*

Moreover, if $\ell \in \mathbb{Z}$ and if $s, t \in 2^{-\ell}\mathbb{N}_0$, then the same estimates hold with R_t replaced by $R_{t,\ell}$.

Proof. (1) Fix $\ell \in \mathbb{Z}$ and let $\delta = 2^{-\ell}$. Suppose $s = n\delta$ and $t = n'\delta$, where $n, n' \in \mathbb{N}_0$. By the previous lemma,

$$
\begin{aligned}
R_{(j+1)\delta,\ell} u &= P_\delta Q_\delta R_{j\delta,\ell} u \\
&\leq Q_\delta R_{j\delta,\ell} u + \frac{Cm\delta}{2(1 + C(j+1)\delta)} \\
&\leq R_{j\delta,\ell} u + \frac{Cm\delta}{2(1 + C(j+1)\delta)},
\end{aligned}
$$

where we have used the fact that $Q_\delta R_{j\delta,\ell} u \in \mathcal{E}(0, C(1 + C(j+1)\delta)^{-1})$. Therefore,

$$
R_{n'\delta,\ell} u \leq R_{n\delta,u} + \sum_{j=n}^{n'-1} \frac{Cm\delta}{2(1 + C(j+1)\delta)}.
$$

Since the sum on the right-hand side is a lower Riemann sum for the function $Cm\delta/(2(1 + C\tau))$ for $\tau \in [s, t]$, we obtain

$$
R_{t,\ell} u \leq R_{s,\ell} u + \frac{m}{2}[\log(1 + Ct) - \log(1 + Cs)].
$$

We obtain (1) by letting $\ell \to +\infty$ and using Lemma 6.13.

(2) Let $\ell, \delta, s, t, n, n'$ be as above. By the previous lemma,

$$R_{(j+1)\delta,\ell}u = P_\delta Q_\delta R_{j\delta,\ell}u \geq Q_\delta R_{j\delta,\ell}u$$

$$\geq R_{j\delta,\ell}u - \frac{\delta}{2}\|D(R_{j\delta,\ell}u)\|_2^2$$

$$\geq R_{j\delta,\ell}u - \frac{\delta}{2}(Cm + \|Du\|_2^2),$$

where the last inequality follows from Lemma 6.11(4). So when we sum from $j = n$ to $n' - 1$, we obtain

$$R_t u \geq R_s u - \frac{t-s}{2}(Cm + \|Du\|_2^2).$$

Then (2) follows by taking $\ell \to +\infty$.

(3) Assume that $s, t \in 2^{-\ell}\mathbb{N}_0$. Choose $k \in \mathbb{Z}$ such that $2^{-k-1} \leq t - s \leq 2^{-k}$. Then we may write $t - s$ in a binary expansion

$$t - s = \sum_{j=k+1}^{\ell} a_j 2^{-j},$$

where $a_j \in \{0, 1\}$ for each j and $a_{k+1} = 1$. Let

$$t_j = s + a_{k+1}2^{-k-1} + \cdots + a_j 2^{-j}.$$

Let $u_j = R_{t_j,\ell}u$. We will estimate $\|Du_j(x) - Du_{j-1}(x)\|_2$ for each j. Of course, if $a_j = 0$, then $u_j = u_{j-1}$, so there is nothing to prove. On the other hand, suppose that $a_j = 1$. Now we estimate (at our given point x, suppressed in the notation)

$$\|D(R_{2^{-j},\ell}u_{j-1}) - Du_{j-1}\|_2 \leq \|D(R_{2^{-j},\ell}u_{j-1}) - D(P_{2^{-j}}Q_{2^{-j}}u_{j-1})\|_2$$

$$+ \|D(P_{2^{-j}}Q_{2^{-j}}u_{j-1}) - D(Q_{2^{-j}}u_{j-1})\|_2$$

$$+ \|D(Q_{2^{-j}}u_{j-1}) - Du_{j-1}\|_2. \tag{6-12}$$

The first term on the right-hand side may be estimated as follows. Recall that we proved Lemma 6.13(3) from the estimate (6-11) by summing the geometric series. The same reasoning shows that if $\ell \geq j$ and $\delta \in 2^{-\ell}\mathbb{N}_0$, then

$$\|D(R_{\delta,\ell}u_{j-1}) - D(R_{\delta,j}u_{j-1})\|_{L^\infty} \leq \left[\frac{\delta}{2} + C\left(\frac{\delta}{2}\right)^2\right]C^2 m^{1/2}(2 \cdot 2^{-j/2} + 2^{-3j/2}C)$$

since $u_{j-1} \in \mathcal{E}(0, C)$. If we substitute $\delta = 2^{-j}$, then $R_{2^{-j},j}$ is simply equal to $P_{2^{-j}}Q_{2^{-j}}$. Thus, at the point x,

$$\|D(R_{2^{-j},\ell}u_{j-1}) - D(P_{2^{-j}}Q_{2^{-j}}u_{j-1})\|_2 \leq C^2 m^{1/2}\left[\frac{2^{-j}}{2} + \frac{C2^{-2j}}{4}\right][2 \cdot 2^{-j/2} + 2^{-3j/2}C].$$

By our assumption $C2^{-j} \leq C(t - s) \leq 1$ and hence we may replace $C2^{-2j}/4$ by $2^{-j}/2$ and replace $2^{-3j/2}C$ by $2^{-j/2}$ and hence

$$\|D(R_{2^{-j},\ell}u_{j-1}) - D(P_{2^{-j}}Q_{2^{-j}}u_{j-1})\|_2 \leq 3C^2 m^{1/2}2^{-3j/2} \leq 3Cm^{1/2}2^{-j/2}.$$

The second term on the right-hand side of (6-12) can be estimated by Lemma 6.3(2) by

$$\|D(P_{2^{-j}}Q_{2^{-j}}u_{j-1}) - D(Q_{2^{-j}}u_{j-1})\|_2 \le Cm^{1/2}2^{-j/2}$$

since $Q_{2^{-j}}u_{j-1} \in \mathcal{E}(0, C)$. The third term on the right-hand side of (6-12) can be estimated using Lemma 6.15(3) by

$$\|D(Q_{2^{-j}}u_{j-1}) - Du_{j-1}\|_2 \le C2^{-j}\|Du_{j-1}\|_2.$$

Meanwhile, by Lemma 6.11(4) and the triangle inequality

$$\|Du_{j-1}\|_2 \le \sqrt{Cm + \|Du\|_2^2} \le C^{1/2}m^{1/2} + \|Du\|_2.$$

So using the fact $C2^{-j} \le 1$, we have

$$\|D(Q_{2^{-j}}u_{j-1}) - Du_{j-1}\|_2 \le C^{3/2}m^{1/2}2^{-j} + C2^{-j}\|Du\|_2 \le Cm^{1/2}2^{-j/2} + C(t_j - t_{j-1})\|Du\|_2.$$

Therefore, plugging all our estimates into (6-12), we get

$$\|Du_j - Du_{j+1}\|_2 \le 5Cm^{1/2}2^{-j/2} + C(t_j - t_{j-1})\|Du\|_2.$$

Then summing from $j = k + 1$ to ℓ we obtain

$$\|Du_\ell - Du_k\|_2 \le 5Cm^{1/2}2^{-k/2} + C(t - s)\|Du\|_2$$
$$\le 5Cm^{1/2}2^{1/2}(t - s)^{1/2} + C(t - s)\|Du\|_2.$$

Because $u_\ell = R_{t,\ell}u$ and $u_k = R_{s,\ell}u$, we have shown that (3) holds for $R_{s,\ell}$ and $R_{t,\ell}$ instead of R_s and R_t. Thus, (3) follows by taking $\ell \to +\infty$. $\square$

Proof of Theorem 6.1. Lemma 6.16 shows that if $t \ge 0$ and if t_ℓ is a sequence of dyadic rationals converging to t as $\ell \to \infty$, then $R_{t_\ell}u$ converges to some function v and this function is independent of the approximating sequence, so we define $R_t u = v$. Claims (1), (3), and (4) of the theorem were proved for dyadic t in Corollary 6.14(1), Lemma 6.16, and Corollary 6.14(2)–(4) respectively, and each of these claims can be extended to real $t \ge 0$ in light of the continuity estimates Lemma 6.16. Claim (2) of the theorem is Lemma 6.13.

Thus, it remains to show that R_t is a semigroup. That is, we must show that $R_s R_t u = R_{s+t}u$ for $u \in \mathcal{E}(0, C)$ (and we have not even checked this for dyadic s, t yet). First, we check this property for real $s, t \ge 0$ under the additional restriction that $Ct \le \frac{1}{2}$. For each $\ell \in \mathbb{Z}$, there exist s_ℓ and $t_\ell \in 2^{-\ell}\mathbb{N}_0$ such that $s - 2^{-\ell} < s_\ell \le s$ and $t - 2^{-\ell} < t_\ell \le t$. By Lemma 6.16(1) and (2) we have

$$|R_{t_\ell}u - R_t u| \le \frac{|t_\ell - t|}{2}(Cm + \|Du\|_2^2) \le 2^{-\ell}\frac{1}{2}(Cm + \|Du\|_2^2),$$

since $|\log(1 + Ct_\ell) - \log(1 + Ct)| \le C|t_\ell - t|$ (from computation of the derivative of $\log(1 + Ct)$). By Lemma 6.13(1) and (2), if $C2^{-\ell-1} \le 1$, then

$$|R_{t_\ell,\ell}u - R_{t_\ell}u| \le 2^{-\ell}\left(\frac{3}{2}\frac{C^2mt}{1 + Ct} + \log(1 + Ct_\ell)(m + Cm + \|Du\|_2^2)\right).$$

Since $t_\ell \leq t$, we can replace t_ℓ by t on the right-hand side. By the triangle inequality, we obtain

$$|R_{t_\ell,\ell}u - R_t u| \leq 2^{-\ell} K_t(1 + \|Du\|_2^2) \tag{6-13}$$

for some constant K_t depending on t (and C). Using Lemma 6.16(3), or rather its extension to real values of t,

$$\|D(R_t u) - Du\|_2 \leq 5Cm^{1/2}2^{1/2}t^{1/2} + Ct\|Du\|_2$$
$$\leq 5Cm^{1/2}2^{1/2}t^{1/2} + Ct\|D(R_t u) - Du\|_2 + Ct\|D(R_t u)\|_2.$$

Hence,

$$\|D(R_t u) - Du\|_2 \leq (1 - Ct)^{-1}[5Cm^{1/2}2^{1/2}t^{1/2} + Ct\|D(R_t u)\|_2],$$

so by the triangle inequality,

$$\|Du\|_2 \leq \|D(R_t u)\|_2 + (1 - Ct)^{-1}[5Cm^{1/2}2^{1/2}t^{1/2} + Ct\|D(R_t u)\|_2.$$

By squaring and applying the arithmetic-geometric mean inequality, we get

$$\|Du\|_2^2 \leq A_t + B_t\|D(R_t u)\|_2^2$$

for some constants A_t and B_t depending on t. The same reasoning applies to $R_{t_\ell,\ell}$ since Lemma 6.16(3) holds for $R_{t_\ell,\ell}$ also. We thus obtain

$$\|Du\|_2 \leq \|D(R_{t_\ell,\ell}u)\|_2 + (1 - Ct_\ell)^{-1}[5Cm^{1/2}2^{1/2}t_\ell^{1/2} + Ct_\ell\|D(R_{t_\ell,\ell}u)\|_2$$
$$\leq \|D(R_{t_\ell,\ell}u)\|_2 + (1 - Ct)^{-1}[5Cm^{1/2}2^{1/2}t^{1/2} + Ct\|D(R_{t_\ell,\ell}u)\|_2$$

and so

$$\|Du\|_2^2 \leq A_t + B_t\|D(R_{t_\ell,\ell}u)\|_2^2.$$

Overall,

$$R_t u \leq R_{t_\ell,\ell}u + 2^{-\ell} K_t(1 + A_t + B_t\|D(R_{t_\ell,\ell}u)\|_2^2),$$
$$R_{t_\ell,\ell}u \leq R_t u + 2^{-\ell} K_t(1 + A_t + B_t\|D(R_t u)\|_2^2).$$

So by Lemma 6.11(3) and (4)

$$R_{s_\ell,\ell}R_t u \leq R_{s_\ell,\ell}R_{t_\ell,\ell}u + 2^{-\ell} K_t(1 + A_t + B_t\|D(R_{s_\ell,\ell}R_{t_\ell,\ell}u)\|_2^2)$$
$$\leq R_{s_\ell,\ell}R_{t_\ell,\ell}u + 2^{-\ell} K_t(1 + A_t + CmB_t + B_t\|Du\|_2^2),$$

and the same holds with R_t and $R_{t_\ell,\ell}$ switched, so that

$$|R_{s_\ell,\ell}R_t u - R_{s_\ell+t_\ell,\ell}u| \leq 2^{-\ell} K_t(1 + A_t + CmB_t + B_t\|Du\|_2^2), \tag{6-14}$$

where we have used that $R_{s_\ell+t_\ell,\ell}u = R_{s_\ell,\ell}R_{t_\ell,\ell}u$.

By the same token as (6-13), since $R_t u \in \mathcal{E}(0, C)$, we have

$$|R_{s_\ell,\ell}R_t u - R_s R_t u| \leq 2^{-\ell} K_s(1 + \|D(R_t u)\|_2^2). \tag{6-15}$$

Similarly, since $(s + t) - (s_\ell + t_\ell) \leq 2 \cdot 2^{-\ell}$, we have

$$|R_{s_\ell+t_\ell,\ell}u - R_{s+t}u| \leq 2^{-\ell} \cdot 2K_{s+t}(1 + \|Du\|_2^2). \tag{6-16}$$

Combining these with (6-14) using the triangle inequality, we get

$$|R_s R_t u - R_{s+t} u| \leq 2^{-\ell} K_t (1 + A_t + Cm B_t + B_t \|Du\|_2^2)$$
$$+ 2^{-\ell} K_s (1 + \|D(R_t u)\|_2^2) + 2^{-\ell} \cdot 2 K_{s+t} (1 + \|Du\|_2^2).$$

Taking $\ell \to \infty$, we get $R_s R_t u = R_{s+t} u$ as desired. This completes the case when $Ct \leq \frac{1}{2}$.

In the general case, suppose $s, t \geq 0$ and $u \in \mathcal{E}(0, C)$. Choose n large enough that $Ct/n \leq \frac{1}{2}$. Then for $j = 1, \ldots, n-1$, we have $R_{t/n}^{n-j} u \in \mathcal{E}(0, C)$. Therefore, by the previous argument

$$R_{s+jt/n} R_{t/n}^{n-j} u = (R_{s+jt/n} R_{t/n})(R_{t/n}^{n-j-1} u) = R_{s+(j+1)t/n} R_{t/n}^{n-j-1} u,$$

so by induction $R_{s+t} u = R_s R_{t/n}^n u$. Since this also holds with s replaced by 0, we have $R_{t/n}^n u = R_t u$. Thus, $R_{s+t} u = R_s R_t u$. $\qquad\qquad\square$

6G. *Solution to the differential equation.*
It remains to show that the semigroup R_t produces solutions to the differential equation $\partial_t u = (2N)^{-1} \Delta u - \frac{1}{2} \|Du\|_2^2$, and that the result agrees with the solution produced by solving the heat equation for $\exp(-N^2 u)$. More precisely, we will prove the following.

Theorem 6.17. *Let $u_0 : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$ be a given function in $\mathcal{E}(c, C)$ for some $c \geq 0$. Let $u(x, t) = R_t u(x)$. Then u is a smooth function on $M_N(\mathbb{C})_{\mathrm{sa}}^m \times (0, +\infty)$ and it solves the equation $\partial_t u = (2N)^{-1} \Delta u - \frac{1}{2} \|Du\|_2^2$. Moreover, $\exp(-N^2 \cdot R_t u_0) = P_t[\exp(-N^2 u_0)]$.*

At this point, we have not proved enough smoothness for $R_t u$ to show that it solves the equation in the classical sense. Therefore, as an intermediate step, we show that u solves the equation in the *viscosity sense* (for background on viscosity solutions, see [Crandall et al. 1992]). We will then deduce that $\exp(-N^2 u)$ is a viscosity solution of the heat equation and hence show it agrees with the smooth solution of the heat equation.

The definition of viscosity solution for parabolic equations is as follows. Here we continue to use the vector space $M_N(\mathbb{C})_{\mathrm{sa}}^m$ with the normalized inner product (rather than $\mathbb{R}^n$ for some n). For smooth $u : M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$, we denote by Du and Hu the gradient and Hessian with respect to the inner product $\langle \cdot, \cdot \rangle_2$; in other words, if $x_0 \in M_N(\mathbb{C})_{\mathrm{sa}}^m$, then $Du(x_0)$ is the vector in $M_N(\mathbb{C})_{\mathrm{sa}}^m$ and $Hu(x_0)$ is the linear transformation $M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})_{\mathrm{sa}}^m$ such that

$$u(x) = u(x_0) + \langle Du(x_0), x - x_0 \rangle_2 + \frac{1}{2} \langle Hu(x_0)[x - x_0], x - x_0 \rangle_2 + o(\|x - x_0\|_2^2).$$

We denote the space of linear transformations $M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})_{\mathrm{sa}}^m$ by $B(M_N(\mathbb{C})_{\mathrm{sa}}^m)$, and we denote the self-adjoint elements by $B(M_N(\mathbb{C})_{\mathrm{sa}}^m)_{\mathrm{sa}}$.

Definition 6.18. Let $F : B(M_N(\mathbb{C})_{\mathrm{sa}}^m)_{\mathrm{sa}} \times M_N(\mathbb{C})_{\mathrm{sa}} \times \mathbb{R} \times M_N(\mathbb{C})_{\mathrm{sa}} \to \mathbb{R}$ be continuous, and consider the partial differential equation

$$\partial_t u = F(Hu, Du, u, x). \tag{6-17}$$

We say that a function $u : M_N(\mathbb{C})_{\mathrm{sa}}^m \times [0, +\infty) \to \mathbb{R}$ is a *viscosity subsolution* if it is upper semicontinuous and if the following condition holds: Suppose that

$$x_0 \in M_N(\mathbb{C})_{\mathrm{sa}}^m, \quad t_0 > 0, \quad A \in B(M_N(\mathbb{C})_{\mathrm{sa}}^m)_{\mathrm{sa}}, \quad p \in M_N(\mathbb{C})_{\mathrm{sa}}^m, \quad \alpha \in \mathbb{R},$$

and suppose that u satisfies

$$u(x, t) \leq u(x_0, t_0) + \alpha(t - t_0) + \langle p, x - x_0 \rangle_2 + \tfrac{1}{2}\langle A(x - x_0), x - x_0 \rangle_2 + o(|t - t_0| + \|x - x_0\|_2^2). \quad (6\text{-}18)$$

Then we also have

$$\alpha \leq F(A, p, u(x_0), x_0). \quad (6\text{-}19)$$

Definition 6.19. With the same setup as above, we say that $u : M_N(\mathbb{C})_{\mathrm{sa}}^m \times [0, +\infty) \to \mathbb{R}$ is a *viscosity supersolution* if it is lower semicontinuous and the following condition holds: if x_0, t_0, A, p, α are as above and if

$$u(x, t) \geq u(x_0, t_0) + \alpha(t - t_0) + \langle p, x - x_0 \rangle_2 + \tfrac{1}{2}\langle A(x - x_0), x - x_0 \rangle_2 + o(|t - t_0| + \|x - x_0\|_2^2), \quad (6\text{-}20)$$

then

$$\alpha \geq F(A, p, u(x_0), x_0). \quad (6\text{-}21)$$

Definition 6.20. We say that u is a *viscosity solution* if it is both a subsolution and a supersolution.

Remark 6.21. Roughly speaking, being a viscosity solution means that whenever there exist upper or lower second-order Taylor approximations to u, then we can evaluate the differential operator F on the Taylor approximation and get an inequality in one direction.

Example 6.22. The heat equation $\partial_t u = (2N)^{-1}\Delta u$ is obtained by taking

$$F(A, p, u, x) = \frac{1}{2N^2}\operatorname{Tr}(A).$$

To understand why $1/N^2$ is the correct normalization on the right-hand side, suppose that u is smooth and $A = Hu(x_0)$ and $p = Du(x_0)$, so that

$$u(x) = u(x_0) + \langle p, x - x_0 \rangle_2 + \tfrac{1}{2}\langle A(x - x_0), x - x_0 \rangle_2 + o(\|x - x_0\|_2^2).$$

In terms of the nonnormalized inner product (which we denote by the dot product), this means that

$$u(x) = u(x_0) + \frac{1}{N} p \cdot (x - x_0) + \frac{1}{2N}(A(x - x_0)) \cdot (x - x_0).$$

Thus, the Hessian with respect to the nonnormalized inner product is $(1/N)A$. Hence, $(1/N)\Delta u(x_0) = (1/N^2)\operatorname{Tr}(A)$. Similarly, the equation $\partial_t u = (2N)^{-1}\Delta u - \tfrac{1}{2}\|Du\|_2^2$ is obtained by taking

$$F(A, p, u, x) = \frac{1}{2N^2}\operatorname{Tr}(A) - \frac{1}{2}\|p\|_2^2.$$

Proposition 6.23. *Let $u_0 \in \mathcal{E}(0, C)$ and define $u(x, t) = R_t u_0(x)$. Then u is a viscosity solution of the equation $\partial_t u = (2N)^{-1}\Delta u - \tfrac{1}{2}\|Du\|_2^2$.*

Proof. First, note that u is continuous. Indeed, by Theorem 6.1(3), u is continuous in t with a modulus of continuity that is uniform for x in a bounded region (this follows because the term $\|Du_0\|_2^2$ on the right-hand side of Lemma 6.16(2) is bounded on bounded regions since Du_0 is C-Lipschitz). Also, $u(\,\cdot\,, t)$ is continuous for each t since it is in $\mathcal{E}(0, C)$. Together, this implies u is jointly continuous in (x, t).

To show that u is a viscosity supersolution, suppose that we have a lower second-order approximation at the point (x_0, t_0), where $x_0 \in M_N(\mathbb{C})_{\mathrm{sa}}^m$ and $t_0 > 0$, given by

$$u(x, t) \geq u(x_0, t_0) + \alpha(t - t_0) + \langle p, x - x_0 \rangle + \tfrac{1}{2}\langle A(x - x_0), x - x_0 \rangle_2 + o(|t - t_0| + \|x - x_0\|_2^2).$$

Then we must show that $\alpha \geq 1/(2N^2) \operatorname{Tr}(A) - \tfrac{1}{2}\|p\|_2^2$.

Our first goal is to replace the soft bound $o(|t - t_0| + \|x - x_0\|_2^2)$ by a more explicit error bound, at the cost of modifying α and A by some positive ϵ. Pick $\epsilon > 0$. Then there exists $r > 0$ such that if $|t - t_0| + \|x - x_0\|_2^2 < 2r$, then we have

$$u(x, t) \geq u(x_0, t_0) + \alpha(t - t_0) - \epsilon|t - t_0| + \langle p, x - x_0 \rangle + \tfrac{1}{2}\langle (A - \epsilon I)(x - x_0), x - x_0 \rangle_2. \tag{6-22}$$

Let us assume that $t_0 - r < t \leq t_0$, so that the above inequality holds for $\|x - x_0\|^2 < r$ and we have $\alpha(t - t_0) - \epsilon|t - t_0| = (\alpha + \epsilon)(t - t_0)$. For x such that $\|x - x_0\|_2^2 \geq r$, we may use Theorem 6.1(3b), the fact that Du is C-Lipschitz, and the convexity of u to conclude that

$$u(x, t) \geq u_0(x) - \frac{t}{2}(Cm + \|Du\|_2^2)$$
$$\geq u_0(x_0) + \langle Du(x_0), x - x_0 \rangle_2 - \frac{t}{2}(Cm + (\|Du(x_0)\|_2 + C\|x - x_0\|_2)^2).$$

In other words, u is bounded below by a quadratic in $x - x_0$, and the estimate holds uniformly for t in a bounded interval. Moreover, the right-hand side of (6-22) is also bounded by a quadratic in $x - x_0$ uniformly for $t \in [t_0 - r, t_0 + r]$. It follows that for a large enough constant K_ϵ, we have

$$u(x_0, t_0) + (\alpha + \epsilon)(t - t_0) + \langle p, x - x_0 \rangle + \tfrac{1}{2}\langle (A - \epsilon I)(x - x_0), x - x_0 \rangle_2 - u(x, t) \leq K_\epsilon \|x - x_0\|_2^4$$

whenever $t \in (t_0 - t, t_0]$ and $\|x - x_0\|_2 \geq r$. Therefore, overall, assuming that $t \in (t_0 - r, t_0]$, we have

$$u(x, t) \geq u(x_0, t_0) + (\alpha + \epsilon)(t - t_0) + \langle p, x - x_0 \rangle + \tfrac{1}{2}\langle (A - \epsilon I)(x - x_0), x - x_0 \rangle_2 - K_\epsilon \|x - x_0\|_2^4. \tag{6-23}$$

For $t \in \mathbb{R}$, let us write $u_t(x) = u(x, t) = R_t u_0(x)$. Now the strategy for proving that $\alpha + \epsilon \geq (1/(2N^2)) \operatorname{Tr}(A - \epsilon I) - \tfrac{1}{2}\|p\|_2^2$ is roughly to use the fact that $u_{t_0}(x_0) = R_\delta u_{t_0 - \delta}(x_0)$ and estimate $u_{t_0 - \delta}(x_0)$ from above using the upper Taylor approximation for small $\delta > 0$. However, for the sake of computation, it is easier to estimate $Q_\delta P_\delta u_{t_0 - \delta}$ rather than R_δ (and then we will control the error between R_δ and $Q_\delta P_\delta$ using Lemmas 6.12 and 6.13).

Let $\delta \in (0, r)$. Then using the above inequality and monotonicity of P_δ, we have

$$P_\delta u_{t_0 - \delta}(x) \geq u_{t_0}(x_0) - (\alpha + \epsilon)\delta + \langle p, x - x_0 \rangle$$
$$+ \frac{1}{2N^2} \operatorname{Tr}(A - \epsilon I)\delta + \tfrac{1}{2}\langle (A - \epsilon I)(x - x_0), x - x_0 \rangle$$
$$- K_\epsilon \left(\|x - x_0\|_2^4 + 2\left(1 + \frac{2}{N^2}\right)m\delta\|x - x_0\|_2^2 + m^2\left(1 + \frac{2}{N^2}\right)\delta^2 \right).$$

Here we have evaluated P_δ applied to $\|x - x_0\|_2^4$ using Example 3.22 and the translation-invariance of P_δ. Now recall that $Q_\delta P_\delta u_{t_0 - \delta}(x_0)$ is obtained by evaluating $P_\delta u_{t_0 - \delta}$ at $x_0 - \delta D(Q_\delta P_\delta u_{t_0 - \delta})(x_0)$. Also, in

light of Lemma 6.11(4) and Corollary 6.14(4), $\|D(Q_\delta P_\delta u_{t_0-\delta})(x_0)\|_2^2$ is bounded by $\|Du_0(x_0)\|_2^2$ plus a constant. In particular, $\|D(Q_\delta P_\delta u_{t_0-\delta})(x_0)\|_2$ is bounded as $\delta \to 0$. Therefore,

$$Q_\delta P_\delta u_{t_0-\delta}(x_0) = P_\delta u_{t_0-\delta}(x_0 - \delta D(Q_\delta P_\delta u_{t_0-\delta})(x_0)) + \frac{\delta}{2}\|D(Q_\delta P_\delta u_{t_0-\delta})(x_0)\|_2^2$$

$$\geq u_{t_0}(x_0) + \frac{\delta}{2}\|D(Q_\delta P_\delta u_{t_0-\delta})(x_0)\|_2^2 + (\alpha + \epsilon)(-\delta)$$
$$- \langle p, D(Q_\delta P_\delta u_{t_0-\delta})(x_0)\rangle \delta + \frac{1}{2N^2}\operatorname{Tr}(A - \epsilon I)\delta + O(\delta^2). \quad (6\text{-}24)$$

(Here the implicit constant in $O(\delta^2)$ depends on ϵ.)

Because $u_{t_0-\delta} \in \mathcal{E}(0, C)$, Lemma 6.12(2) and (3) imply that if $C\delta \leq 1$, then

$$|Q_\delta P_\delta u_{t_0-\delta}(x_0) - P_\delta Q_\delta u_{t_0-\delta}(x_0)| \leq 2C^2 m\delta^2 + 2C\delta^2\|D(P_\delta Q_\delta u_{t_0-\delta})(x_0)\|_2.$$

Again by Lemma 6.11(4) and Theorem 6.1(4c), $\|D(Q_\delta P_\delta u_{t_0-\delta})(x_0)\|_2^2$ is bounded by $\|Du_0(x_0)\|_2^2$ plus a constant, so that

$$Q_\delta P_\delta u_{t_0-\delta}(x_0) = P_\delta Q_\delta u_{t_0-\delta}(x_0) + O(\delta^2).$$

Also, if we let $\delta_\ell = 2^{-\ell}$ for $\ell \in \mathbb{Z}$, then Lemma 6.13 implies that when $2C\delta_\ell \leq 1$ and $\delta_\ell < r$, we have

$$|P_{\delta_\ell} Q_{\delta_\ell} u_{t_0-\delta}(x_0) - R_{\delta_\ell,\ell} u_{t_0-\delta_\ell}(x_0)| = |R_{\delta_\ell,\ell} u_{t_0-\delta_\ell}(x_0) - R_{\delta_\ell} u_{t_0-\delta_\ell}(x_0)|$$
$$\leq \left(\frac{3}{2}\frac{C^2 m\delta_\ell}{1 - C\delta_\ell} + \log(1 + C\delta_\ell)(m + Cm + \|Du(x_0)\|_2^2)\right)2^{-\ell} = O(\delta_\ell^2).$$

So overall

$$Q_{\delta_\ell} P_{\delta_\ell} u_{t_0-\delta_\ell}(x_0) = R_{\delta_\ell} u_{t_0-\delta_\ell}(x_0) + O(\delta_\ell^2) = u_{t_0}(x_0) + O(\delta_\ell^2). \quad (6\text{-}25)$$

Using similar reasoning, Lemma 6.12(1) shows that

$$D(Q_{\delta_\ell} P_{\delta_\ell} u_{t_0-\delta_\ell})(x_0) = D(P_{\delta_\ell} Q_{\delta_\ell} u_{t_0-\delta_\ell})(x_0) + O(\delta_\ell^{3/2}).$$

Then using Lemma 6.13(3), we obtain

$$D(P_{\delta_\ell} Q_{\delta_\ell} u_{t_0-\delta_\ell})(x_0) = D(R_{\delta_\ell} u_{t_0-\delta_\ell})(x_0) + O(\delta^{3/2}).$$

Finally, because $u_{t_0-\delta} \in \mathcal{E}(0, C)$, it is differentiable everywhere; the upper Taylor approximation (6-22) implies that $u_{t_0}(x) \leq u_{t_0}(x_0) + \langle p, x - x_0\rangle_2 + o(\|x - x_0\|_2)$ and therefore p must equal $Du_{t_0}(x_0)$. Thus, overall

$$D(Q_{\delta_\ell} P_{\delta_\ell} u_{t_0-\delta_\ell})(x_0) = p + O(\delta_\ell^{3/2}). \quad (6\text{-}26)$$

Substituting (6-25) and (6-26) into (6-24), we obtain

$$u_{t_0}(x_0) \geq u_{t_0}(x_0) + \frac{1}{2}\|p\|_2^2\delta_\ell + (\alpha + \epsilon)(-\delta_\ell) - \|p\|_2^2\delta_\ell + \frac{1}{2N^2}\operatorname{Tr}(A - \epsilon I)\delta_\ell + O(\delta_\ell^2).$$

We cancel $u_{t_0}(x_0)$ from both sides, divide by δ_ℓ, and move $\alpha + \epsilon$ to the left-hand side to conclude that

$$\alpha + \epsilon \geq \frac{1}{2N^2}\operatorname{Tr}(A - \epsilon I)\delta_\ell - \frac{1}{2}\|p\|_2^2 + O(\delta_\ell).$$

Then taking $\ell \to \infty$, we get $\alpha + \epsilon \geq (1/(2N^2))\operatorname{Tr}(A - \epsilon I) - \frac{1}{2}\|p\|_2^2$. Since ϵ was arbitrary, we have $\alpha \geq (1/(2N^2))\operatorname{Tr}(A) - \frac{1}{2}\|p\|_2^2$. This shows that u is a viscosity supersolution.

To show that the u is a viscosity subsolution, the argument is symmetrical for the most part. However, to obtain the constant K_ϵ in (6-23), we used the one-sided estimate Theorem 6.1(3b) to show that u is bounded below by a quadratic in $x - x_0$ that is independent of t, so long as $t \in (t_0 - r, t_0]$. To show that u is a viscosity subsolution, we want to prove an analogous quadratic upper bound. But by Theorem 6.1(3a) and semiconcavity of u_0, we have for $t \le t_0$ that

$$u_t(x) \le u_0(x) + \frac{m}{2} \log(1 + Ct_0)$$

$$\le u_0(x_0) + \langle Du_0(x_0), x - x_0 \rangle + \frac{C}{2} \|x - x_0\|_2^2 + \frac{m}{2} \log(1 + Ct_0),$$

which is the desired upper bound. The rest of the argument is symmetrical except that $\alpha + \epsilon$ is replaced by $\alpha - \epsilon$ and $A - \epsilon I$ is replaced by $A + \epsilon I$. $\qquad\square$

Lemma 6.24. *Let $u : M_N(\mathbb{C})_{\mathrm{sa}}^m \times [0, +\infty) \to \mathbb{R}$. Then u is a viscosity solution to $\partial_t u = (2N)^{-1} \Delta u - \frac{1}{2} \|Du\|_2^2$ if and only if $\exp(-N^2 u)$ is a viscosity solution to $\partial_t u = (2N)^{-1} \Delta u$.*

Proof. More precisely, we claim that u is a viscosity subsolution if and only if $\exp(-N^2 u)$ is viscosity supersolution and vice versa. Suppose that u is a subsolution, and let us show that $v = \exp(-N^2 u)$ is a supersolution. If u is upper semicontinuous, then v is lower semicontinuous. Now suppose that we have a lower Taylor approximation at (x_0, t_0)

$$v(x, t) \ge v(x_0, t_0) + \alpha(t - t_0) + \langle p, x - x_0 \rangle_2 + \frac{1}{2} \langle A(x - x_0), x - x_0 \rangle_2 + o(|t - t_0| + \|x - x_0\|_2^2).$$

Note that $v > 0$ and $u = (-1/N^2) \log v$. The function $h \mapsto \log h$ is increasing and analytic for $h > 0$ and we have

$$\log(h + \delta) = \log(h) + \log\left(1 + \frac{\delta}{h}\right) = \log(h) + \frac{\delta}{h} - \frac{1}{2}\left(\frac{\delta}{h}\right)^2 + O(\delta^3).$$

Substituting $h = v(x_0, t_0) = \exp(-N^2 u(x_0, t_0))$ and

$$\delta = v(x, t) - v(x_0, t_0) = \alpha(t - t_0) + \langle p, x - x_0 \rangle_2 + \frac{1}{2} \langle A(x - x_0), x - x_0 \rangle_2 + o(|t - t_0| + |x - x_0|^2),$$

we get

$$-N^2 u(x, t) \ge -N^2 u(x_0, t_0) + \frac{\alpha}{v(x_0, t_0)}(t - t_0) + \frac{1}{v(x_0, t_0)} \langle p, x - x_0 \rangle_2$$

$$+ \frac{1}{2v(x_0, t_0)} \langle A(x - x_0), x - x_0 \rangle_2 - \frac{1}{2v(x_0, t_0)^2} \langle p, x - x_0 \rangle_2^2 + o(|t - t_0| + \|x - x_0\|_2^2),$$

since $\langle p, x - x_0 \rangle_2 / v(x_0, t_0)^2$ is the only term from $-(\delta/h)^2/2 + O(\delta^3)$ that is not $o(|t - t_0| + \|x - x_0\|_2^2)$ (here we use the fact that $|t - t_0| \|x - x_0\|_2 \le \frac{2}{3} |t - t_0|^{3/2} + \frac{1}{3} \|x - x_0\|_2^3$). Let us denote by P the linear map $P(x - x_0) = p\langle p, x - x_0 \rangle_2$. Then the above inequality becomes

$$u(x, t) \le u(x_0, t_0) - \frac{\alpha}{N^2 v(x_0, t_0)}(t - t_0) - \frac{1}{N^2 v(x_0, t_0)} \langle p, x - x_0 \rangle_2$$

$$- \frac{1}{2N^2 v(x_0, t_0)} \langle A(x - x_0), x - x_0 \rangle_2 + \frac{1}{2N^2 v(x_0, t_0)^2} \langle P(x - x_0), (x - x_0) \rangle + o(|t - t_0| + \|x - x_0\|_2^2).$$

Because u is a subsolution, we have

$$-\frac{\alpha}{N^2 v(x_0, t_0)} \le -\frac{1}{2N^4}\operatorname{Tr}(A) + \frac{1}{2N^4 v(x_0, t_0)^2}\operatorname{Tr}(P) - \frac{1}{2N^4 v(x_0, t_0)^2}\|p\|_2^2.$$

But $\operatorname{Tr}(P) = \|p\|_2^2$, so the last two terms cancel. Thus,

$$\alpha \ge \frac{1}{2N^2}\operatorname{Tr}(A)$$

as desired. So v is a supersolution.

A symmetrical argument shows that if v is a supersolution, then u is a subsolution. The other two claims are proved in the same way except using the Taylor expansion of the exponential function instead of the logarithm. $\qquad\square$

Now we are ready to prove Theorem 6.17 in the special case where u_0 is bounded below.

Lemma 6.25. *Suppose that $u_0 \in \mathcal{E}(0, C)$ is bounded below. Then*

$$\exp(-N^2 R_t u_0) = P_t[\exp(-N^2 u_0)].$$

Proof. Let $v(x, t) = \exp(-N^2 R_t u_0(x))$ and let $w(x, t) = P_t[\exp(-N^2 u_0)](x)$. Since u_0 is bounded below by some constant K, we have $R_t u_0 \ge K$ by monotonicity of R_t (see Corollary 6.14(3)) and the fact that it does not affect constant functions (since the same is true of P_t and Q_t). Hence, $v = \exp(-N^2 R_t u_0) \le \exp(-N^2 K)$. We also have $\exp(-N^2 u_0) \le \exp(-N^2 K)$ and hence $w \le \exp(-N^2 K)$.

Thus, v and w are both bounded, w is a smooth solution to the heat equation, and v is a viscosity solution by the previous lemma. We will conclude from this that $v = w$ (and this is nothing but a standard argument for the maximum principle together with the basic philosophy of viscosity solutions).

To show that $v \le w$, choose $\epsilon > 0$, and consider the function

$$\phi(x, t) = v(x, t) - w(x, t) - \frac{\epsilon}{2}\|x\|_2^2 - 2m\epsilon t.$$

Suppose for contradiction that $\phi > 0$ at some point. Since ϕ is continuous on $M_N(\mathbb{C})_{\mathrm{sa}}^m \times [0, +\infty)$ and since w and v are bounded, ϕ achieves a maximum at some (x_0, t_0). Since the maximum is strictly positive, we have $t_0 > 0$. Let

$$\psi(x, t) = w(x, t) + \frac{\epsilon}{2}\|x\|_2^2 + 2m\epsilon t,$$

so that $\phi(x, t) = v(x, t) - \psi(x, t)$. Then $\phi(x, t) \le \phi(x_0, t_0)$ implies that

$$v(x, t) \le v(x_0, t_0) + \psi(x, t) - \psi(x_0, t_0)$$
$$= v(x_0, t_0) + \partial_t \psi(x_0, t_0)(t - t_0) + \langle D\psi(x_0, t_0), x - x_0 \rangle_2$$
$$+ \tfrac{1}{2}\langle H\psi(x_0, t_0)(x - x_0), x - x_0 \rangle_2 + o(|t - t_0| + \|x - x_0\|_2^2),$$

where the last step follows because ψ is smooth. Because v is a viscosity subsolution,

$$\partial_t \psi(x_0, t_0) \le \frac{1}{2N}\Delta\psi(x_0, t_0).$$

However, this is a contradiction because at every point (x, t), we have

$$\partial_t \psi = \partial_t w + 2m\epsilon > \frac{1}{2N} \Delta w + m\epsilon = \frac{1}{2N} \Delta \psi,$$

by computation and the fact that w solves the heat equation. It follows that $\phi \leq 0$ and hence $v(x, t) \leq w(x, t) + (\epsilon/2)\|x\|_2^2 + 2m\epsilon t$. Since ϵ was arbitrary, $v \leq w$. Then a symmetrical argument shows that $v \geq w$. □

Thus, to prove Theorem 6.17, it only remains to remove the boundedness assumption on u_0. We achieve this by replacing u_0 with the function

$$\tilde{u}_0(x) = u_0(x) - \langle Du_0(0), x\rangle_2, \tag{6-27}$$

which is nonnegative by the convexity of u_0 and hence it is bounded below.

Lemma 6.26. *Let $u_0 \in \mathcal{E}(0, C)$ and let $\tilde{u}_0$ be given by (6-27). Let $v_0 = \exp(-N^2 u_0)$ and $\tilde{v}_0 = \exp(-N^2 \tilde{u}_0)$. Then the integral defining $P_t \exp(-N^2 u_0)$ is well-defined and also*

$$P_t v_0(x) = \exp(-N^2 \langle Du_0(0), x\rangle + \frac{N^2 t}{2}\|Du_0(0)\|_2^2)P_t \tilde{v}_0(x - t Du_0(0)).$$

Proof. We can write

$$d\sigma_{t,N}(y) = \frac{1}{Z_N} \exp\left(-\frac{N^2}{2t}\|y\|_2^2\right) dy.$$

Also, set $p = Du_0(0)$. Then

$$
\begin{aligned}
P_t v_0(x) &= \frac{1}{Z_N} \int \exp(-N^2 u_0(x + y)) \exp\left(-\frac{N^2}{2t}\|y\|_2^2\right) dy \\
&= \frac{1}{Z_N} \int \exp\left(-N^2 \tilde{u}_0(x + y) - N^2\langle p, x + y\rangle - \frac{N^2}{2t}\|y\|_2^2\right) dy \\
&= \frac{1}{Z_N} \int \exp\left(-N^2 \tilde{u}_0(x + y) - N^2\langle p, x\rangle + \frac{N^2 t}{2}\|p\|_2^2 - \frac{N^2}{2t}\|y + tp\|_2^2\right) dy \\
&= \frac{1}{Z_N} \int \exp\left(-N^2 \tilde{u}_0(x - tp + z) - N^2\langle p, x\rangle + \frac{N^2 t}{2}\|p\|_2^2 - \frac{N^2}{2t}\|z\|_2^2\right) dz \\
&= \exp\left(-N^2\langle p, x\rangle + \frac{N^2 t}{2}\|p\|_2^2\right)P_t \tilde{v}_0(x - tp). \qquad \square
\end{aligned}
$$

Lemma 6.27. *Let $u_0 \in \mathcal{E}(0, C)$, let $p \in M_N(\mathbb{C})_{\text{sa}}^m$, and let $\tilde{u}_0(x) = u_0(x) - \langle p, x\rangle_2$. Then:*

(1) $P_t u_0(x) = P_t \tilde{u}_0(x) + \langle p, x\rangle_2$.

(2) $Q_t u_0(x) = Q_t \tilde{u}_0(x - tp) + \langle p, x\rangle_2 - (t/2)\|p\|_2^2$.

(3) $R_t u_0(x) = R_t \tilde{u}_0(x - tp) + \langle p, x\rangle_2 - (t/2)\|p\|_2^2$.

Proof. (1) holds because P_t is a linear operator and it also does not affect linear functions. To prove (2), fix x and let y be the point where the infimum defining $Q_t u_0(x)$ is achieved and let $\tilde{y}$ be the point where the infimum defining $Q_t \tilde{u}_0(x - tp)$ is achieved. By Corollary 6.7(1), the points y and $\tilde{y}$ are characterized respectively by the relations

$$y = x - t Du_0(y), \quad \tilde{y} = x - tp - t D\tilde{u}_0(\tilde{y}).$$

But $D\tilde{u}_0(\tilde{y}) = Du_0(\tilde{y}) - p$. Thus, $x - t\,Du_0(\tilde{y}) = \tilde{y}$, so that $y = \tilde{y}$. Then

$$
\begin{aligned}
Q_t u_0(x) &= u_0(y) + \frac{1}{2t} \|y - x\|_2^2 \\
&= \tilde{u}_0(y) + \langle p, y \rangle_2 + \frac{1}{2t} \|y - x\|_2^2 \\
&= \tilde{u}_0(y) + \langle p, x \rangle_2 - \frac{t}{2} \|p\|_2^2 + \frac{1}{2t} \|y - (x - tp)\|_2^2 \\
&= Q_t \tilde{u}_0(x - tp) + \langle p, x \rangle_2 - \frac{t}{2} \|p\|_2^2.
\end{aligned}
$$

(3) It follows by iteration (after some computation) that for $t \in 2^{-\ell} \mathbb{N}_0$, we have

$$
R_{t,\ell} u_0(x) = R_t \tilde{u}_0(x - tp) + \langle p, x \rangle_2 - \frac{t}{2} \|p\|_2^2.
$$

Then by Lemma 6.13, we may take $\ell \to \infty$, and by Theorem 6.1(3), we may extend the inequality to all real t. □

Proof of Theorem 6.17. We have already proved the case where u_0 is bounded. For the general case, let $u_0 \in \mathcal{E}(0, C)$. Define $p = Du_0(0)$ and $\tilde{u}_0(x) = u_0(x) - \langle p, x \rangle_2$. As remarked above, $\tilde{u}_0$ is bounded below by zero. By Lemma 6.26, the bounded case, and Lemma 6.27,

$$
\begin{aligned}
P_t \exp(-N^2 u_0)(x) &= \exp\!\Big(-N^2 \langle p, x \rangle + \frac{N^2 t}{2} \|p\|_2^2\Big)[P_t \exp(-N^2 \tilde{u}_0)](x - tp) \\
&= \exp\!\Big(-N^2 \langle p, x \rangle + \frac{N^2 t}{2} \|p\|_2^2\Big) \exp(-N^2 R_t \tilde{u}_0(x - tp)) \\
&= \exp\!\Big(-N^2 \Big(R_t \tilde{u}_0(x - tp) + \langle p, x \rangle - \frac{t}{2} \|p\|_2^2\Big)\Big) \\
&= \exp(-N^2 R_t u_0(x)).
\end{aligned}
$$

In particular, since $P_t \exp(-N^2 \tilde{u}_0)$ is smooth for $t > 0$, we see that all the functions in the above equation are smooth for $t > 0$, and hence $R_t u_0(x)$ is smooth function of (x, t). Also, $P_t[\exp(-N^2 u_0)] = \exp(-N^2 R_t u_0)$ as desired. □

6H. *Approximation by trace polynomials.* Now we are ready to prove that R_t preserves asymptotic approximability by trace polynomials.

Proposition 6.28. *Let $\{V_N\}$ be a sequence of functions $M_N(\mathbb{C})_{\mathrm{sa}}^m \to \mathbb{R}$ such that V_N is convex and $V_N(x) - (C/2)\|x\|_2^2$ is concave, and $\{DV_N\}$ is asymptotically approximable by trace polynomials. Then for every $t \geq 0$, the sequences $\{D(P_t V_N)\}$, $\{D(Q_t V_N)\}$, and $\{D(R_t V_N)\}$ are asymptotically approximable by trace polynomials.*

Proof. The fact that $\{D(P_t V_N)\}$ is asymptotically approximable by trace polynomials follows from Lemma 3.28.

Now consider $D(Q_t V_N)$. Note that by Corollary 6.7(1), $D(Q_t V_N)(x)$ is the solution of the fixed point equation

$$
y = DV_N(x - ty).
$$

But if $t < 1/C$, then $y \mapsto DV_N(x - ty)$ is a contraction and thus iterates of this function will converge to the fixed point. Let us define $\phi_{N,0}(x) = 0$ and $\phi_{N,\ell+1}(x) = DV_N(x - t\phi_{N,\ell}(x))$. By Lemma 6.6(5), the distance from 0 to the fixed point $D(Q_t V_N)(x)$ is bounded by $\|DV_N(x)\|_2$; hence

$$\|\phi_{N,\ell}(x) - D(Q_t V_N)(x)\|_2 \le C^\ell t^\ell \|DV_N(x)\|_2.$$

Because $DV_{N,t}$ is C-Lipschitz, Lemma 3.27 implies that $\{\phi_{N,\ell}\}_N$ is asymptotically approximable by trace polynomials.

Now $\|DV_N(0)\|_2$ is bounded by some constant A as $N \to \infty$ because DV_N is asymptotically approximable by trace polynomials. Since DV_N is also C-Lipschitz, $\|DV_N(x)\|_2 \le A + C\|x\|_2$. In particular, $\|\phi_{N,\ell}(x) - D(Q_t V_N)(x)\|_2 \le C^\ell t^\ell (A + C\|x\|_2)$. Thus, by Observation 3.26, $\{D(Q_t V_N)\}$ is asymptotically approximable by trace polynomials.

This holds whenever $t < 1/C$. But for general t, we can write $Q_t = Q_{t/n}^n$, where n is large enough that $t/n < 1/C$, and then iterating the previous statement shows that $\{Q_t V_N\}$ is asymptotically approximable by trace polynomials.

For the sequence $\{D(R_t V_N)\}$, first note that when $t \in \mathbb{Q}_2^+$, we know $\{D(R_{t,\ell} V_N)\}$ is asymptotically approximable by trace polynomials (where ℓ is large enough that $R_{t,\ell}$ is defined). By Theorem 6.1(1c) and Observation 3.26, the sequence $\{D(R_t V_N)\}$ is asymptotically approximable by trace polynomials for $t \in \mathbb{Q}_2^+$. Finally, by Theorem 6.1(2d) and Observation 3.26, the sequence $\{D(R_t V_N)\}$ is asymptotically approximable by trace polynomials for all $t \in \mathbb{R}^+$. $\qquad\square$

7. Main theorem on free entropy

We are now ready to prove the following theorem which shows that $\chi = \chi^*$ for a law which is the limit of log-concave random matrix models.

Theorem 7.1. *Let μ_N be a sequence of probability measures on $M_N(\mathbb{C})_{\mathrm{sa}}^m$ given by the potential V_N. Assume*:

(A) *The potential $V_N(x)$ is convex and $V_N(x) - (C/2)\|x\|_2^2$ is concave for some $C > 0$ independent of N.*

(B) *The sequence μ_N concentrates around some noncommutative law λ with $\lambda(X_j^2) > 0$.*

(C) *For some $R_0 > 0$, we have $\lim_{N \to \infty} \int_{\|x\| \ge R_0} (1 + \|x\|_2^2)\, d\mu_N(x) = 0$.*

(D) *The sequence $\{DV_N\}$ is asymptotically approximable by trace polynomials.*

Then $\lambda \in \Sigma_{m,R_0}$ and moreover:

(1) *The law λ has finite Fisher information $\Phi^*(\lambda)$, and for all $t \ge 0$, we have*

$$\lim_{N \to \infty} \frac{1}{N^3} \mathcal{I}(\mu_N * \sigma_{t,N}) \to \Phi^*(\lambda \boxplus \sigma_t).$$

(2) *We have for all $t \ge 0$*

$$\chi(\lambda \boxplus \sigma_t) = \underline{\chi}(\lambda \boxplus \sigma_t) = \lim_{N \to \infty} \frac{1}{N^2}\left(h(\mu_N * \sigma_{t,N}) + \frac{m}{2}\log N\right) = \chi^*(\lambda \boxplus \sigma_t).$$

(3) *The functions* $t \mapsto (1/N^3)\mathcal{I}(\mu_N * \sigma_{t,N})$ *and* $t \mapsto \Phi^*(\lambda \boxplus \sigma_t)$ *are decreasing and Lipschitz and the absolute value of the derivative* (*where defined*) *is bounded by* $C^2 m(1 + Ct)^{-2}$.

Remark 7.2. If $V_N(x) - (c/2)\|x\|_2^2$ is convex and $V_N(x) - (C/2)\|x\|_2^2$ is concave and if $\{DV_N\}$ is asymptotically approximable by trace polynomials, then Theorem 4.1 implies that μ_N satisfies the hypotheses of Theorem 7.1 for some noncommutative law λ.

However, Theorem 7.1 holds in a slightly more general situation than Theorem 4.1 in that we do not have to assume uniform convexity, finite moments, or exponential concentration.

In preparation for the proof of Theorem 7.1, we have already verified that the hypotheses (A), (C), and (D) are preserved under Gaussian convolution. Now we show that (B) is preserved in Lemma 7.4. This is straightforward apart from one subtlety — although we have assumed that for every noncommutative polynomial p, the noncommutative moment $\tau_N(p(x))$ concentrates around $\lambda(p)$ under μ_N, we have *not* assumed that $|\tau_N(p(x))|$ has finite expectation. To deal with this issue, we first prove an auxiliary lemma.

Lemma 7.3. *Let* λ *be a noncommutative law in* Σ_m, *let* $p(X, Y) = p(X_1, \ldots, X_m, Y_1, \ldots, Y_m)$ *be a noncommutative polynomial of* $2m$ *variables, and let* $R > 0$. *Then there exists a neighborhood* $\mathcal{V}$ *of* λ *in* Σ_m *and a constant* K *such that, for all* $N \in \mathbb{N}$, *for all* $x \in \Gamma_N(\mathcal{V})$, *the function* $y \mapsto \tau_N(p(x, y))$ *is* K-*Lipschitz with respect to* $\| \cdot \|_2$ *for self-adjoint tuples* y *in the operator-norm ball* $\{y : \|y_j\| \leq R\}$.

Proof. To prove the lemma, it suffices to consider the case of a noncommutative monomial. Indeed, if $p = \sum_{j=1}^n p_j$, where p_j is a monomial, and if we find neighborhoods $\mathcal{V}_j$ and Lipschitz constants K_j for each p_j, then the result will also hold for p with $\mathcal{V} = \bigcap_{j=1}^n \mathcal{V}_j$ and $K = \sum_{j=1}^n K_j$.

Thus, assume without loss of generality that $p(X, Y)$ is a noncommutative monomial. Then it can be written in the form

$$p(X, Y) = q_0(X)Y_{i_1}q_1(X)Y_{i_2} \cdots q_{\ell-1}(X)Y_{i_\ell}q_\ell(X),$$

where $i_j \in \{1, \ldots, m\}$ and $q_j(X)$ is a noncommutative monomial in X (which of course is allowed to be 1). Consider $x, y, y' \in M_N(\mathbb{C})_{\text{sa}}^m$, and suppose that $\|y_i\|_\infty \leq R$ and $\|y_i'\|_\infty \leq R$ for each i. Then

$$p(x, y) - p(x, y') = \sum_{j=1}^\ell q_0(x)y_{i_1} \cdots y_{i_{j-1}}q_{j-1}(x)(y_{i_j} - y_{i_j}')q_i(x)y_{i_{j+1}} \cdots y_{i_\ell}q_\ell(x).$$

Recalling the noncommutative L^α norms and Hölder's inequality (see Section 2C), we have

$$\|p(x, y) - p(x, y')\|_1 \leq \left(\sum_{j=1}^\ell \prod_{k \neq j} \|q_j(x)\|_{2(\ell+1)} \prod_{k<j} \|y_{i_k}\|_\infty \prod_{k>j} \|y_{i_k}'\|_\infty \right) \|y_j - y_j'\|_2.$$

This implies that

$$|\tau_N(p(x, y)) - \tau_N(p(x, y'))| \leq \left(\sum_{j=1}^\ell \prod_{k \neq j} \|q_j(x)\|_{2(\ell+1)} \right) R^{\ell-1} \|y - y'\|_2.$$

Now

$$\|q_j(x)\|_{2(\ell+1)} = (\tau_N[(q_j(x)^* q_j(x))^{\ell+1}])^{1/(2(\ell+1))}.$$

We can define

$$\mathcal{V} = \{\lambda' : \lambda'[(q_j^* q_j)^{\ell+1}] < \lambda[(q_j^* q_j)^{\ell+1}] + 1 \text{ for } j = 0, \dots, \ell\}.$$

Then $\|q_j(x)\|_{2(\ell+1)}$ is uniformly bounded for $x \in \Gamma_N(\mathcal{V})$ for each $j = 0, \dots, \ell$. Suppose that each of these quantities is bounded by K. Then the above estimate shows that

$$|\tau_N(p(x, y)) - \tau_N(p(x, y))| \leq \ell K^{\ell+1} R^{\ell-1} \|y - y'\|_2$$

whenever $x \in \Gamma_N(\mathcal{V})$ and y, y' are in the operator-norm ball of radius R. □

Lemma 7.4. *Suppose that $\{\mu_N\}$ concentrates around a noncommutative law λ. Then $\{\mu_N * \sigma_{t,N}\}$ concentrates around $\lambda \boxplus \sigma_t$ for every $t > 0$.*

Proof. Fix t. Let $X_N = (X_{N,1}, \dots, X_{N,m})$ and $Y_N = (Y_{N,1}, \dots, Y_{N,m})$ be independent random variables with the laws μ_N and $\sigma_{t,N}$ respectively. Because the topology on the space Σ_m of noncommutative laws is generated by noncommutative moments, it suffices to show that for each noncommutative polynomial p and $\delta > 0$

$$\lim_{N \to \infty} P(|\tau_N(p(X_N + Y_N)) - \lambda \boxplus \sigma_t(p)| \geq \delta) = 0.$$

Fix p and let d be its degree. By the previous lemma, there is a neighborhood $\mathcal{V}$ of λ and a constant K such that for every $x \in \Gamma_N(\mathcal{V})$, the function $y \mapsto \tau_N(p(x + y))$ is K-Lipschitz with respect to $\|\cdot\|_2$ on the operator-norm ball $\{y : \|y\|_\infty \leq 4t^{1/2}\}$. By shrinking $\mathcal{V}$ if necessary, we may also assume that $\tau_N(q(x))$ is uniformly bounded for every noncommutative monomial $q(x)$ of degree less than or equal to d.

Choose a C_c^∞ function $\psi : \mathbb{R} \to \mathbb{R}$ such that $\psi(z) = z$ for $|z| \leq 3t^{1/2}$ and $|\psi(z)| \leq 4t^{1/2}$. Then

$$\Psi : (y_1, \dots, y_m) \mapsto (\psi(y_1), \dots, \psi(y_m))$$

is globally Lipschitz in $\|\cdot\|_2$ and it also maps $M_N(\mathbb{C})_{sa}^m$ into the operator-norm ball of radius $4t^{1/2}$ (which is the region where $z \mapsto \tau_N(p(x, z))$ was assumed to be K-Lipschitz with respect to $\|\cdot\|_2$ whenever $x \in \Gamma_N(\mathcal{V})$). This implies that there is some constant K' such that $y \mapsto \tau_N(p(x, \Psi(y)))$ is K'-Lipschitz for all $x \in \Gamma_N(\mathcal{V})$.

Let

$$\alpha_N(x) = E[\tau_N(p(x + \Psi(Y_N)))],$$

$$\beta_N(x) = E[\tau_N(p(x + Y + N))] = \exp\left(\frac{tL_N}{2}\right)[\tau(p)](x),$$

$$\beta(x) = \exp\left(\frac{tL}{2}\right)[\tau(p)](x).$$

By Theorem 2.10 applied to Y_N,

$$x \in \Gamma_N(\mathcal{V}) \implies P\left(|\tau_N(p(x + \Psi(Y_N))) - \alpha_N(x)| \geq \frac{\delta}{3}\right) \leq 2e^{-\delta^2 N^2/(18t(K')^2)}.$$

On the other hand, we know by standard tail estimates on the GUE (see Corollary 2.12) that

$$\lim_{N \to \infty} E[\tau_N(q(Y_N))\mathbf{1}_{\|Y_N\| \geq 3t^{1/2}}] = 0$$

for every noncommutative polynomial q. This implies that $|\alpha_N(x) - \beta_N(x)| \to 0$ uniformly for $x \in \Gamma_N(\mathcal{V})$. On the other hand, by Lemma 3.21,

$$\beta_N(x) = \exp\left(\frac{tL_N}{2}\right)[\tau(p)](x) \to \exp\left(\frac{tL}{2}\right)[\tau(p)](x) = \beta(x)$$

where the convergence occurs coefficientwise. Now $\exp(tL_N/2)[\tau(p)]$ is a sum of products of traces of noncommutative monomials q of degree $\leq d$ and for every such q, we know $\tau_N(q(x))$ is uniformly bounded on $\Gamma_N(\mathcal{V})$ by our choice of $\mathcal{V}$. Thus, coefficientwise convergence of $\beta_N \to \beta$ implies uniform convergence for $x \in \Gamma_N(\mathcal{V})$. Therefore, for sufficiently large N we have $|\beta_N(x) - \beta(x)| \leq \delta/3$ for $x \in \Gamma_N(\mathcal{V})$, and hence

$$P\left(|\tau_N(p(X_N + Y_N)) - \tau(\beta(X_N))| \geq \frac{2\delta}{3}, \ X_N \in \Gamma_N(\mathcal{V}), \ \|Y_N\| \leq 3t^{1/2}\right) \leq 2e^{-\delta^2 N^2/(18t(K')^2)},$$

where we have applied the Fubini–Tonelli theorem for the product measure $\mu_N \otimes \sigma_{t,N}$. By our concentration assumption,

$$P\left(|\tau_N(\beta(X_N)) - \lambda(\beta)| \geq \frac{\delta}{3}\right) \to 0, \quad P(X_N \in \Gamma_N(\mathcal{V})) \to 1,$$

and by Corollary 2.12 also $P(\|Y_k\| \geq 3t^{1/2}) \to 0$. Altogether, we have

$$P(|\tau_N(p(X_N + Y_N)) - \lambda(\beta)| \geq \delta) \to 0.$$

But note that $\lambda(\beta) = \lambda(\exp(tL/2)[\tau(p)]) = (\lambda \boxplus \sigma_t)(p)$ by Lemma 3.23. Thus, the proof is complete. $\square$

Proof of Theorem 7.1. Let $V_{N,t} = R_t V_N$ be the potential associated to $\mu_N * \sigma_{t,N}$. Let us verify that $V_{N,t}$ satisfies the assumptions (A)–(D) for every $t > 0$.

(A) This follows from Theorem 6.1(1) because $V_{N,t} = R_t V_N$; hence $V_{N,t} \in \mathcal{E}(0, C)$.

(B) This follows from Lemma 7.4.

(C) This follows from tail bounds on the GUE (Corollary 2.12).

(D) This follows from Proposition 6.28.

   Next, the fact that $\lambda \in \Sigma_{m, R_0}$ follows from Proposition 5.5 with $n = 1$.
   Claim (1) of the theorem follows by applying Proposition 5.10 to $\mu_N * \sigma_{t,N}$ with $n = 1$.
   For claim (2), recall that by Lemma 5.7, (5-6),

$$\frac{1}{N^2}h(\mu_N) + \frac{m}{2}\log N = \frac{1}{2}\int_0^\infty \left(\frac{m}{1+t} - \frac{1}{N^3}\mathcal{I}(\mu_N * \sigma_{t,N})\right) ds + \frac{m}{2}\log 2\pi e. \qquad (7\text{-}1)$$

Because $N^{-3}\mathcal{I}(\mu_N)$ converges as $N \to \infty$, there is some constant K with $N^{-3}\mathcal{I}(\mu_N) \leq K$ for all N. Also, because of assumptions (B) and (C), we have $\int \|x\|_2^2 \, d\mu_N(x) \to \sum_{j=1}^m \lambda_j(X_j^2) > 0$. Therefore, there is a constant a such that $\int \|x\|_2^2 \, d\mu_N(x) \geq ma$ for large enough N. Thus, (5-4), we have for sufficiently large N that

$$\frac{m}{a+t} \leq \frac{1}{N^3}\mathcal{I}(\mu_N * \sigma_{t,N}) \leq \min\left(M, \frac{m}{t}\right).$$

Thus, we can apply the dominated convergence theorem to take the limit as $N \to \infty$ inside the integral on the right-hand side of (7-1) and apply claim (1) to conclude that

$$\lim_{N \to \infty} \left(\frac{1}{N^2} h(\mu_N) + \frac{m}{2} \log N \right) \to \chi^*(\lambda).$$

On the left-hand side of (7-1), we will apply Proposition 5.5 with $n = 1$. We may replace V_N by $V_N - V_N(0)$ without changing μ_N (because the definition of μ_N includes the normalizing constant Z_N anyway). Then because $\{DV_N\}$ is asymptotically approximable by trace polynomials, we know that $\{V_N\}$ is asymptotically approximable by trace polynomials (Lemma 3.29). Therefore, the hypotheses of Proposition 5.5 are satisfied and so

$$\chi(\lambda) = \limsup_{N \to \infty} \left(\frac{1}{N^2} h(\mu_N) + \frac{m}{2} \log N \right) = \chi^*(\lambda)$$

and the same holds for $\underline{\chi}(\lambda)$. Moreover, this holds for $\mu_N * \sigma_{t,N}$ just as well as μ_N because $\mu_N * \sigma_{t,N}$ satisfies the same assumptions (A)–(D).

For claim (3), first fix N and let X be a random variable with law μ_N, and let Y_t be an independent Hermitian Brownian motion (here $Y_t \sim \sigma_{t,N}$). Let $\Xi_t = DV_{N,t}(X + Y_t)$, which is the conjugate variable of $X + Y_t$. Then

$$\frac{1}{N^3} \mathcal{I}(\mu_N * \sigma_{t,N}) = E \|\Xi_t\|_2^2$$

Suppose $0 \leq s \leq t \leq T$. Then $X + Y_t$ is the sum of the independent random variables $X + Y_s$ and $Y_t - Y_s$, and thus $\Xi_t = E[\Xi_s | X + Y_t]$ by Lemma 5.6. In other words, Ξ_t is the orthogonal projection of $DV_{N,s}(X + Y_s)$ onto the space of L^2 random variables that are functions of $X + Y_t$, or in other words it is the function of $X + Y_t$ that is closest to Ξ_s in L^2. This implies that

$$[\|\Xi_s - \Xi_t\|_2^2] \leq E[\|DV_{N,s}(X + Y_s) - DV_{N,s}(X + Y_t)\|_2^2]$$
$$\leq E\left[\frac{C^2}{(1 + Cs)^2} \|Y_s - Y_t\|_2^2 \right]$$
$$= \frac{C^2}{(1 + Cs)^2} m(t - s)$$

using the fact that $V_{N,s} \in \mathcal{E}(0, C(1 + Cs)^{-1})$ and hence $DV_{N,s}$ is $C(1 + Cs)^{-1}$-Lipschitz. Since Ξ_t is the orthogonal projection of Ξ_s onto this subspace, we know $\Xi_s - \Xi_t$ is orthogonal to Ξ_t and hence

$$E[\|\Xi_s\|_2^2] - E[\|\Xi_t\|_2^2] = E[\|\Xi_s - \Xi_t\|_2^2].$$

Overall,

$$0 \leq \frac{1}{N^3} \mathcal{I}(\mu_N * \sigma_{s,N}) - \frac{1}{N^3} \mathcal{I}(\mu_N * \sigma_{t,N}) \leq \frac{C^2}{(1 + Cs)^2} m(t - s).$$

This immediately proves that $t \mapsto N^{-3} \mathcal{I}(\mu_N * \sigma_{t,N})$ is a decreasing function of t, it is Lipschitz, and the absolute value of the derivative is bounded by $C^2 m / (1 + Ct)^2$. The same holds for $\Phi^*(\lambda \boxplus \sigma_t)$ by taking the limit as $N \to \infty$. $\qquad \square$

8. Free Gibbs laws

In the situation of Theorem 4.1, we want to interpret the law λ as the free Gibbs state for a potential which is the limit of the V_N. To this end, we will define a noncommutative function space where each element is a limit of functions on $M_N(\mathbb{C})_{\text{sa}}^m$. We will then give several characterizations of the closure of trace polynomials in this space, as well as the class of potentials to which our previous results apply.

8A. *Asymptotic approximation and function spaces.* Let $Y_\bullet = \{Y_N\}$ be a sequence of normed vector spaces. We define a (possibly infinite) seminorm on sequences $\phi_\bullet = \{\phi_N\}$ of functions $M_N(\mathbb{C})_{\text{sa}}^m \to Y_N$ by

$$\|\phi_\bullet\|_{R,Y_\bullet} = \limsup_{N\to\infty} \sup_{\|x\|\leq R} \|\phi_N(x)\|_{Y_N}.$$

Let $\mathcal{F}_m(Y_\bullet)$ be the vector space

$$\{\phi_\bullet : \|\phi_\bullet\|_{R,Y_\bullet} < +\infty \text{ for all } R\}/\{\phi_\bullet : \|\phi_\bullet\|_{R,Y_\bullet} = 0 \text{ for all } R\}.$$

For a sequence $\phi_\bullet$, we denote its equivalence class by $[\phi_\bullet]$.

We equip $\mathcal{F}_m(Y_\bullet)$ with the topology generated by the seminorms $\|\cdot\|_{R,Y_\bullet}$, or equivalently given by the metric

$$d_{\mathcal{F}_m(Y_\bullet)}(\phi_\bullet, \psi_\bullet) = \sum_{n=1}^{\infty} \frac{1}{2^n} \min(\|\phi_\bullet - \psi_\bullet\|_{n,Y}, 1). \tag{8-1}$$

Note that $\mathcal{F}_m(Y_\bullet)$ is a complete metric space in this metric and is a locally convex topological vector space.

There is a canonical map from the vector space of scalar-valued trace polynomials TrP_m^0 into $\mathcal{F}_m^0 := \mathcal{F}_m(\mathbb{C})$ by the map that sends a trace polynomial to the corresponding sequence of functions it defines on $M_N(\mathbb{C})_{\text{sa}}^m$. A sequence $\phi_\bullet$ is asymptotically approximable by trace polynomials if and only if $[\phi_\bullet]$ is in the closure of the image of TrP_m^0 in $\mathcal{F}_m^0$, which we will denote by $\mathcal{T}_m^0$. (Unfortunately, we do not know whether the map $\text{TrP}_m^0 \to \mathcal{F}_m^0$ is injective, but this point is irrelevant for our purposes.)

Similarly, let $M_\bullet(\mathbb{C})^m$ be the sequence $\{M_N(\mathbb{C})^m\}$ equipped with $\|\cdot\|_2$. There is a canonical map from TrP_m^1 into $\mathcal{F}_m^1 := \mathcal{F}_m(M_\bullet(\mathbb{C}))$ given by mapping a trace polynomial to the corresponding sequence of functions on matrices. A sequence $\phi_\bullet$ of functions $M_N(\mathbb{C})_{\text{sa}}^m \to M_N(\mathbb{C})_{\text{sa}}$ is asymptotically approximable by trace polynomials if and only if $[\phi_\bullet]$ is in the closure of the image of TrP_m^1, which we denote by $\mathcal{T}_m^1$.

The spaces $\mathcal{T}_m^0$ and $\mathcal{T}_m^1$ can be viewed as noncommutative function spaces through the following alternative characterization. Here $\mathcal{R}$ denotes the hyperfinite II$_1$ factor and $\mathcal{R}^\omega$ denotes its ultrapower; for an explanation, see [Anantharaman and Popa 2016, §1.6 and §5.4] or [Capraro 2010, pp. 5–7].

Lemma 8.1. *Let $f \in \text{TrP}_m^0$. Then we have*

$$\limsup_{\substack{N\to\infty \\ \substack{x\in M_N(\mathbb{C})_{\text{sa}}^m \\ \|x\|_\infty\leq R}}} |f(x)| = \sup_N \sup_{\substack{x\in M_N(\mathbb{C})_{\text{sa}}^m \\ \|x\|_\infty\leq R}} |f(x)| = \sup_{\substack{x\in(\mathcal{R}_{\text{sa}}^\omega)^m \\ \|x\|_\infty\leq R}} |f(x)|. \tag{8-2}$$

If we denote the common value by $\|f\|_{\mathcal{T}_m^0, R}$, then this family of seminorms defines a metrizable topology on TrP_m^0 with the metric given as in (8-1), and $\mathcal{T}_m^0$ is the completion of TrP_m^0 in this metric. The same

result holds for $\mathcal{T}_m^1$ using the seminorm

$$\limsup_{\substack{N \to \infty \\ \substack{x \in M_N(\mathbb{C})_{\text{sa}}^m \\ \|x\|_\infty \leq R}}} \|f(x)\|_2 = \sup_N \sup_{\substack{x \in M_N(\mathbb{C})_{\text{sa}}^m \\ \|x\|_\infty \leq R}} \|f(x)\|_2 = \sup_{\substack{x \in (\mathcal{R}_{\text{sa}}^\omega)^m \\ \|x\|_\infty \leq R}} \|f(x)\|_2. \tag{8-3}$$

Proof. Fix f and let A, B, and C be the three quantities in (8-2) from left to right. It is clear that $A \leq B$. Moreover, $B \leq C$ because there is an isometric trace-preserving embedding of $M_N(\mathbb{C})$ into $\mathcal{R}^\omega$. To show that $C \leq A$, pick $x \in (\mathcal{R}_{\text{sa}}^\omega)^m$ with $\|x\| \leq R$. Then there exists $x_n \in \mathcal{R}_{\text{sa}}^m$ with $\|x_n\| \leq R$ and $x = \lim_{n \to \omega} x_n$. For each n, we can choose an N_n, an embedding $M_{N_n}(\mathbb{C}) \to \mathcal{R}$ and a $y_n \in M_{N_n}(\mathbb{C})$ such that $\|y_n\| \leq R$ and $\|x_n - y_n\|_2 \leq 1/2^n$ and $\lim_{n \to \infty} N_n = +\infty$. Then $x = \lim_{n \to \omega} y_n$ and $|f(x)| = \lim_{n \to \omega} |f(y_n)| \leq A$. This shows that the three seminorms in (8-2) are equal, and the other claims follow because these seminorms are the same as the seminorms for $\mathcal{F}_m^0$. □

From this point of view, every $f \in \mathcal{T}_m^0$ has a canonical sequence that represents its equivalence class in $\mathcal{F}_m^0$, constructed as follows. If we write f as the limit of a sequence of trace polynomials $f^{(k)}$, then $f_{(k)}|_{M_N(\mathbb{C})_{\text{sa}}^m}$ converges locally uniformly on $M_N(\mathbb{C})_{\text{sa}}^m$ as $k \to \infty$ and the limit is independent of the approximating sequence $f^{(k)}$. We can therefore define $f|_{M_N(\mathbb{C})_{\text{sa}}^m}$ to be this limit.

Similarly, f defines a function on $(\mathcal{R}_{\text{sa}}^\omega)^m$. Moreover, if $(\mathcal{M}, \tau)$ is a tracial von Neumann algebra and there is a trace-preserving embedding $\iota : \mathcal{M} \to \mathcal{R}^\omega$, then we may define $f|_\mathcal{M} = f \circ \iota$. It is easy to see that this is independent of the choice of trace-preserving embedding if f is a trace polynomial, and this holds for general $f \in \mathcal{T}_m^0$ or $\mathcal{T}_m^1$ by density of trace polynomials. In this sense, $\mathcal{T}_m^0$ and $\mathcal{T}_m^1$ represent spaces of universal scalar- or operator-valued functions that can be applied to self-adjoint operators in every $\mathcal{R}_{\text{sa}}^\omega$-embeddable tracial von Neumann algebra.

In the scalar-valued case, we have yet another characterization of $\mathcal{T}_m^0$:

Lemma 8.2. *Let $\Sigma_{m,\text{bdd}} = \bigcup_{R>0} \Sigma_{m,R}$. Let $C(\Sigma_{m,\text{bdd}})$ be the space of functions $g : \Sigma_{m,\text{bdd}} \to \mathbb{C}$ such that $g \in C(\Sigma_{m,R})$ for every R, equipped with the family of seminorms $\| \cdot \|_{C(\Sigma_{m,R})}$. Then $\mathcal{T}_m^0$ is isomorphic to $C(\Sigma_{m,\text{bdd}})$ as a topological vector space.*

Proof. For a scalar-valued trace polynomial f, the value $f(x)$ only depends on the law of x, so that $f(x) = g(\lambda_x)$ for some function $g : \Sigma_m \to \mathbb{R}$ such that $g \in C(\Sigma_{m,R})$ for all R, and we have

$$\|f\|_{\mathcal{T}_m^0, R} = \|g\|_{C(\Sigma_{m,R})}.$$

Passing to the completion with respect to the metric defined as in (8-1), we have a map $\iota : \mathcal{T}_m^0 \to C(\Sigma_{m,\text{bdd}})$ which is an isomorphism onto its image. To show that ι is surjective, note the algebra of trace polynomials is self-adjoint and separates points in $\Sigma_{m,R}$, and hence by the Stone–Weierstrass theorem, trace polynomials are dense in $C(\Sigma_{m,R})$ for every R. Therefore, if $g \in C(\Sigma_{m,R})$, we can choose a trace polynomial $g^{(k)}(\lambda_x) = f^{(k)}(x)$ such that $\|g - g^{(k)}\|_{C(\Sigma_{m,k})} \leq 1/2^k$. Then $f^{(k)}$ converges to some f in $\mathcal{T}_m^0$, and we have $\iota(f) = g$. □

8B. *Convex differentiable functions.* Now we are ready to characterize the type of convex functions which occur in Theorem 7.1. First of all, we let $\mathcal{T}_m^{0,1}$ be the completion of the trace polynomials with

respect to the metric

$$d(f, g) = \sum_{n=1}^{\infty} \frac{1}{2^n} [\min(1, \|f - g\|_{\mathcal{T}_m^0, n}) + \min(1, \|Df - Dg\|_{(\mathcal{T}_m^1)^m, n})].$$

Observe that if $f \in \mathcal{T}_m^{0,1}$ and $f^{(k)}$ is a sequence of trace polynomials converging to f in $\mathcal{T}_m^{0,1}$ as $k \to \infty$, then $Df^{(k)}$ converges in $(\mathcal{T}_m^1)^m$ and the limit is independent of the choice of approximating sequence. We denote this limit by Df.

Remark 8.3. If f and $f^{(k)}$ are as above, then since $Df^{(k)}$ is a tuple of trace polynomials, it is continuous on the operator norm ball $\{y \in M_N(\mathbb{C})_{\text{sa}}^m : \|y\|_\infty \leq R\}$ with a modulus of continuity that only depends on R and does not depend on N. Because $Df^{(k)} \to Df$ uniformly on the operator-norm ball (with rate of convergence independent of N), we know Df is also continuous on this operator-norm ball with modulus of continuity independent of N.

It follows that for every $x, y \in M_N(\mathbb{C})_{\text{sa}}^m$ with $\|x\|, \|y\| \leq R$ we have

$$f(y) - f(x) = \langle Df(x), y - x \rangle_2 + o(\|y - x\|_2),$$

where the error estimate only depends on R and not on N. In particular, this shows Df is uniquely determined by f. Also, it shows that $Df|_{M_N(\mathbb{C})_{\text{sa}}^m}$ is equal to the normalized gradient of $f|_{M_N(\mathbb{C})_{\text{sa}}^m}$ in the ordinary sense of functions on $M_N(\mathbb{C})_{\text{sa}}^m \cong \mathbb{R}^{mN^2}$.

Lemma 8.4. *Let $f \in \mathcal{T}_m^{0,1}$ be real-valued. The following are equivalent:*

(1) *The function $f|_{M_N(\mathbb{C})_{\text{sa}}^m}$ is convex for every N.*

(2) *The function f is convex as a function on $(\mathcal{R}_{\text{sa}}^\omega)^m$.*

(3) *There exists a sequence of differentiable convex functions $V_N : M_N(\mathbb{C})_{\text{sa}}^m \to \mathbb{R}$ such that $[V_\bullet] = f$ and $[DV_\bullet] = Df$. (Here $DV_\bullet$ denotes the sequence $(DV_N)_{N \in \mathbb{N}}$, where D is the normalized gradient understood in the standard sense of calculus.)*

Proof. The implication (1) $\implies$ (2) follows from an argument similar to the proof of Lemma 8.1.

The implication (1) $\implies$ (3) holds because we can take $V_N = f|_{M_N(\mathbb{C})_{\text{sa}}^m}$.

Now we will prove (3) $\implies$ (1). Fix N. To prove that $f|_{M_N(\mathbb{C})_{\text{sa}}^m}$ is convex, it suffices to show that $\langle Df(x) - Df(y), x - y \rangle_2 \geq 0$ for every $x, y \in M_N(\mathbb{C})_{\text{sa}}^m$. For $k \in \mathbb{N}$, consider $x \otimes I_k$ and $y \otimes I_k$ in $M_{Nk}(\mathbb{C})_{\text{sa}}^m$. Then, as $k \to \infty$,

$$\langle Df(x) - Df(y), x - y \rangle_2 = \langle Df(x \otimes I_k) - Df(y \otimes I_k), x \otimes I_k - y \otimes O_k \rangle_2;$$

meanwhile, if $R = \max(\|x\|, \|y\|)$, then since $DV_N - Df \to 0$ in $\|\cdot\|_2$ uniformly on the operator norm ball of radius R, we have as $k \to \infty$ that

$$\langle Df(x \otimes I_k) - Df(y \otimes I_k), x \otimes I_k - y \otimes I_k \rangle_2 - \langle DV_{Nk}(x \otimes I_k) - DV_{Nk}(y \otimes I_k), x \otimes I_k - y \otimes I_k \rangle_2 \to 0.$$

Because V_{Nk} is convex, the second inner product is ≥ 0 and therefore $\langle Df(x) - Df(y), x - y \rangle_2 \geq 0$. $\square$

Let $\mathcal{E}_m(c, C)^{0,1}$ denote the class of $V \in \mathcal{T}_m^{0,1}$ such that $V(x) - (c/2)\|x\|_2^2$ is convex and $V(x) - (C/2)\|x\|_2^2$ is concave. If $0 < c < C$ and $V \in \mathcal{E}_m(c, C)^{0,1}$ and $V_N = V|_{M_N(\mathbb{C})_{\text{sa}}^m}$, then the sequence of normalized gradients DV_N is asymptotically approximable by trace polynomials. If we let μ_N be the corresponding measure on $M_N(\mathbb{C})_{\text{sa}}^m$, then Theorem 4.1 (the hypothesis (4-1) being trivially satisfied by unitary invariance) implies that μ_N concentrates around a noncommutative law λ_V, which we will call the *free Gibbs state for the potential V*.

Furthermore, the free Gibbs state λ_V is independent of the choice of representative sequence in the following sense. Let μ_N be the measure on $M_N(\mathbb{C})_{\text{sa}}^m$ given by the potential $V_N = V|_{M_N(\mathbb{C})_{\text{sa}}^m}$. Let W_N be another sequence of potentials satisfying the hypotheses of Theorem 4.1 such that $[W_\bullet] = V$ in $\mathcal{T}_m^{0,1}$, and let ν_N be the sequence of random matrix measures given by W_N. By Theorem 4.1, ν_N concentrates around some noncommutative law λ. We claim that $\lambda = \lambda_V$. To prove this, consider the sequence $\widetilde{V}_N$ which equals V_N for odd N and W_N for even N. Then $[\widetilde{V}_\bullet] = V$ in $\mathcal{T}_m^{0,1}$, which means that $\{D\widetilde{V}_N\}_{N \in \mathbb{N}}$ is asymptotically approximable by trace polynomials. Therefore,

$$\lambda_V(p) = \lim_{\substack{N \text{ even} \\ N \to \infty}} \int \tau_N(p) \, d\mu_N = \lim_{\substack{N \text{ odd} \\ N \to \infty}} \int \tau_N(p) \, d\nu_N = \lambda(p).$$

In fact, Lemma 8.4 implies that the noncommutative laws λ which occur as limits in Theorem 4.1 are precisely the free Gibbs laws for potentials $V \in \mathcal{E}_m(c, C)^{0,1}$. In particular, Theorem 7.1 implies that $\chi = \underline{\chi} = \chi^*$ for every such law.

Remark 8.5. We have *not* proved that the law λ_V is uniquely characterized by the Schwinger–Dyson equation $\lambda[DV(X)f(X)] = \lambda \otimes \lambda[\mathcal{D}f(X)]$, although something like this is implied by [Dabrowski 2016]. One could hope to prove this by letting the semigroup T_t^V act on an abstract space of Lipschitz functions which is the completion of trace polynomials (where the metric now allows x to come from any tracial von Neumann algebra rather than only the $\mathcal{R}^\omega$-embeddable algebras). We would want to show that if λ satisfies the Schwinger–Dyson equation, then $\lambda(T_t^V u) = \lambda(u)$, but to justify the computation, we need to show more regularity of $T_t^V u$ than we have done in this paper. In the SDE approach as well, the proof that λ_V is characterized by Schwinger–Dyson is subtle when we do not assume more regularity for V (see [Dabrowski 2010; 2016]).

8C. *Examples of convex potentials.* A natural class of examples of functions in $\mathcal{E}_m(c, C)^{0,1}$ are those of the form

$$V(x) = \tfrac{1}{2}\|x\|_2^2 + \epsilon f(u)$$

where ϵ is a small positive parameter,

$$u = (u_1, \ldots, u_m), \quad u_j = \frac{x_j + 4i}{x_j - 4i},$$

and f is a real-valued trace polynomial in u and u^*. Computations similar to those of Section 3B show that the normalized Hessian of $\text{Jac}(Df(u(x)))$ with respect to x is bounded uniformly in N. Therefore, $V \in \mathcal{E}_m\left(\tfrac{1}{2}, \tfrac{3}{2}\right)^{0,1}$ for sufficiently small ϵ. Similar examples are described in the introduction of [Dabrowski

2016]. More generally, we can replace the trace polynomial $f(u)$ by a power series where the individual terms are trace monomials in u.

The class $\mathcal{E}_m(c, C)^{0,1}$ does not include trace polynomials in x because if g is a trace polynomial of degree ≥ 3, then we cannot have $g(x)$ convex and $g(x) - (C/2)\|x\|_2^2$ concave (globally). However, if we consider a potential which is a small perturbation of a quadratic (as considered in [Guionnet and Maurel-Segala 2006; Guionnet and Shlyakhtenko 2014]), we can fix this problem by introducing an operator-norm cut-off as follows.

Let f be a scalar-valued trace polynomial and let us define

$$V^{(\epsilon)}(x) = \|x\|_2^2 + \epsilon f(x). \tag{8-4}$$

Let $\phi : \mathbb{R} \to \mathbb{R}$ be a C_c^∞ function such that $\phi(t) = t$ for $|t| \leq R$ and $\phi(t) = 0$ for $|t| \geq 2R$. Let $\Phi : M_N(\mathbb{C})_{\mathrm{sa}}^m \to M_N(\mathbb{C})_{\mathrm{sa}}^m$ be given by $\Phi_N(x) = (\phi(x_1), \ldots, \phi(x_m))$.

$$\widetilde{V}_N^{(\epsilon)}(x) = \|x\|_2^2 + \epsilon f_N(\Phi_N(x)). \tag{8-5}$$

We will prove the following.

Proposition 8.6. *Let $\widetilde{V}_N^{(\epsilon)}$ be given as above. Then $[\widetilde{V}_\bullet^{(\epsilon)}] \in \mathcal{T}_m^{0,1}$. Moreover, given $\delta > 0$, we have $[\widetilde{V}_\bullet^{(\epsilon)}] \in \mathcal{E}_m(1 - \delta, 1 + \delta)^{0,1}$ for sufficiently small ϵ (depending on f, R, and δ).*

As a consequence, we will deduce the following result about measures defined by $V^{(\epsilon)}$ restricted to an operator-norm ball (without the smooth cut-off Φ).

Proposition 8.7. *Let $2 < R' < R$, let f be a trace polynomial, and let $V^{(\epsilon)}$ be as in (8-4). Let*

$$d\mu_N^{(\epsilon)}(x) = \frac{1}{Z_N} \exp(-N^2 V_N^{(\epsilon)}(x)) \mathbf{1}_{\|x\| \leq R} \, dx.$$

For sufficiently small ϵ (depending on f, R, and R'), we have the following. The measure $\mu_N^{(\epsilon)}$ exhibits exponential concentration around a noncommutative law $\lambda^{(\epsilon)} \in \Sigma_{m,R'}$. If $X \in (\mathcal{M}, \tau)$ is a noncommutative m-tuple realizing the law $\lambda^{(\epsilon)}$, then the conjugate variable is given by $DV^{(\epsilon)}(X)$. Moreover, we have

$$\chi(\lambda^{(\epsilon)}) = \underline{\chi}(\lambda^{(\epsilon)}) = \chi^*(\lambda^{(\epsilon)}) = \lim_{N \to \infty} \left(\frac{1}{N^2} h(\mu_N^{(\epsilon)}) + \frac{m}{2} \log N \right).$$

To fix notation for the remainder of this section, functions without a subscript, such as f, will denote elements of $\mathcal{T}_m^0$ or $\mathcal{T}_m^{0,1}$, and Df will denote the "gradient" defined in the abstract space $\mathcal{T}_m^{0,1}$ as the limit of the "gradients" of trace polynomials approximating f. However, f_N will denote $f|_{M_N(\mathbb{C})_{\mathrm{sa}}^m}$, and Df_N will denote the normalized gradient $N\nabla f_N$ defined in the usual sense of calculus with respect to $\langle \cdot, \cdot \rangle_2$ on $M_N(\mathbb{C})_{\mathrm{sa}}^m$. Moreover, $Hf_N = \mathrm{Jac}(Df_N)$ will denote the Hessian of f_N with respect to $\langle \cdot, \cdot \rangle_2$.

In order to prove Proposition 8.6, we must understand $D[f_N \circ \Phi_N]$ and $H[f_N \circ \Phi_N]$. To this end, we recall some results of [Peller 2006] on noncommutative derivatives of $\phi(x)$, where ϕ is a smooth function on the real line.

For a polynomial ϕ in one variable, the noncommutative derivative $\mathcal{D}\phi \in \mathbb{C}\langle X \rangle \otimes \mathbb{C}\langle X \rangle$ defined by Definition 3.6 can be written as the difference quotient

$$\mathcal{D}\phi(s, t) = \frac{\phi(s) - \phi(t)}{s - t},$$

where we view $\mathbb{C}\langle X \rangle \otimes \mathbb{C}\langle X \rangle$ as a subset of functions on $\mathbb{R}^2$ with the variables s and t. However, the above difference quotient makes sense whenever $\phi : \mathbb{R} \to \mathbb{C}$ is smooth. Thus, it defines an extension of $\mathcal{D}$ to continuously differentiable functions ϕ of one variable.

Similarly, if ϕ is a polynomial, then the higher-order noncommutative derivatives $\mathcal{D}^n \phi$ can be viewed as functions of $n + 1$ variables, which are obtained through iterated difference quotients and thus their definition can be extended to smooth functions ϕ. (However, beware that we have *not* defined $\mathcal{D}_j^n \phi$ if ϕ is a nonpolynomial function of *multiple* variables.)

If ϕ is a polynomial, then to estimate $\phi(X) - \phi(Y)$ for operators X and Y with norm bounded by R, one seeks to control the norm of $\mathcal{D}\phi$ in the projective tensor product $L^\infty[-R, R] \widehat{\otimes} L^\infty[-R, R]$. Similarly, if ϕ is a smooth function and $\phi(X)$ and $\phi(Y)$ are defined through functional calculus, one can estimate the operator norm $\|\phi(X) - \phi(Y)\|$ by representing ϕ as an integral of simpler functions (e.g., by Fourier analysis) whose noncommutative derivatives are easier to analyze. In this case, it is convenient to write $\mathcal{D}\phi$ as an *integral* rather than a *sum* of simple tensors.

We thus consider the integral projective tensor powers of the space of bounded Borel functions $\mathcal{B}(\mathbb{R})$. The *integral projective tensor product* $\mathcal{B}(\mathbb{R})^{\widehat{\otimes}_i n}$ consists of Borel functions G on $\mathbb{R}^n$ which admit a representation

$$G(x_1, \ldots, x_n) = \int_\Omega G_1(x_1, \omega) \cdots G_n(x_n, \omega) \, d\mu(\omega) \tag{8-6}$$

for some measure space (Ω, μ) such that

$$\int_\Omega \|G_1(\,\cdot\,, \omega)\|_{\mathcal{B}(\mathbb{R})} \cdots \|G_1(\,\cdot\,, \omega)\|_{\mathcal{B}(\mathbb{R})} \, d\mu(\omega) < +\infty \tag{8-7}$$

and we define $\|G\|_{\mathcal{B}(\mathbb{R})^{\widehat{\otimes}_i n}}$ to be the infimum of (8-7) over all representations (8-6).

Given $G \in \mathcal{B}(\mathbb{R})^{\widehat{\otimes}_i n}$, bounded self-adjoint operators $x_0, \ldots, x_n$ and bounded operators $y_1, \ldots, y_n$, we define

$$G(x_0, \ldots, x_n) \# (y_1 \otimes \cdots \otimes y_n) = \int_\Omega G_0(x_0, \omega) y_1 G_1(x_1, \omega) \cdots y_n G_n(x_n, \omega) \, d\mu(\omega), \tag{8-8}$$

where $G_0, \ldots, G_n$ satisfy (8-6). This is well-defined by [Peller 2006, Lemma 3.1]. If the x_j and y_j are elements of a tracial von Neumann algebra $(\mathcal{M}, \tau)$, we have by the noncommutative Hölder's inequality (see Section 2C) that if $1/\alpha = 1/\alpha_1 + \cdots + 1/\alpha_n$, then

$$\|G(x_0, \ldots, x_n) \# (y_1 \otimes \cdots \otimes y_n)\|_\alpha \leq \|G\|_{\mathcal{B}(\mathbb{R})^{\widehat{\otimes}_i (n+1)}} \|y_1\|_{\alpha_1} \cdots \|y_n\|_{\alpha_n}. \tag{8-9}$$

Moreover, we have the following bounds on the noncommutative derivatives of ϕ as a corollary of the results of [Peller 2006].

Proposition 8.8. *There exists a constant K_n such that for all $\phi \in C_c^\infty(\mathbb{R})$*

$$\|\mathcal{D}^n \phi\|_{\mathcal{B}(\mathbb{R})^{\widehat{\otimes}_i (n+1)}} \le K_n \int_{\mathbb{R}} |\hat{\phi}(\xi)\xi^n| \, d\xi. \tag{8-10}$$

Proof. As in [Peller 2006, §2], choose $w \in C_c^\infty$ such that $0 \le w \le \chi_{[-1/2,2]}$ and $\sum_{k \in \mathbb{Z}} w(2^{-k}\xi) = 1$ for $\xi > 0$. Let W_k and $W_k^\#$ be given by $\widehat{W}_k(\xi) = w(2^{-n}\xi)$ and $\widehat{W}_k^\#(\xi) = w(-2^{-k}x)$, where $\hat{\ }$ denotes the Fourier transform. It is shown in [Peller 2006, Theorem 5.5] that

$$\|\mathcal{D}^n \phi\|_{\mathcal{B}(\mathbb{R})^{\widehat{\otimes}_i (n+1)}} \le K_n \sum_{k \in \mathbb{Z}} 2^{nk} (\|W_k * \phi\|_{L^\infty(\mathbb{R})} + \|W_k^\# * \phi\|_{L^\infty(\mathbb{R})}).$$

This can be estimated by the right-hand side of (8-10) (for a possibly different constant) by a standard Fourier analysis computation. $\qquad\square$

Proof of Proposition 8.6. Recall that $\widetilde{V}_N^{(\epsilon)}(x) = \frac{1}{2}\|x\|_2^2 + \epsilon f_N \circ \Phi_N$. Thus, to show that the sequence $V_N^{(\epsilon)}$ defines an element of $\mathcal{T}_0^m$, it suffices to prove this for $f_N \circ \Phi_N$. To this end, it is sufficient to show that for each $r > 0$ there is a sequence of trace polynomials $\{g^{(k)}\}_{k \in \mathbb{N}}$ such that

$$\lim_{k \to \infty} \sup_{N \in \mathbb{N}} \sup_{x \in M_N(\mathbb{C})_{\mathrm{sa}}^m : \|x\|_\infty \le r} |g^{(k)}(x) - f_N \circ \Phi_N(x)| = 0$$

and

$$\lim_{k \to \infty} \sup_{N \in \mathbb{N}} \sup_{x \in M_N(\mathbb{C})_{\mathrm{sa}}^m : \|x\|_\infty \le r} \|Dg^{(k)}(x) - D[f_N \circ \Phi_N(x)]\|_2.$$

Fix $r > 0$. By standard approximation techniques, there exist Schwarz functions $\phi^{(k)} : \mathbb{R} \to \mathbb{R}$ such that $\phi^{(k)}|_{[-r,r]}$ is a polynomial and $\phi^{(k)} \to \phi$ in the Schwarz space as $k \to \infty$. By Proposition 8.8, we have $\mathcal{D}^n \phi^{(k)} \to \mathcal{D}^n \phi$ in $\mathcal{B}(\mathbb{R})^{\widehat{\otimes}_i (n+1)}$ as $k \to \infty$ for every n.

Let $\Phi_N^{(k)}(x_1, \ldots, x_m) = (\phi^{(k)}(x_1), \ldots, \phi^{(k)}(x_m))$. Then $f_N \circ \Phi_N^{(k)}$ is given by a trace polynomial $g^{(k)}$ on $\{\|x\|_\infty \le r\}$. Because of the spectral mapping theorem,

$$\sup_{\|x\| \le r} \|\Phi_N^{(k)}(x) - \Phi_N(x)\|_\infty \le m \sup_{t \in [-r,r]} |\phi^{(k)}(t) - \phi(t)|,$$

which is independent of N and vanishes as $k \to \infty$. Thus, our trace polynomials $g^{(k)}$ approximate $f_N \circ \Phi_N$ uniformly on the operator norm ball $\{x : \|x\|_\infty \le r\}$.

Next, we must show that $Dg^{(k)}$ approximates $D[f_N \circ \Phi_N]$ uniformly in $\|\cdot\|_2$ on the operator-norm ball $\{\|x\|_\infty \le r\}$. By the chain rule, we have

$$D_j[f_N \circ \Phi_N] = \mathrm{Jac}_j(\Phi_N)^t [D_j f_N],$$

where D_j and Jac_j are the normalized gradient and Jacobian with respect to the variable $x_j \in M_N(\mathbb{C})_{\mathrm{sa}}$. Now

$$\mathrm{Jac}_j(\Phi_N)(x)y = \mathcal{D}\phi(x_j) \# y.$$

Now $\mathcal{D}\phi$ viewed as an element of the tensor product $\mathbb{C}[X] \otimes \mathbb{C}[X]$ is invariant under the flip map that switches the order of the tensorands; this is because $\mathcal{D}\phi$ is represented as a difference quotient for

one-variable functions. Flip invariance implies that

$$\tau_N[(\mathcal{D}\phi(x_j) \,\#\, y)z] = \tau_N[y(\mathcal{D}\phi(x_j) \,\#\, z)],$$

which means that the operator $\mathrm{Jac}_j(\Phi_N)(x)$ on $M_N(\mathbb{C})_{\mathrm{sa}}$ is self-adjoint. Hence,

$$D_j[f_N \circ \Phi_N](x) = \mathrm{Jac}_j(\Phi_N(x))[D_j f_N](x) = \mathcal{D}\phi(x_j) \,\#\, D_j f_N(\Phi_N(x)).$$

This function is given by a trace polynomial on $\{\|x\|_\infty \le r\}$, and it is also equal to the trace polynomial $D_j g^{(k)}$ when evaluated on any tuple of matrices because both functions are equal to the gradient of $g^{(k)}|_{M_N(\mathbb{C})^m_{\mathrm{sa}}}$. Moreover, for $\|x\|_\infty \le r$, we have

$$\mathcal{D}\phi_k(x_j) \,\#\, D_j f(\Phi_k(x)) = \mathcal{D}\phi_k(x_j) \,\#\, D_j f(\Phi(x)) + \mathcal{D}\phi_k(x_j) \,\#\, [D_j f(\Phi_k(x)) - D_j f(\Phi(x))].$$

The first term converges to $\mathcal{D}\phi(x_j) \,\#\, D_j f(\Phi(x))$ in $\|\cdot\|_2$ uniformly on $\{\|x\|_\infty \le r\}$ using (8-9) with estimates independent of N. Similarly, because the images of Φ_k and Φ are contained in an operator norm ball and $D_j f$ is K-Lipschitz in $\|\cdot\|_2$ on this ball for some $K > 0$, we have $D_j f(\Phi_k(x)) - D_j f(\Phi(x)) \to 0$ uniformly. This in turn implies that the second term goes to zero because $\mathcal{D}\phi_k(x_j)$ is uniformly bounded in $\mathcal{B}(\mathbb{R}) \widehat{\otimes}_i \mathcal{B}(\mathbb{R})$. Thus, for every $r > 0$, there is a sequence of trace polynomials $g^{(k)}$ such that $g_k \to f \circ \Phi$ and $Dg^{(k)} \to D(f \circ \Phi)$ uniformly on $\{\|x\|_\infty \le r\}$. This means that $f \circ \Phi \in \mathcal{T}^{1,0}_m$.

It follows that the sequence $\widetilde{V}^{(\epsilon)}_N$ defines a function in $\mathcal{T}^{0,1}_m$ for every ϵ. It remains to show that this function is in $\mathcal{E}_m(1 - \delta, 1 + \delta)^{0,1}$ for sufficiently small ϵ. To this end, it suffices to show that $f_N \circ \Phi_N$ defines a function in $\mathcal{E}_m(-a, a)^{0,1}$ for some real $a > 0$. Thus, we only need to obtain some upper and lower bounds on the operator norm of $H[f_N \circ \Phi_N]$ that are independent of N. However, this is equivalent to showing that $D_j(f_N \circ \Phi_N) = \mathcal{D}\phi_N(x_j) \,\#\, D_j f_N(\Phi_N(x))$ is Lipschitz in $\|\cdot\|_2$ for each j (uniformly in N). Because $\mathcal{D}^2\phi$ is bounded in $\mathcal{B}(\mathbb{R}) \widehat{\otimes}_i \mathcal{B}(\mathbb{R}) \widehat{\otimes}_i \mathcal{B}(\mathbb{R})$, we see that

$$\|\mathcal{D}\phi(x_j) \,\#\, y - \mathcal{D}\phi(x_j') \,\#\, y\|_2 \le K \|x_j - x_j'\|_2 \|y\|_\infty$$

for some constant K; we may apply this to $y = D_j f_N(\Phi_N(x))$, which is bounded in $\|\cdot\|_\infty$ because $D_j f_N$ is a trace polynomial and $\Phi_N(x)$ is bounded in $\|\cdot\|_\infty$. Together with the fact that $D_j f_N(\Phi_N(x))$ is Lipschitz in $\|\cdot\|_2$, this implies that $D_j(f_N \circ \Phi_N)$ is Lipschitz in $\|\cdot\|_2$ as desired. $\qquad \square$

Proof of Proposition 8.7. Let $\tilde{\mu}^{(\epsilon)}_N$ be the measure on $M_N(\mathbb{C})^m_{\mathrm{sa}}$ given by the potential $\widetilde{V}^{(\epsilon)}_N$. Let δ be a number in $(0, 1)$ to be chosen later. By Proposition 8.6, we have that $\widetilde{V}^{(\epsilon)} \in \mathcal{E}_m(1 - \delta, 1 + \delta)^{0,1}$ for sufficiently small ϵ. By Theorem 4.1, the laws $\tilde{\mu}_N$ concentrate around a noncommutative law λ. Furthermore, in Theorem 4.1(1), we can take $M = 0$ and $c = 1 - \delta$ and $C = 1 + \delta$, so that

$$\limsup_{N \to \infty} R_N \le \frac{2}{(1 - \delta)^{1/2}} + \frac{\|D\widetilde{V}^{(\epsilon)}(0)\|_2}{1 - \delta} + \frac{\delta}{(1 - \delta)^{3/2}}.$$

Note that $D\widetilde{V}^{(\epsilon)}(0) = DV^{(\epsilon)}(0) = \epsilon Df(0)$ is a scalar multiple of the identity matrix since f is a trace polynomial. Because $R' > 2$, we may choose δ sufficiently small that

$$\frac{2}{(1 - \delta)^{1/2}} + \frac{\delta}{(1 - \delta)^{3/2}} < R'.$$

Then by choosing ϵ (and hence $\|D\widetilde{V}^{(\epsilon)}(0)\|_2$) sufficiently small, we can arrange that

$$R_* = \limsup_{N \to \infty} R_N < R'.$$

This implies that the measures $\tilde{\mu}_N^{(\epsilon)}$ concentrate on the ball $\{\|x\|_\infty \le R'\}$. For $\|x\|_\infty \le R$, we have $\widetilde{V}^{(\epsilon)}(x) = V^{(\epsilon)}(x)$, and therefore $\mu_N^{(\epsilon)}$ is the (normalized) restriction of $\tilde{\mu}_N^{(\epsilon)}$ to $\{\|x\|_\infty \le R\}$. It follows that $\mu_N^{(\epsilon)}$ concentrates around the law $\lambda^{(\epsilon)}$ as well.

If $X \in (\mathcal{M}, \tau)$ realizes the law $\lambda^{(\epsilon)}$, then $\|X\|_\infty \le R'$ since $\lambda \in \Sigma_{m,R_*} \subseteq \Sigma_{m,R'}$ by Theorem 4.1(2). Moreover, by Proposition 5.10, the conjugate variables for λ are given by $D\widetilde{V}(X) = DV(X)$. Moreover, by Theorem 7.1 applied to $\tilde{\mu}_N^{(\epsilon)}$, we have

$$\chi(\lambda^{(\epsilon)}) = \underline{\chi}(\lambda^{(\epsilon)}) = \chi^*(\lambda^{(\epsilon)}) = \lim_{N \to \infty} \left(\frac{1}{N^2} h(\tilde{\mu}_N^{(\epsilon)}) + \frac{m}{2} \log N \right).$$

In the last equality, we can replace $\tilde{\mu}_N$ by μ_N as in the proof of Proposition 5.5 because $\tilde{\mu}_N$ concentrates on $\{\|x\|_\infty \le R'\}$. $\qquad \square$

Remark 8.9. The approach given here probably does not give the optimal range of ϵ for Proposition 8.7. To get the best result, one would want a more direct way to extend the potential $V^{(\epsilon)} : \{\|x\|_\infty \le R\} \to \mathbb{R}$ to a potential $\widetilde{V}^{(\epsilon)}$ defined everywhere. This leads us to ask the following question.

Suppose that V is a real-valued function in the closure of trace polynomials with respect to the norm $\|f\|_{\mathcal{T}_m^0, R} + \|Df\|_{\mathcal{T}_m^1, R}$, and hence V defines a function $\{x : \|x\|_\infty \le R\} \to \mathbb{R}$ for $x \in M_N(\mathbb{C})_{\mathrm{sa}}^m$. If $V(x) - (c/2)\|x\|_2^2$ is convex and $V(x) - (C/2)\|x\|_2^2$ is concave on $\{\|x\| \le R\}$, then does V extend to a potential $\widetilde{V} \in \mathcal{E}_m(c, C)^{0,1}$? What if we allow $\widetilde{V}$ to have slightly worse constants c and C?

The construction of extensions that preserve the convexity properties is not difficult, but it is less obvious how to construct an extension that one can verify preserves the approximability by trace polynomials.

Acknowledgements

Note added in proof

Since this paper was first submitted, the author has extended the techniques to cover conditional expectations and entropy in [Jekel 2020a] and in particular obtained an alternative proof of Theorem 7.1.

Moreover, the Ph.D. thesis [Jekel 2020b] contains the results of this paper and [Jekel 2020a] with more detail and historical background.

References

[Anantharaman and Popa 2016] C. Anantharaman and S. Popa, "An introduction to II$_1$ factors", preprint, 2016, available at https://www.math.ucla.edu/~popa/Books/IIun.pdf. To be published by Cambridge University Press.

[Anderson et al. 2010] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics **118**, Cambridge University Press, 2010. MR Zbl

[Barbu 1986] V. Barbu, "A semigroup approach to Hamilton–Jacobi equations in Hilbert space", pp. 9–18 in *Semigroups, theory and applications* (Trieste, 1984), vol. I, edited by M. G. Crandall and F. Kappel, Pitman Res. Notes Math. Ser. **141**, Longman, Harlow, 1986. MR Zbl

[Barron 1986] A. R. Barron, "Entropy and the central limit theorem", *Ann. Probab.* **14**:1 (1986), 336–342. MR Zbl

[Biane 1997] P. Biane, "Segal–Bargmann transform, functional calculus on matrix spaces and the theory of semi-circular and circular systems", *J. Funct. Anal.* **144**:1 (1997), 232–286. MR Zbl

[Biane et al. 2003] P. Biane, M. Capitaine, and A. Guionnet, "Large deviation bounds for matrix Brownian motion", *Invent. Math.* **152**:2 (2003), 433–459. MR Zbl

[Brascamp and Lieb 1976] H. J. Brascamp and E. 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", *J. Functional Analysis* **22**:4 (1976), 366–389. MR Zbl

[Cabanal Duvillard and Guionnet 2001] T. Cabanal Duvillard and A. Guionnet, "Large deviations upper bounds for the laws of matrix-valued processes and non-communicative entropies", *Ann. Probab.* **29**:3 (2001), 1205–1261. MR Zbl

[Capraro 2010] V. Capraro, "A survey on Connes' embedding conjecture", preprint, 2010. arXiv

[Cébron 2013] G. Cébron, "Free convolution operators and free Hall transform", *J. Funct. Anal.* **265**:11 (2013), 2645–2708. MR Zbl

[Crandall et al. 1992] M. G. Crandall, H. Ishii, and P.-L. Lions, "User's guide to viscosity solutions of second order partial differential equations", *Bull. Amer. Math. Soc.* (*N.S.*) **27**:1 (1992), 1–67. MR Zbl

[Dabrowski 2010] Y. Dabrowski, "A non-commutative path space approach to stationary free stochastic differential equations", preprint, 2010. arXiv

[Dabrowski 2016] Y. Dabrowski, "A Laplace principle for Hermitian Brownian motion and free entropy, I: The convex functional case", preprint, 2016. arXiv

[Dabrowski et al. 2016] Y. Dabrowski, A. Guionnet, and D. Shlyakhtenko, "Free transport for convex potentials", preprint, 2016. arXiv

[Driver et al. 2013] B. K. Driver, B. C. Hall, and T. Kemp, "The large-N limit of the Segal–Bargmann transform on $\mathbb{U}_N$", *J. Funct. Anal.* **265**:11 (2013), 2585–2644. MR Zbl

[Ekeland and Lasry 1980] I. Ekeland and J.-M. Lasry, "On the number of periodic trajectories for a Hamiltonian flow on a convex energy surface", *Ann. of Math.* (2) **112**:2 (1980), 283–319. MR Zbl

[Evans 2010] L. C. Evans, *Partial differential equations*, 2nd ed., Graduate Studies in Mathematics **19**, American Mathematical Society, Providence, RI, 2010. MR Zbl

[Guionnet and Maurel-Segala 2006] A. Guionnet and E. Maurel-Segala, "Combinatorial aspects of matrix models", *ALEA Lat. Am. J. Probab. Math. Stat.* **1** (2006), 241–279. MR Zbl

[Guionnet and Shlyakhtenko 2009] A. Guionnet and D. Shlyakhtenko, "Free diffusions and matrix models with strictly convex interaction", *Geom. Funct. Anal.* **18**:6 (2009), 1875–1916. MR Zbl

[Guionnet and Shlyakhtenko 2014] A. Guionnet and D. Shlyakhtenko, "Free monotone transport", *Invent. Math.* **197**:3 (2014), 613–661. MR Zbl

[Hargé 2004] G. Hargé, "A convex/log-concave correlation inequality for Gaussian measure and an application to abstract Wiener spaces", *Probab. Theory Related Fields* **130**:3 (2004), 415–440. MR Zbl

[Jekel 2020a] D. Jekel, "Conditional expectation, entropy, and transport for convex Gibbs laws in free probability", *Int. Mat. Res. Not.* (online publication August 2020).

[Jekel 2020b] D. Jekel, *Evolution equations in non-commutative probability*, Ph.D. thesis, University of California, Los Angeles, 2020, available at https://escholarship.org/uc/item/8n39f7mt.

[Kato 1978] T. Kato, "Trotter's product formula for an arbitrary pair of self-adjoint contraction semigroups", pp. 185–195 in *Topics in functional analysis*, edited by I. Gohberg and M. Kac, Adv. in Math. Suppl. Stud. **3**, Academic Press, New York, 1978. MR Zbl

[Lasry and Lions 1986] J.-M. Lasry and P.-L. Lions, "A remark on regularization in Hilbert spaces", *Israel J. Math.* **55**:3 (1986), 257–266. MR Zbl

[Ledoux 1992] M. Ledoux, "A heat semigroup approach to concentration on the sphere and on a compact Riemannian manifold", *Geom. Funct. Anal.* **2**:2 (1992), 221–224. MR Zbl

[Nica and Speicher 2006] A. Nica and R. Speicher, *Lectures on the combinatorics of free probability*, London Mathematical Society Lecture Note Series **335**, Cambridge University Press, 2006. MR Zbl

[Peller 2006] V. V. Peller, "Multiple operator integrals and higher operator derivatives", *J. Funct. Anal.* **233**:2 (2006), 515–544. MR Zbl

[Pisier and Xu 2003] G. Pisier and Q. Xu, "Non-commutative L^p-spaces", pp. 1459–1517 in *Handbook of the geometry of Banach spaces*, vol. 2, edited by W. B. Johnson and J. Lindenstrauss, North-Holland, Amsterdam, 2003. MR Zbl

[Rains 1997] E. M. Rains, "Combinatorial properties of Brownian motion on the compact classical groups", *J. Theoret. Probab.* **10**:3 (1997), 659–679. MR Zbl

[Razmyslov 1974] J. P. Razmyslov, "Identities with trace in full matrix algebras over a field of characteristic zero", *Izv. Akad. Nauk SSSR Ser. Mat.* **38** (1974), 723–756. In Russian; translated in *Math. USSR-Izv.* **8**:4 (1974), 727–760. MR Zbl

[Razmyslov 1985] Y. P. Razmyslov, "Trace identities and central polynomials in matrix superalgebras $M_{n,k}$", *Mat. Sb.* (*N.S.*) **128(170)**:2 (1985), 194–215. In Russian; translated in *Math. USSR-Sb.* **56**:1 (1987), 187–206. MR Zbl

[Correa da Silva 2018] R. Correa da Silva, "Lecture notes on non-commutative L_p spaces", preprint, 2018. arXiv

[Simon 1979] B. Simon, *Functional integration and quantum physics*, Pure and Applied Mathematics **86**, Academic Press, New York, 1979. MR Zbl

[Simon 2005] B. Simon, *Trace ideals and their applications*, 2nd ed., Mathematical Surveys and Monographs **120**, American Mathematical Society, Providence, RI, 2005. MR Zbl

[Trotter 1959] H. F. Trotter, "On the product of semi-groups of operators", *Proc. Amer. Math. Soc.* **10** (1959), 545–551. MR Zbl

[Voiculescu 1986] D. Voiculescu, "Addition of certain noncommuting random variables", *J. Funct. Anal.* **66**:3 (1986), 323–346. MR Zbl

[Voiculescu 1991] D. Voiculescu, "Limit laws for random matrices and free products", *Invent. Math.* **104**:1 (1991), 201–220. MR Zbl

[Voiculescu 1993] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory, I", *Comm. Math. Phys.* **155**:1 (1993), 71–92. MR Zbl

[Voiculescu 1994] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory, II", *Invent. Math.* **118**:3 (1994), 411–440. MR Zbl

[Voiculescu 1998] D. Voiculescu, "The analogues of entropy and of Fisher's information measure in free probability theory, V: Noncommutative Hilbert transforms", *Invent. Math.* **132**:1 (1998), 189–227. MR Zbl

[Voiculescu 2002] D. Voiculescu, "Free entropy", *Bull. London Math. Soc.* **34**:3 (2002), 257–278. MR Zbl

[Voiculescu et al. 1992] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables*: *a noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups*, CRM Monograph Series **1**, American Mathematical Society, Providence, RI, 1992. MR Zbl

DAVID JEKEL: davidjekel@math.ucla.edu
Department of Mathematics, University of California, Los Angeles, CA, United States

Analysis & PDE

EDITORS

EDITOR-IN-CHIEF

Patrick Gérard

patrick.gerard@math.u-psud.fr

Université Paris Sud XI

Orsay, France

BOARD OF EDITORS

PRODUCTION

ANALYSIS & PDE

Volume 13   No. 8   2020