An Elementary Approach to Free Entropy Theory for Convex Potentials *

David Jekel †

May 24, 2018

Abstract

We present an alternative approach to the theory of free Gibbs states with convex potentials developed in several papers of Guionnet, Shlyakhtenko, and Dabrowski. Instead of solving SDE’s, we combine PDE techniques with a notion of asymptotic approximability by trace polynomials for a sequence of functions on $M_N(C)^{sa}$ to prove the following. Suppose $\mu_N$ is a probability measure on $M_N(C)^{sa}$ given by uniformly convex and semi-concave potentials $V_N$, and suppose that the sequence $DV_N$ is asymptotically approximable by trace polynomials. Then the moments of $\mu_N$ converge to a non-commutative law $\lambda$. Moreover, the free entropies $\chi(\lambda)$, $\chi(\lambda)$, and $\chi^*(\lambda)$ agree and equal the limit of the normalized classical entropies of $\mu_N$. A key step is to show that the property of asymptotic approximation by trace polynomials is preserved under several operations, including limits, composition, Gaussian convolution, and ultimately evolution under certain parabolic PDE. This allows us to prove convergence of the moments of $\mu_N$ and of the Fisher information of Gaussian perturbations of $\mu_N$.

1 Introduction

1.1 Motivation and Main Ideas

Since Voiculescu introduced free entropy of a non-commutative law in [28, 29, 30], a number of open problems have prevented a satisfying unification of the theory (as explained in [31]). The free entropy $\chi$ was defined by taking the lim sup of the log volume of the space of micro states in the algebra of $N \times N$ matrices, and it is unclear whether using the lim inf instead of lim sup would yield the same quantity. Voiculescu also defined a non-microstates free entropy $\chi^*$ 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 [5] showed that $\chi \leq \chi^*$ as a consequence of their large deviation principle for the GUE (see also [13]). The proof relied on stochastic differential equations relative to Hermitian Brownian motion and exponential functionals of Brownian motion. Recent work of Dabrowski [9] 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 non-commutative laws.

---

*MSC 2010 Subj. Class. 46L54
†Department of Mathematics, UCLA. Email: davidjekel@math.ucla.edu
This paper will prove a similar result 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 non-microstates entropy $\chi^\ast$.

Consider a non-commutative law $\lambda$ of an $m$-tuple and suppose $\lambda$ is the limit of a sequence of random $N \times N$ matrix distributions $\mu_N$ given by convex, semi-concave potentials $V_N : M_N(\mathbb{C})^m_{sa} \to \mathbb{R}$. Let $\sigma_{t,N}$ be the distribution of $m$ independent GUE matrices with normalized variance $t$, and let $\kappa_t$ be the non-commutative law of $m$ free semicircular variables of variance $t$. Let $V_{N,t}$ be the potential corresponding to $\mu_N * \sigma_{t,N}$. The classical Fisher information $I$ satisfies

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

and from this we deduce that

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

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

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

where $\Phi^\ast$ is the free Fisher information $[30]$.

Under suitable assumptions on $V_N$, the microstates free entropy $\chi(\lambda)$ is the lim sup of normalized classical entropies of $\mu_N$. On the right hand side, we want to show that $N^{-3} I(\mu_N * \sigma_{t,N}) \to \Phi^\ast(\lambda \boxplus \sigma_t)$ 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 showed that the entrywise Segal-Bargman transform of non-commutative functions evaluated on $N \times N$ matrices can be approximated by the free Segal-Bargman transform computed through analytic functional calculus $[4]$. Similarly, Guionnet and Shlyakhtenko showed that classical monotone transport maps for certain random matrix models approximate the free monotone transport $[17]$ Theorem 4.7]. Moreover, Dabrowski’s approach to prove $\chi = \chi^\ast$ involved constructing solutions to free SDE as ultraproducts of the solutions to classical SDE $[9]$.

In section $3.4$, we make precise the idea that a sequence of functions on $M_N(\mathbb{C})^m_{sa}$ has a “well-defined, non-commutative asymptotic behavior” by defining *asymptotically approximability by trace polynomials* (Definition $3.14$). 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 Gaussian $(3.4$). Then in $(4.4$ we analyze the PDE that describes the evolution of $V_{N,t}$. We show that for all $t$ the solution can be approximated in a dimension-independent way by applying a sequence of simpler operations, each of which
preserves asymptotic approximability by trace polynomials. We conclude that 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 \( \mu_N \) converges to \( \lambda \) 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 \( \mu_N \) to converge in moments to a non-commutative law \( \lambda \), so that Theorem 7.1 can be applied. As a consequence, we show that \( \chi = \chi^* \) for a class of free Gibbs states.

1.2 Main Results

To fix notation, let \( M_N(\mathbb{C})_{sa}^m \) be space of \( m \)-tuples \( x = (x_1, \ldots, x_m) \) of self-adjoint \( N \times N \) matrices and let \( \|x\|_2 = (\sum_j \tau_N(x_j^2))^{1/2} \), where \( \tau_N = (1/N) \text{Tr} \). We denote by \( \|x\| \) 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 non-commutative polynomials in \( x_1, \ldots, x_m \).

Consider a sequence of potentials \( V_N : M_N(\mathbb{C})_{sa}^m \to \mathbb{R} \) 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 \( \mu_N \) by

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

Assume that the sequence of normalized gradients \( DV_N = N\nabla V_N(x) \) 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

\[
\limsup_{N \to \infty} \sup_{\|x\| \leq R} \|DV_N(x) - f(x)\|_2 \leq \epsilon.
\]

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 \( \mu_N \) is unitarily invariant or has expectation zero). In this case, we have the following.

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

2. There exists a non-commutative law \( \lambda \) such that

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

for every non-commutative polynomial \( p \).

3. The measures \( \mu_N \) exhibit exponential concentration around \( \lambda \), in the sense that

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

for every \( R > 0 \) and every non-commutative polynomial \( p \).
(4) The law $\lambda$ has finite free entropy and we have
\[
\chi(\lambda) = \chi^*(\lambda) = \lim_{N \to \infty} \frac{1}{N^2} \left( h(\mu_N) + \frac{m}{2} \log N \right),
\]
where $\chi$ and $\chi^*$ are respectively the lim sup and lim inf versions of microstates free entropy, $\chi^*$ is the non-microstates 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 $\sigma_t$ is the law of $m$ free semicircular variables with variance $t$.

(6) The law $\lambda$ has finite free Fisher information. If $I$ is the classical Fisher information and $\Phi^*$ is the free Fisher information, then
\[
\lim_{N \to \infty} \frac{1}{N^3} I(\mu_N * \sigma_{t,N}) = \Phi^*(\lambda \boxplus \sigma_t).
\]

(7) The functions $t \mapsto \frac{1}{N^3} I(\mu_N * \sigma_{t,N})$ and $t \mapsto \Phi^*(\lambda \boxplus \sigma_t)$ are Lipschitz equicontinuous on any finite time interval.

Here claims (1) through (3) come from Theorem 4.1, which is similar to the earlier results [13, Theorem 4.4], [11, Proposition 50 and Theorem 51], [9, Theorem 4.4] plus standard results on concentration of measure. Claims (4) through (7) come from Theorem 7.1 which is similar to [3, Theorem A].

In particular, we recover Dabrowski’s result [3, Theorem A] that $\chi(\lambda) = \chi(\lambda) = \chi^*(\lambda)$ when the law $\lambda$ is a free Gibbs state given by a sufficiently regular convex non-commutative potential $V(X)$, because taking $V_N = V$ will define a sequence of random matrix models $\mu_N$ which concentrate around the non-commutative law $\lambda$.

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 $L^2$ (and hence being Hölder 1/2 continuous) as shown by Dabrowski. Our results 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 the same formula for every $N$ or that $V_N$ is exactly unitarily invariant.

1.3 Organization of Paper

Section 2 establishes notation and reviews basic facts from non-commutative 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 $M_N(\mathbb{C})_{sa}^m \to M_N(\mathbb{C})_{sa}^m$ or $\mathbb{C}$ 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 4.3 concerning the convergence of moments for the measure $\mu_N$ (claims (1) - (3) of [1,2]). We evaluate $\int u \, d\mu_N$ for a Lipschitz function $u$ as $\lim_{t \to \infty} T^{V_N}_t u$, where $T^{V_N}_t$ is the semigroup such that $u_t = T^{V_N}_t u$ solves the equation $\partial_t u_t = (1/2N)\Delta u_t - (N/2)\nabla V \cdot \nabla u_t$. We approximate $T^{V_N}_t$ 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 $T^{V_N}_t u_N$, 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 $\chi^*(\lambda)$ are the lim sup and lim inf of normalized classical
entropies of $\mu_N$, provided that $\mu_N$ concentrates around $\lambda$ and satisfies some mild operator norm tail bounds, and that $\{V_N\}$ is asymptotically approximable by trace polynomials. Similarly, if $N\nabla V_N$ 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 $\mu_N \ast \sigma_{t, N}$, where $\sigma_{t, N}$ is the law of $m$ independent GUE of variance $t$. Our goal is to show that if $N\nabla V_N(x, 0)$ 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 §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.

Section 7 concludes the proof of our main theorem on free entropy and Fisher’s information (Theorem 7.1), which establishes claims (4) - (7) of §1.2, assuming a weakened version of the hypothesis and conclusion of Theorem 4.1.

In section 8, we characterize the limiting laws $\lambda$ 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.

1.4 Acknowledgements

I thank Timothy Austin, Guillaume Cébron, Yoann Dabrowski, Alice Guionnet, Benjamin Hayes, Dimitri Shlyakhtenko, Terence Tao, and Dan Voiculescu for various useful conversations. I especially thank Shlyakhtenko for his mentorship and ongoing conversations about free entropy, and Dabrowski for detailed discussions of his own results and other recent literature.

I acknowledge the support of the NSF grants DMS-1344970 and DMS-1500035. Part of this research was conducted at the Institute for Pure and Applied Mathematics (IPAM) during the long program on Quantitative Linear Algebra in Spring 2018. IPAM provided hospitality and a stimulating research environment where many of the conversations mentioned above took place.

2 Preliminaries

2.1 Notation for Matrix Algebras

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

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

$$B_N = \{ E_{k,k} \}_{k=1}^N \cup \left\{ \frac{1}{\sqrt{2}} E_{k,\ell} + \frac{1}{\sqrt{2}} E_{\ell,k} \right\}_{k<\ell} \cup \left\{ \frac{i}{\sqrt{2}} E_{k,\ell} - \frac{i}{\sqrt{2}} E_{\ell,k} \right\}_{k<\ell}. \quad (2.1)$$

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

$$\sum_{b \in B_N} xbybz = xz \text{Tr}(y), \quad (2.2)$$
which follows from an elementary computation.

We denote the norm corresponding to $\text{Tr}$ by $|·|$. We denote the normalized trace by $\tau_N = \frac{1}{N} \text{Tr}$. We denote the corresponding inner product by $\langle x, y \rangle_2 = \sum_{j=1}^{N} \tau_N (x_jy_j)$ and the norm by $\|\cdot\|_2$. For $x \in M_N(\mathbb{C})$, we denote the operator norm by $\|x\|$. Similarly, if $x = (x_1, \ldots, x_m) \in M_N(\mathbb{C})^m$, we denote $\|x\| = \max_j \|x_j\|$.

The symbols $\nabla$ and $\Delta$ will represent the gradient and Laplacian operators with respect to the coordinates of $M_N(\mathbb{C})_{sa}$ in the non-normalized inner product $\langle \cdot, \cdot \rangle_{\tau_N}$. 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 notation will be explained and justified in §3.2.

### 2.2 Non-Commutative Probability Spaces and Laws

The following are standard definitions in free probability. For further background, see [32, 22].

**Definition 2.1.** A *von Neumann algebra* is a unital $\mathbb{C}$-algebra $M$ 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 *non-commutative probability space* is a von Neumann algebra $M$ together with a bounded linear map $\tau : M \rightarrow \mathbb{C}$ 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 $\tau$ is called a *trace*.

**Definition 2.3.** For $m \geq 1$, we denote by $\text{NCP}_m = \mathbb{C}\langle X_1, \ldots, X_m \rangle$ the algebra of non-commutative polynomials in $X_1, \ldots, X_m$. A *non-commutative law* (for an $m$-tuple) is a map $\lambda : \text{NCP}_m \rightarrow \mathbb{C}$ such that

1. $\lambda$ is linear,
2. $\lambda$ is unital (that is, $\mu(1) = 1$),
3. $\lambda$ is completely positive, that is, for any matrix $P$ with entries in $\mathbb{C}\langle X_1, \ldots, X_m \rangle$, the matrix $\lambda(P^*P)$ is positive semi-definite.
4. $\lambda$ is tracial, that is, $\lambda(p(X)q(X)) = \lambda(q(X)p(X))$.

We denote by $\Sigma_m$ the space of non-commutative laws equipped with the topology of pointwise convergence on $\mathbb{C}\langle X_1, \ldots, X_m \rangle$, that is, convergence in non-commutative moments.

**Definition 2.4.** We say that a non-commutative law $\lambda$ is *bounded by $R$* if we have

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

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 $(M, \tau)$. Then the *law of* $x = (x_1, \ldots, x_m)$ is the map

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

**Definition 2.6.** Let $M_N(\mathbb{C})$ be the algebra of $N \times N$ matrices over $\mathbb{C}$. Let $\tau_N = \frac{1}{N} \text{Tr}$ be the normalized trace. Then $(M_N(\mathbb{C}), \tau_N)$ is a tracial von Neumann algebra, and hence for any tuple of self-adjoint matrices $x = (x_1, \ldots, x_m)$, the law $\lambda_x$ 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 $\lambda_x$ for some tuple $x = (x_1, \ldots, x_m)$ of self-adjoint elements of a tracial von Neumann algebra $(M, \tau)$ with $\|x_j\| \leq R$. 

6
2.3 Free Independence, Semicircular Law, and GUE

Recall that self-adjoint elements $X_1, \ldots, X_m$ of a tracial von Neumann algebra $(M, \tau)$ 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 $	au(f_j(X_{i_j}) = 0$, we have also $\tau(f_1(X_{i_1}) \ldots f_k(X_{i_k})) = 0$.

If $X_1, \ldots, X_m$ are freely independent, then their joint law is determined by the individual laws of the $X_j$'s, each of which is represented by a compactly supported probability measure on $\mathbb{R}$. The *semicircle law* (mean zero and variance 1) is the probability measure given by density
\[
(1/2\pi)\sqrt{1-x^2}1_{[-2,2]}(x)\]
We denote by $\sigma_t$ the non-commutative law of $m$ freely independent semicircular random variables of mean zero and variance $t$.

We denote by $\sigma_{t,N}$ be the probability distribution on $M_N(\mathbb{C})_{sa}^m$ for $m$ independent GUE matrices of normalized variance $t$, that is,
\[
d\sigma_{t,N}(x) = \frac{1}{Z_{N,t}} \exp \left(-N \sum_{j=1}^{m} \text{Tr}(x_j^2)/2t \right) \, dx,
\]
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 limit like freely independent semicircular random variables, although we shall prove the specific properties we use in §3. For further background, refer to [1], [27, 32, 22, 1].

2.4 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 [1, §4.4.2]), although it can also be proved directly using the heat semigroup directly as in [21]. We state the theorem here with free probabilistic normalizations.

**Theorem 2.8.** Suppose that $V : M_N(\mathbb{C})_{sa}^m \to \mathbb{R}$ 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 $f : M_N(\mathbb{C})_{sa}^m \to \mathbb{R}$ is $L$-Lipschitz with respect to $||\cdot||_2$. Then
\[
\mu(x : f(x) - \int f \, d\mu \geq \delta) \leq e^{-cN^2\delta^2/2L^2}.
\]

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 [13, Theorem 1.1] and the application to random matrix models is taken from the proof of [13, Theorem 3.4].

**Theorem 2.9.** Let $V$ and $\mu$ be as in Theorem 2.8, and suppose that $f : M_N(\mathbb{C})_{sa}^m \to \mathbb{R}$ 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, let $||x||_p = \left(\tau_N(|x|^p)\right)^{1/p}$ and $||x||_\infty$ be the operator norm. Then for every $\alpha \geq 1$ and $p \in [1, +\infty]$, we have
\[
\int ||x_j - a_j||_p^\alpha \, d\mu(x) \leq \int ||y_j||_p^\alpha \, d\sigma_{c^{-1},N}(y).
\]
Proof. The convexity assumption on $V$ means that $\mu$ has a log-concave density with respect to the Gaussian measure $\sigma_{c^{-1},N}(y)$. Therefore, the first claim follows from Hargé’s theorem [18, Theorem 1.1]. The second claim follows because norms on vector spaces are convex functions.

Corollary 2.10. Let $V_N : M_N(\mathbb{C})_{sa} \to \mathbb{R}$ be a function such that $V_N(x) - (c/2)\|x\|_2^2$ is convex and let $\mu_N$ 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

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

Proof. For the first claim, it suffices to check the case of $\sigma_{c^{-1},N}$. This is a standard result. See for instance the proof of [1, Theorem 2.1.22]. The second claim follows from Theorem 2.8 after we observe that $\|x_j\|$ is $N^{1/2}$-Lipschitz with respect to $\|\cdot\|_2$.

3 Trace Polynomials

3.1 Definition

We define the algebra of trace polynomials $\text{TrP}_m^0$ as follows. Let $V$ be the vector space $\text{NCP}_m / \text{Span}(pq - qp : p,q \in \text{NCP}_m)$. We define the vector space

$$\text{TrP}_m^0 = \bigoplus_{n=0}^{\infty} V^\otimes n,$$

(3.1)

where $\otimes$ is the symmetric tensor power over $\mathbb{C}$. Then $\text{TrP}_m^0$ forms a commutative algebra with the tensor operator $\otimes$ as the multiplication. We denote the element $p_1 \otimes \cdots \otimes p_n$ by $\tau(p_1) \cdots \tau(p_n)$ where $\tau$ is a formula symbol.

To state the definition more suggestively, an element of $\text{TrP}_m^0$ 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 non-commutative polynomials in $X_1, \ldots, X_m$ and $\tau$ 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 $\text{TrP}_m^0$ over $\mathbb{C}$ where the $*$-operation is

$$(\tau(p_1(X)) \cdots \tau(p_n(X))^\ast = \tau(p_1(X))^\ast \cdots \tau(p_n(X))^\ast$$

(3.2)

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

We define $\text{TrP}_m^k$ to be the vector space

$$\text{TrP}_m^k := \text{TrP}_m^0 \otimes \text{C}(X_1, \ldots, X_m)^\otimes k.$$ 

We call the elements of $\text{TrP}_m^1$ non-commutative trace polynomials. Note that $\text{TrP}_m^1$ forms a $*$-algebra because it is the tensor products of two $*$-algebras.

Suppose that $M$ is a von Neumann algebra with trace $\sigma$. Given $f \in \text{TrP}_m^1$ and a self-adjoint tuple $x = (x_1, \ldots, x_m)$ of elements of $M$, we define $f(x)$ to the element of $M$ given by formally evaluating $f$ with $X = x$ and $\tau = \sigma$. For instance if $f(X) = p_0(X) \otimes \tau(p_1(X)) \cdots \tau(p_n(X))$ in $\text{TrP}_m^1$, 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 \( \tau = \tau_N \). Similarly, we can evaluate a non-commutative law \( \lambda \) on any trace polynomial.

Moreover, there is a composition operation \((\text{TrP}_m^0)^m \times (\text{TrP}_m^1)^m \to (\text{TrP}_m^1)^m\) defined just as one would expect from manipulations in \( M_N(\mathbb{C}) \). Namely, we multiply elements out by treating the terms \( \tau(p) \) like scalars.

### 3.2 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 \( N \to \infty \). The results in \[3.2\] and \[3.3\] are standard and related to \[7\] and \[12, \S 3\]. We first recall the non-commutative difference operators of Voiculescu.

**Definition 3.1.** We define the **non-commutative derivative** \( D_j : \text{NCP}_m \to \text{NCP}_m \otimes \text{NCP}_m \) by

\[
D_j[X_{i_1} \ldots X_{i_n}] = \sum_{k: i_k = j} X_{i_{k-1}} X_{i_{k+1}} \ldots X_{i_n}.
\]

We also define \( D_j : \text{NCP}_m \otimes \text{n} \to \text{NCP}_m \otimes \text{n+1} \) by

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

Then of course \( D_j^k \) is a well-defined map \( \text{NCP}_m \otimes \text{n} \to \text{NCP}_m \otimes \text{n+k} \).

**Remark 3.2.** We caution the reader that the normalization for \( D_j^n f \) here differs from that of \[30\] by a factor of \( n! \).

**Definition 3.3.** We define the **cyclic derivative** \( D^\circ_j : \text{NCP}_m \to \text{NCP}_m \) as the linear map given by

\[
D^\circ_j[X_{i_1} \ldots X_{i_n}] = \sum_{k: i_k = j} X_{i_{k+1}} \ldots X_{i_n} X_{i_1} \ldots X_{i_{k-1}}.
\]

**Definition 3.4.** Given an algebra \( \mathcal{A} \) (e.g. \( \text{NCP}_m \)), we define the **hash operation** as the bilinear map \( \mathcal{A}^\otimes n+1 \times \mathcal{A}^\otimes n \) given by

\[
(a_0 \otimes \cdots \otimes a_n) \# (b_1 \otimes \cdots \otimes b_n) = a_0 b_1 a_1 \ldots b_n a_n.
\]

Now we can define derivatives for scalar-valued and non-commutative trace polynomials that correspond with differentiation on \( \mathbb{R}^{mN^2} \) using the coordinates given in \[2.1\]. We begin with the gradient.

**Definition 3.5.** Define the \( j \)th differentiation operator \( \text{TrP}_m^0 \to \text{TrP}_m^1 \) by

\[
D_j \left[ \prod_{k=1}^n \tau(p_k) \right] = \sum_{k=1}^n D^\circ_j p_k \prod_{\ell \neq k} \tau(p_\ell).
\]

**Lemma 3.6.** For \( M_N(\mathbb{C})^m_{sa} \), let \( \nabla_j \) denote the gradient with respect to the coordinates of \( x_j \). If \( f \in \text{TrP}_m^0 \) is viewed as a function \( M_N(\mathbb{C})^m_{sa} \to M_N(\mathbb{C})^m = \mathbb{C} \otimes \mathbb{R} M_N(\mathbb{C})^m_{sa} \), then we have

\[
\nabla_j[f(x)] = \frac{1}{N} [D_j f](x).
\]
Similarly, for \( F : M_N(\mathbb{C})^{m}_{sa} \to M_N(\mathbb{C})^{m} \), let \( J_j F \) denote the Jacobian linear transformation with respect to the coordinates of \( x_j \). Then for a non-commutative polynomial \( p \), we have

\[
J_j p(x)[y] = D_j p(x)y, \tag{3.5}
\]

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

\[
J_j[p(x)f(x)][y] = (D_j p(x)y)f(x) + p(x) \cdot \tau_N(D_j f(x)y). \tag{3.6}
\]

**Proof.** By standard computations, for a non-commutative polynomial \( p \) and \( y \in M_N(\mathbb{C})_{sa} \), we have

\[
J_j p(x)[y] = D_j p(x)y
\]

\[
\nabla_j [\tau_N(p(x))] = \frac{1}{N}D_j^p p(x).
\]

The claims (3.4) and (3.6) now follow from the product rule. \( \square \)

Now we can define the Laplacian on trace polynomials, starting with the case of \( \text{TrP}_m^0 \). Motivated by (2.2) and the computation in Lemma 3.9 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 3.7.** 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

\[
L_j[\tau(p)] = L_{N,j}[\tau(p)] = \tau \circ \eta[D_j^2 p] \text{ for } p \in \text{NCP}_m. \tag{3.7}
\]

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 \( L_N = \sum_{j=1}^{m} L_{N,j} \).

**Definition 3.8.** We also define \( L_j \) and \( L_{N,j} : \text{TrP}_m^1 \to \text{TrP}_m^1 \) to be the unique linear operators such that

\[
L_j[p] = L_{N,j}[p] = \eta[D_j^2 p] \text{ for } p \in \text{NCP}_m \tag{3.10}
\]

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

\[
L_j[p \cdot f] = L_j[p] \cdot f + p \cdot L_j[f] \tag{3.11}
\]

\[
L_{N,j}[p \cdot f] = L_{N,j}[p] \cdot f + p \cdot L_{N,j}[f] + \frac{2}{N^2} D_j p \# D_j f. \tag{3.12}
\]

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

**Lemma 3.9.** Let \( f \in \text{TrP}_m^0 \). For functions on \( M_N(\mathbb{C})^{m}_{sa} \), let \( \Delta_j \) be the Laplacian with respect to the coordinates of the \( j \)-th matrix \( x_j \). Viewing \( f \) is a function \( M_N(\mathbb{C})^{m}_{sa} \to \mathbb{C} \), we have

\[
\Delta_j f(x) = N[L_{N,j} f](x) \quad \Delta f(x) = N[L_N f](x). \tag{3.13}
\]

10
Lemma 3.12. Together with Corollary 3.10 yield the following.

\[ \exp(0 \text{formations TrP}) \]

The space of trace polynomials of degree \( \leq 1 \) can verify that \( \partial \) the heat equation

\[ \Delta \]

Let \( \eta \) the operator \( L \) of \( \sigma \)

Therefore, in the rest of the paper, we will freely write

\[ L \]

Corollary 3.10. Let \( f \in \text{TrP}_m^0 \) or \( \text{TrP}_m^1 \). If we view \( f \) as a function on \( M_N(\mathbb{C})_{sa}^m \), then \( \Delta f \) is a trace polynomial of lower degree than \( f \), and we have coefficient-wise

\[ \lim_{N \to \infty} \frac{1}{N} \Delta f(x) = \lim_{N \to \infty} L_N f(x) = L f(x). \]

Remark 3.11. We have shown that if \( f \) is a scalar-valued trace polynomial, then viewed as a map \( M_N(\mathbb{C})_{sa}^m \to \mathbb{C} \), we have

\[ Du = N \nabla u, \quad L_N u = \frac{1}{N} \Delta u. \]

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 \( f : M_N(\mathbb{C})_{sa}^m \to \mathbb{C} \).

3.3 Convolution of Trace Polynomials and Gaussians

If \( f \) is a trace polynomial in \( \text{TrP}_m^0 \) or \( \text{TrP}_m^1 \) with degree \( d \), then the function \( f_t = f \ast \sigma_{t,N} \) satisfies the heat equation \( \partial_t f_t = \frac{1}{2N} \Delta f_t \), where the normalization \( 1/2N \) comes from the normalization of \( \sigma_{t,N} \).

We know that \( L_N = \frac{1}{N} \Delta \) on trace polynomials is given by a purely combinatorial computation. The operator \( L_N \) 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 \( \text{TrP}_m^0 \to \text{TrP}_m^0 \) and \( \text{TrP}_m^1 \to \text{TrP}_m^1 \). Moreover, a standard computation shows that \( \exp(tL_N/2)f \) satisfies the heat equation for the normalized Laplacian \( L_N \). These observations, together with Corollary 3.10 yield the following.

Lemma 3.12. Let \( f \) be a trace polynomial in \( \text{TrP}_m^0 \) or \( \text{TrP}_m^1 \). Then we have

\[ \sigma_{t,N} \ast f(x) = [\exp(tL_N/2)f](x), \quad (3.14) \]

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

\[ \lim_{N \to \infty} \exp(tL_N/2)f = \exp(tL/2)f \text{ coefficient-wise.} \quad (3.15) \]
For a trace polynomial $f$, we know that $\sigma_{t,N} * f(x) = \exp(tL_N/2)$ is the expectation of $f(x + t^{1/2}Y)$, where $Y$ is an $m$-tuple of independent GUE of variance 1. Moreover, for any probability measure $\mu$ on $M_N(\mathbb{C})_{sa}$ with finite moments, we have

$$\int f(x) d(\mu * \sigma_{t,N})(x) = \int (\sigma_{t,N} * f)(x) d\mu(x) = \int [\exp(tL_N/2)f](x) d\mu(x).$$

(3.16)

The operator $\exp(tL/2)$ has a similar relationship with the free semicircular family, a fact which we will need for Lemmas 3.18 and 7.3 below.

**Lemma 3.13.** Let $\lambda$ be a non-commutative law with $\text{rad}(\lambda) < +\infty$. Then for any trace polynomial $f \in \text{TrP}_m^0$, we have

$$\lambda \boxplus \sigma_t(f) = \lambda(\exp(tL/2)f).$$

(3.17)

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

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

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

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

Let $X = (X_1, \ldots, X_m)$ be a random variable with law $\lambda$ and let $S = (S_1, \ldots, S_m)$ be a freely independent tuple of semi-circulars realized together in a von Neumann algebra $(M, \tau)$. We want to compute $\frac{d}{dt} \bigg|_{t=0} \tau(p(X + t^{1/2}S))$. But note that

$$p(X + t^{1/2}S) = p(X) + t^{1/2} \sum_{j=1}^m D_j p(X) \# S_j + \frac{1}{2} t \sum_{j,k=1}^m D_j D_k p(X) \# (S_j \otimes S_k) + O(t^{3/2})$$

A moment computation with free independence shows that the first order terms have expectation zero, and so do the second order terms with $j \neq k$. We are left with

$$\frac{d}{dt} \bigg|_{t=0} \tau(p(X + t^{1/2}S)) = \frac{1}{2} \sum_{j=1}^m \tau(D^2_j p(X) \# (S_j \otimes S_j)),$$

which using freeness evaluates to $(1/2) \sum_{j=1}^m \tau(\eta D^2_j p(X)) = \tau((1/2)Lp(X)).$ \hfill $\Box$

### 3.4 Asymptotic Approximation by Trace Polynomials

Now we are ready to define the approximation property which captures the asymptotic behavior of functions on $M_N(\mathbb{C})_{sa}^N$. This is one of the key technical tools in our proof.

**Definition 3.14.** A sequence of functions $\phi_N : M_N(\mathbb{C})_{sa}^m \to M_N(\mathbb{C})^m$ is said to be asymptotically approximable by trace polynomials if for every $\epsilon > 0$ and $R > 0$, there exists a trace polynomial $f \in \text{TrP}_m^1$ such that

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

In this case, we call $f$ an $(\epsilon, R)$-approximation of $\{\phi_N\}$. We make the same definitions for functions $\phi_N : M_N(\mathbb{C})_{sa}^m \to \mathbb{C}$, except that we use scalar-valued trace polynomials (elements of $\text{TrP}_m^0$) and apply the absolute value rather than the 2-norm.
Observation 3.15. If $f_N$ is given exactly by a trace polynomial $p$ for $\|x\| \leq R$, then $f_N$ is asymptotically approximable by trace polynomials. Also, asymptotically approximable sequences form a vector space over $\mathbb{C}$.

Observation 3.16. Let $\{\phi_{N,\ell}\}_{N,\ell \in \mathbb{N}}$ be a sequence of functions where $\phi_{N,\ell} : M_N(\mathbb{C})_{sa} \rightarrow M_N(\mathbb{C})^m$. Suppose that $\{\phi_N\}$ is another sequence such that for every $R > 0$,

$$\lim_{\ell \to \infty} \lim_{N \to \infty} \sup_{\|x\| \leq R} \|\phi_{N,\ell}(x) - \phi_N(x)\|_2 = 0.$$ 

If $\{\phi_{N,\ell}\}_{N \in \mathbb{N}}$ is asymptotically approximable by trace polynomials for each $\ell$, then so is $\{\phi_N\}_{N \in \mathbb{N}}$.

Lemma 3.17. Let $\phi_N, \psi_N : M_N(\mathbb{C})_{sa}^m \rightarrow M_N(\mathbb{C})_{sa}^m$. Suppose that $\{\phi_N\}$ and $\{\psi_N\}$ are both asymptotically approximable by trace polynomials, and furthermore suppose that $\{\phi_N\}_{N \in \mathbb{N}}$ in uniformly Lipschitz in $\|\cdot\|_2$, that is, for some $L > 0$,

$$\|\phi_N(x) - \phi_N(y)\|_2 \leq L \|x - y\|_2 \text{ for all } x, y, \text{ for all } N.$$ 

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

Proof. Choose $\epsilon > 0$ and $R > 0$. Choose a trace polynomial $g$ which is an $(\epsilon/2L, R)$-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

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

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

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

Therefore,

$$\lim_{N \to \infty} \sup_{\|x\| \leq R} \|\phi_N \circ \psi_N(x) - f \circ g(x)\|_2 \leq L \cdot \frac{\epsilon}{2L} + \frac{\epsilon}{2} = \epsilon. \quad \Box$$

Lemma 3.18. Suppose that $\phi_N : M_N(\mathbb{C})_{sa}^m \rightarrow M_N(\mathbb{C})_{sa}^m$ is asymptotically approximable by trace polynomials and that

$$\|\phi_N(x)\|_2 \leq A \left(1 + \sum_j \tau_N(x_j^{2n}) \right) \quad (3.18)$$

for some $A > 0$ and 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 $\|x\| \leq R$, we estimate

$$\|\sigma_{t,N} * \phi_k(x) - \sigma_{t,N} * f(x)\|_2 \leq \int \|\phi_k(x + y) - f(x + y)\|_2 d\sigma_{t,N}(y).$$
We break this integral into two pieces: The integral over the region where \( \|y\| \leq 3t^{1/2} \) is bounded by \( \epsilon \) as \( N \to \infty \) by our choice of \( f \). Furthermore, we claim that the integral over the region where \( \|y\| > 3t^{1/2} \) 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

\[
\sup_{\|x\| \leq R} \|\phi_k(x + y)\|_2 + \|f(x + y)\|_2 \leq C \left( 1 + \sum_j \tau_N(y^{2d}_j) \right).
\]

Therefore, we have

\[
\int_{\|y\| \geq 3t^{1/2}} \|\phi_k(x + y) - f(x + y)\|_2 \, d\sigma_{t,N}(y) \leq C \int_{\|y\| \geq 3t^{1/2}} \left( 1 + \sum_j \tau_N(y^{2d}_j) \right) \, d\sigma_{t,N}(y).
\]

This vanishes as \( N \to \infty \) by Corollary 2.10 applied to the GUE. Therefore, we have

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

On the other hand, by Lemma 3.12 we have \( \sigma_{t,N} * f = \exp(tL_N/2)f \to \exp(tL/2)f \) coefficient-wise, and therefore,

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

so that

\[
\limsup_{N \to \infty} \sup_{\|x\| \leq R} \|\sigma_{t,N} * \phi_k(x) - [\exp(tL/2)f](x)\|_2 \leq \epsilon. \quad \square
\]

**Lemma 3.19.** Suppose that \( \phi_N : M_N(C)^m_{sa} \to C \) 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 \( F \in (\text{TrP}^1_m)^m \), we can define

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

in \( \text{TrP}_m^0 \). Then we have

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

\[
\leq R \sup_{\|y\| \leq R} \|N\nabla \phi_N(y) - F(y)\|_2. \quad \square
\]

## 4 Convergence of Moments

Our goal in this section is to prove the following theorem. The convergence of moments is related to [16, Theorem 4.4], [11, Proposition 50 and Theorem 51], [9, Theorem 4.4], and we include versions of standard concentration estimates in the statement.
Theorem 4.1. Let \( V_N : M_N(\mathbb{C})_sa^m \to \mathbb{R} \) 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 \( \mu_N \) be the associated measure. Suppose that the sequence \( \{DV_N\} \) is asymptotically approximable by trace polynomials, and assume that

\[
M = \limsup_{N \to \infty} \max_j \left\| \int (x_j - \tau_N(x_j)) \, d\mu_N(x) \right\| < +\infty. \tag{4.1}
\]

(1) We have the following bounds on the operator norm. If \( R_N = \max_j \int \|x_j\| \, d\mu_N(x) \), then

\[
\limsup_{N \to \infty} R_N \leq \frac{2}{c^{1/2}} + \frac{1}{c} \limsup_{N \to \infty} \left( \int \tau_N(x_j) \, d\mu(x) \right) + M
\]

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

and as a consequence of concentration we have

\[
\mu_N(\|x\| \geq R_N + \delta) \leq e^{-cN\delta^2/2}.
\]

(2) There exists a non-commutative law \( \lambda \) such that for every non-commutative 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 \( \lambda \) in the sense that for every \( R > 0 \), and every neighborhood \( U \) of \( \lambda \) in \( \Sigma_m \),

\[
\limsup_{N \to \infty} \frac{1}{N^2} \log \mu_N(\|x\| \leq R, \lambda_x \not\in U) < 0.
\]

Remark 4.2. The rather artificial hypothesis that \( \max_j \sup_N \left\| \int (x_j - \tau_N(x_j)) \, d\mu_k(x) \right\| < +\infty \) is trivially satisfied if either (1) \( \mu_N \) has expectation zero or (2) \( \mu_N \) is invariant under unitary conjugation and hence \( \int x_j \, d\mu_N(x) = \int \tau(x_j) \, d\mu_N(x) \).

We have already seen in (2.4) 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 \( \mu_N \) 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 different size matrices.

To prove convergence of moments, we want to express \( \int ud\mu_N \) in terms of \( DV_N \). One of the standard techniques is to show that \( \mu_N \) is the unique stationary distribution for a process \( X_t \) that satisfies the SDE

\[
dX_t = dY_t - \frac{1}{2}DV_N(X_t) \, dt,
\]

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

Specifically, Dabrowski, Guionnet, and Shylakhtenko used the free version of this SDE to show that for a non-commutative potential \( V \) satisfying certain convexity assumptions, there exists a free Gibbs law for \( V \) which is the unique stationary distribution [11 Proposition 5]. As
an application, they show convergence of moments for random matrix models given by $V_N = V_{11}$, Proposition 50 and Theorem 51], essentially a special case of our Theorem 4.1.

Dabrowski 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 [9, Theorem 4.4]. Our theorem has similar convexity assumptions to Dabrowski’s, but we consider a more general sequence of potentials $V_N$. We also perform most of our analysis in the finite-dimensional setting, but we use deterministic rather than stochastic methods.

Instead of solving the SDE, we study the associated semigroup $T_{t/VN}$, acting on Lipschitz functions $u$, given by

$$T_{t/VN} u(x) = E_x[u(X_t)],$$

where $X_t$ is the process solving the SDE 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

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

The semigroup $T_{t/VN}$ will decrease the Lipschitz norms of functions and thus, if $u$ is Lipschitz, then $T_{t/VN} u$ 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 (§4.1), and then we will show (Lemma 4.9) that the iteration procedure preserves approximability by trace polynomials and hence conclude that $\lim_{N \to \infty} \int u \, d\mu_N$ exists.

### 4.1 Iterative Construction of the Semigroup

To simplify notation in this section, we fix $N$ and fix a potential $V : M_N(\mathbb{C})^m \to \mathbb{R}$ such that $V(x) - (c/2)\|x\|^2$ is convex and $V(x) - (C/2)\|x\|^2$ is concave, and we 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 $dY_t = (1/2)DV(Y_t) \, dt$. Recall that the solution to the heat equation $\partial_t u = (1/2N)\Delta u$ 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 $\partial_t u = -(1/2)\langle DV, Du \rangle$ is given by

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

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

$$\partial_t W(x,t) = -\frac{1}{2} DV(W(x,t)) \quad W(x,0) = x. \quad (4.2)$$

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 [22, 19, 24, p. 4 - 6]). But of course, we must show that $(P_{t/n} S_{t/n})^n$ converges and derive dimension-independent error bounds.

We use the following basic properties of the semigroup $S_t$; here $\|u\|_{\text{Lip}}$ denotes the Lipschitz norm with respect to the normalized $L^2$ metric $\| \cdot \|_2$ on $M_N(\mathbb{C})^m$.

**Lemma 4.3.**
(1) The solution \( W(x,t) \) to (1.2) 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) \( \|(W(x,t) - x) - (W(y,t) - y)\|_2 \leq \frac{C}{L}(1 - e^{-ct/2})\|x-y\|_2 \).

(5) \( |S_{tu}|_{\text{lip}} \leq e^{-ct/2}\|u\|_{\text{lip}} \).

**Proof.** (1) The convexity assumptions on \( V \) imply that \( DV \) is \( C \)-Lipschitz and therefore global existence of the solution follows from Picard-Lindelöf.

(2) Let \( \tilde{V}(x) = V(x) - (c/2)\|x\|^2 \). Because \( \tilde{V} \) is convex, we have

\[
\langle D\tilde{V}(x) - D\tilde{V}(y), x-y \rangle_2 \geq 0,
\]

which translates to

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

Now observe that

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

and hence by Gronwall’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

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

Meanwhile, \( \|W(x,t) - x\|_2 \) is Lipschitz 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

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

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) follows from (2).
Thus, we want to estimate $S_t$. We want to show that the sequence $T_{t,\ell}$ exists and we have

$$\lim_{\ell \to \infty} T_{t,\ell} u = (P_{2^{-\ell}S_2}u)^{2^\ell}u.$$

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 $t/n$.

**Lemma 4.4.** For dyadic $t \in 2^{-\ell}\mathbb{N}$, define

$$T_{t,\ell} u = (P_{2^{-\ell}S_2})^{2^\ell}u.$$

Then $T_{t,\ell} u := \lim_{\ell \to \infty} T_{t,\ell} u$ exists and we have

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

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

$$T_{t,\ell+1} u - T_{t,\ell} u = \sum_{j=0}^{n-1} (P_{2^jS_2}P_{2^{j+1}S_2})^n u_{n-j} u.$$

Thus, we want to estimate $S_2P_2 - S_2S_2$ and then control the propagation of the errors through the applications of the other operators. Note that for a Lipschitz function $v$, we have

$$|S_2P_2 v(x) - S_2S_2 v(x)| \leq \int |v(W(x, \delta) + y) - v(W(x + y, \delta))| d\sigma_{\delta,N}(y)$$

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

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

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

and therefore,

$$\|S_2P_2 v - S_2S_2 v\|_{L^\infty} \leq \frac{C}{c} \delta^{1/2} (1 - e^{-\delta/2}) \|v\|_{Lip}.$$

Now we apply this to the telescoping series identity together with the fact that $P_t$ and $S_t$ are contractions in $L^\infty$ to obtain

$$\|T_{t,\ell+1} u - T_{t,\ell} u\|_{L^\infty} \leq \sum_{j=0}^{n-1} \frac{C}{c} \delta^{1/2} (1 - e^{-\delta/2}) \|S_2(P_{2^jS_2})^{n-1-j} u\|_{Lip}$$

$$\leq \sum_{j=0}^{n-1} \frac{C}{c} \delta^{1/2} (1 - e^{-\delta/2}) e^{-\delta/2} e^{-\delta/2} e^{-\delta(n-j-1)} \|u\|_{Lip}$$

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

$$\leq \frac{C}{c} \delta^{1/2} e^{-\delta/2} \|u\|_{Lip} \leq \frac{C}{2c} \delta^{1/2} \|u\|_{Lip}.$$

In other words, we have $\|T_{t,\ell+1} u - T_{t,\ell} u\|_{L^\infty} \leq \frac{C}{2c} 2^{-(\ell+1)/2} \|u\|_{Lip}$. It follows that the sequence is Cauchy and we have the desired estimate from summing the geometric series. \qed
Lemma 4.5. The semigroup $T_t$ defined above extends to a semigroup defined for positive $t$ such that for $s \leq t$,

$$|T_t u(x) - T_s u(x)| \leq e^{-cs/2} \left( \frac{C}{c} (6 + 5\sqrt{2}) (t - s)^{1/2} + \|V(x)\|_2 (t - s) \right) \|u\|_{\text{Lip}},$$

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

Proof. We first prove the estimate on $|T_t u - T_s u|$ for dyadic values of $t$ and $s$. In the general case, let us write $t > s$ in a binary expansion

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

where $a_j \in \{0, 1\}$ and $a_{n+1} = 1$, and let $t_k = s + \sum_{j=n+1}^{\infty} a_j 2^{-j}$. Then

$$|T_t u(x) - T_s u(x)| \leq \sum_{j=n+1}^{\infty} |T_{t_k} u(x) - T_{t_j-1} u(x)|$$

$$\leq \sum_{j=n+1}^{\infty} \left( \frac{C}{c} (2 - 2^{j/2}) 2^{-j/2} + 2^{-j/2} + \frac{2^{-j}}{2} \|V(x)\|_2 \right) \|T_{t_j-1} u\|_{\text{Lip}}$$

$$\leq \left( \left( \frac{C}{c} (2 - 2^{1/2}) + 1 \right) \frac{2^{1/2}}{1 - 2^{-1/2}} \cdot 2^{-(n+1)/2} + \|V(x)\|_2 \cdot 2^{-n-1} \right) \|T_{t_k} u\|_{\text{Lip}}$$

$$\leq e^{-cs/2} \left( \frac{C}{c} (6 + 5\sqrt{2}) (t - s)^{1/2} + \|V(x)\|_2 (t - s) \right) \|u\|_{\text{Lip}}.$$
The first term can be estimated by
\[|(T_{s_n} - T_s)T_t u(x)| \leq e^{-s/2} \left( \frac{C}{c} (6 + 5\sqrt{2})(s_n - s)^{1/2} + \|V(x)\|_2 (s_n - s) \right) \|T_t u\|_{\text{Lip}},\]
which goes to zero as \(n \to \infty\). For the second term, we first note that
\[|(T_{t_n} - T_t)u(x)| \leq e^{-t/2} \left( \frac{C}{c} (6 + 5\sqrt{2})(t_n - t)^{1/2} + \|V(x)\|_2 (t_n - t) \right) \|u\|_{\text{Lip}}\]
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 \(S_s\). 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 a \(e^{-t/2}(t_n - t)C\|u\|_{\text{Lip}}\) Lipschitz function and hence
\[|T_s h(x)| \leq h_n(x) + |(T_s - 1)h_n(x)|\]
\[\leq h_n(x) + e^{-t/2}(t_n - t)C\|u\|_{\text{Lip}} \left( \frac{C}{c} (6 + 5\sqrt{2})s^{1/2} + \|V(x)\|_2 s \right),\]
which goes to zero as \(n \to \infty\). □

**Lemma 4.6.** 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 \(\phi \in C^\infty_c(M_N(\mathbb{C})^m)\), we have

\[\int_{M_N(\mathbb{C})^m} [T_{t_n} u \phi - T_t u \phi] = \int_{t_0}^{t_1} \int_{M_N(\mathbb{C})^m} \left[ -\frac{1}{2N} \nabla(T_s u) \cdot \nabla \phi - \frac{N}{2} (\nabla V \cdot \nabla(T_s u)) \phi \right] ds.\]

**Proof.** Recall by Rademacher’s theorem if \(u\) is Lipschitz, then \(\nabla u\) exists almost everywhere and it is in \(L^\infty\). Moreover, because \(\nabla V\) is Lipschitz, we also know that the second derivatives of \(V\) exist almost everywhere and are in \(L^\infty\).

We begin by considering \(\int (S_\delta P_\delta - 1) u \cdot \phi\) for a Lipschitz \(u_0 : M_N(\mathbb{C})^m \to \mathbb{R}\) and a \(\phi \in C^\infty_c(M_N(\mathbb{C})^m)\) 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 \(L^\infty\). (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 Lemma 4.3, we have \(W(x, \delta) - x = \frac{N\delta}{2} \nabla V(x) + O(\delta^2)\) uniformly on any compact set \(K\). Therefore,
\[(S_\delta - 1) P_\delta u(x) = P_\delta u(W(x, \delta)) - P_\delta u(x) = -\frac{N\delta}{2} \nabla(P_\delta u)(x) \cdot \nabla V(x) + O_K(\delta^{3/2}).\]
Now we have
\[
\int (S_\delta P_\delta - 1)u \cdot \phi = \int (S_\delta - 1)P_\delta u \phi + \int (P_\delta - 1)u \phi \\
= -\frac{N\delta}{2} \int [\nabla (P_\delta u) \cdot \nabla \phi] + \int (P_\delta - 1)u \phi + O(\delta^{3/2}) \\
= -\frac{N\delta}{2} \int P_\delta |\Delta V \phi + \nabla V \cdot \nabla \phi| + \int \frac{\delta}{2N} \Delta \phi + O(\delta^{3/2}) \\
= -\frac{N\delta}{2} \int u P_\delta |\Delta V \phi + \nabla V \cdot \nabla \phi| + \frac{\delta}{2N} \int u \Delta \phi + O(\delta^{3/2}),
\]
where the error estimates depend only on $C$, $N$, $\|u\|_{\text{Lip}}$, the support of $\phi$, and the $L^\infty$ norms of its derivatives. We also know that $(S_\delta P_\delta - P_\delta S_\delta)u$ is bounded by $\|u\|_{\text{Lip}}(C/c)(1 - e^{-c\delta})\delta^{3/2}$ 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 let us take $\delta = 2^{-t}$ and let $t = n\delta$. Recall that $T_{t,\ell} = (P_\delta S_\delta)^n$. Then by a telescoping series argument
\[
\int (T_{t,\ell} - 1)u \cdot \phi = \sum_{j=0}^{n-1} \left( -\frac{N\delta}{2} \int T_{j,\ell} u P_\delta |\Delta V \phi + \nabla V \cdot \nabla \phi| + \frac{\delta}{2N} \int T_{j,\ell} u \Delta \phi \right) + O(\delta^{1/2}).
\]

We fix a dyadic $t$ and take $\ell \to \infty$ (and hence $\delta \to 0$). Using Lemma 4.3, 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 $L^1_{\text{loc}}$. Thus, in the limit, we obtain
\[
\int (T_{t} - 1)u \cdot \phi dx = \int_{0}^{t} \int \left( -\frac{N\delta}{2} T_{s} u |\Delta V \phi + \nabla V \cdot \nabla \phi| + \frac{\delta}{2N} T_{s} u \Delta \phi \right) ds dx.
\]

We pass from dyadic $t$ to all positive $t$ using Lemma 4.3. Finally, after another integration by parts (which is justified by approximation by smooth functions in the appropriate Sobolev spaces), we have
\[
\int (T_{t} - 1)u \cdot \phi dx = \int_{0}^{t} \int \left( -\frac{N\delta}{2} (\nabla T_{s} u \cdot \nabla \phi) - \frac{\delta}{2N} \nabla T_{s} u \cdot \nabla \phi \right) ds dx.
\]

The asserted formula then follows by applying this formula to $T_{t_0}u$ with $t = t_1 - t_0$.

**Lemma 4.7.** If $\mu$ 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 4.6 and approximating $(1/Z) \exp(-N^2V(x))$ by compactly supported smooth functions, we see that
\[
\int T_{t} u \, d\mu - \int u \, d\mu = \frac{1}{Z} \int_{0}^{t} \left[ -\frac{1}{2N} \nabla (T_{s} u) \cdot \nabla [e^{-N^2V}] - \frac{N}{2} (\nabla V \cdot \nabla (T_{s} u)) e^{-N^2V} \right] ds dx = 0. \]
Lemma 4.8. We have $T_t u(x) \to \int u \, d\mu$ as $t \to \infty$ and more precisely

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

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

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

$$\leq \sum_{j=0}^{n-1} e^{-ct^4/2} e^{-crj/2n} \left( \frac{C}{c} (6 + 5\sqrt{2}) (r/n)^{1/2} + \|V(x)\|_2 (r/n) \right) \|u\|_{\text{Lip}}$$

$$\leq e^{-ct^4/2} e^{-crj/2n} \frac{2n}{ct} \left( \frac{C}{c} (6 + 5\sqrt{2}) (r/n)^{1/2} + \|V(x)\|_2 (r/n) \right) \|u\|_{\text{Lip}}$$

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

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

Because this holds for all sufficiently large $r$, this shows that $\lim_{t \to \infty} T_t u(x)$ exists. Because $\|T_t u\|_{\text{Lip}} \leq e^{-ct^4/2} \|u\|_{\text{Lip}}$, 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.  

4.2 Approximability and Convergence of Moments

Now we are ready to show that the semigroup $T^V_t$ 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 $\mu_N$ converge.

Lemma 4.9. Let $V_N : M_N(C)^m \to \mathbb{R}$ 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 $\mu_N$ be the associated measure. Let $S^V_N$ and $T^V_N$ 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 $L$-Lipschitz functions which is asymptotically approximable by trace polynomials. Then

(1) $\{S^V_N u_N\}$ is asymptotically approximable by trace polynomials.

(2) $\{T^V_N u_N\}$ is asymptotically approximable by trace polynomials.

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

Proof. (1) Recall that $S^V_N u_N = u_N(W_N(x,t))$, where $W_N$ is the solution to (4.2). Thus, by Lemma 3.17, 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 of Picard iterates $W_{N,t}$ given by

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

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

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

22
Because $DV_N$ is asymptotically approximable by trace polynomials, we know that $DV_N(0)$ is bounded as $N \to \infty$ and hence $\|DV_N(x)\|_2$ is uniformly bounded on $\|x\|_2 \leq R$ for each $R$, and thus convergence of $W_{N,\ell}$ to $W_N$ occurs on $\|x\|_2 \leq R$ uniformly with respect to $N$. Thus, by Observation 3.16 it suffices to show that each Picard iterate $\{W_{N,\ell}(x,t)\}_N$ is asymptotically approximable by trace polynomials.

In fact, let us fix $T > 0$. We claim that for every $\ell$, 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

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

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

$$R' = \sup_{t \in [0,T]} \sup_{N \in M_N(C)_\epsilon^\infty} \sup_{\|x\| \leq R} \|f(x,t)\| < +\infty.$$ 

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 3.17 we have for $\|x\| \leq R$ and $t \in [0,T]$ that

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

$$\leq \frac{1}{2} \sup_{\|y\| \leq R'} \|DV_N(h(y)) - g(y)\|_2 + \frac{Ct}{2} \sup_{s \in [0,T]} \sup_{\|x\| \leq R} \|W_{N,\ell}(x,s) - f(x,s)\|_2.$$

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

(2) We have shown that $S^V_{\ell_k}$ preserves asymptotic approximability and we also know that $P_\ell$ preserves asymptotic approximability by Lemma 3.18. Therefore, the iterated operator $T^V_{\ell_k}(0,0) = (P_{2^{-\ell}}S^V_{2^{-\ell}})^{2^{\ell}}$ preserves asymptotic approximability for dyadic values of $t$. Taking $\ell \to \infty$, we see by Observation 3.16 and Lemma 4.4 that $T^V_1$ preserves asymptotic approximation for dyadic values of $t$. Finally, we extend the approximability property to $T^V_{1^k}$ for all real $t$ using Observation 3.16 and Lemma 4.5.

(3) We know by Lemma 4.8 that $T^V_1 u_N(x) \to \int u_N \, d\mu_N$ as $t \to \infty$ with estimates that are independent of $N$. It follows by Observation 3.16 that the sequence of constant functions $\{\int u_N \, d\mu_N\}$ is asymptotically approximable by trace polynomials. But since these functions are constant, this simply means that the limit as $N \to \infty$ of $\int u_N \, d\mu_N$ exists. 

**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

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

When we take the $\limsup$ as $N \to \infty$, the first term is bounded by $2/e^{1/2}$ by Corollary 2.10 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

$$\int DV_N(x) \, d\mu_N(x) = 0$$
On the other hand, by our convexity assumptions
\[ \left\| DV_N(x) - \left( DV_N(0) + \frac{C + c}{2} x \right) \right\|_2 \leq \frac{C - c}{2} \|x\|_2. \]
Therefore,
\[ \left\| DV_N(0) + \frac{C + c}{2} a \right\|_2 \leq \frac{C - c}{2} \int \|x\|_2 d\mu_N(x) \]
\[ \leq \frac{C - c}{2} \left(\|a\|_2 + \left( \int \|x - a\|_2^2 d\mu(x) \right)^{1/2} \right) \]
\[ \leq \frac{C - c}{2} \left(\|a\|_2 + c^{-1/2} \right), \]
where the last step follows from Theorem 2.9 Altogether,
\[ \frac{C + c}{2} \|a\|_2 \leq \frac{C - c}{2} \|a\|_2 + \| DV_N(0)\|_2 + \frac{C - c}{2c^{1/2}}. \]
Then we move \((C - c)/2 \cdot \|a\|_2\) to the left hand side and divide the equation by \(c\) to obtain
\[ \left| \int \tau_N(x_j) d\mu_N(x) \right| \leq \|a\|_2 \leq \frac{1}{c} \| DV_N(0)\|_2 + \frac{C - c}{2c^{3/2}}, \]
which proves the asserted estimate on \(R_N\). The tail estimate on \(\mu_N(\|x_j\| \geq R_N + \delta)\) follows from Corollary 2.10

(2) Fix a non-commutative polynomial \(p\). Let \(\psi \in C_c^\infty(\mathbb{R})\) such that \(\psi(t) = t\) for \(|t| \leq 2R_0\), and set \(\Psi(x_1, \ldots, x_m) = (\psi(x_1), \ldots, \psi(x_m))\). Now \(x \mapsto \psi(x)\) is Lipschitz in \(\|\cdot\|_2\) for \(x \in M_N(\mathbb{C}_{sa})\) with constants independent of \(N\) (see for instance Proposition 8.7 below). It follows that \(p(\Psi(x))\) is globally Lipschitz in \(\|\cdot\|_2\) and it equals \(p(x)\) when \(\|x\| \leq 2R_0\). Also, because \(\psi(t)\) can be approximated on compact sets by polynomials, we see that \(\tau_N(p(\Psi(x)))\) is asymptotically approximable by trace polynomials. Therefore, by Lemma 4.9, the limit
\[ \lambda(p) = \lim_{N \to \infty} \int \tau_N(p(\Psi(x))) d\mu_N(x) \]
exists. Clearly, \(\lambda\) satisfies all the conditions to be a non-commutative law. Furthermore, because of the operator norm bounds (1), we know that \(\int_{\|x\| \geq 2R_0} \tau_N(p(x)) dx\) is finite and approaches zero as \(k \to \infty\) and the same holds for the integral of \(\tau_N(p(\Psi(x)))\). Therefore,
\[ \lim_{N \to \infty} \int \tau_N(p(x)) d\mu_N(x) = \lim_{N \to \infty} \int \tau_N(p(\Psi(x))) d\mu_N(x) = \lambda(p). \]

(3) It suffices to prove the concentration claim (3) for sufficiently large \(R\), for instance, \(R > 2 \sup_N R_N\). Because the topology of \(\Sigma_{m,R}\) is generated by non-commutative polynomials, it suffices to consider the case where \(\mathcal{U} = \{ \mathcal{N}: |\mathcal{N}'(p) - \lambda(p)| < \epsilon \}\) for some non-commutative polynomial \(p\). Choose a function \(\psi \in C_c^\infty(\mathbb{R})\) with \(\psi(t) = t\) for \(|t| \leq R\), and let \(\Psi\) be as above. Then by Theorem 2.8
\[ \mu_k \left( \left| \tau_N(p(\Psi(x))) - \int \tau_N(p \circ \Psi) d\mu_N \right| \geq \epsilon/2 \right) \leq 2e^{-2N^2 \epsilon^2 / \|\tau_N(p(\Psi))\|^2_{L^\infty}}. \]
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
\[
\limsup_{N \to \infty} \frac{1}{N^2} \log \mu_N (\|x\| \leq R, |\tau_N(p(x)) - \lambda(p)| \geq \epsilon) < 0.
\]

5 Entropy and Fisher’s Information

5.1 Classical and Free Entropy

In this section, we will state sufficient conditions for the microstates free entropies \( \chi \) and \( \chi \) to be evaluated as the lim sup and lim inf of renormalized classical entropies. Recall that the (classical, continuous) entropy of a measure \( d\mu = \rho(x) \, dx \) on \( \mathbb{R}^n \) is defined as
\[
h(\mu) := \int_{\mathbb{R}^n} -\log \rho,
\]
whenever the integral makes sense. Free entropy is defined as follows \cite{28, 29}:

**Definition 5.1.** For \( U \subseteq \Sigma_m \), we define the microstate space
\[
\Gamma_N(U) = \{ x \in M_N(\mathbb{C})_sa : \lambda_x \in U \}
\]
\[
\Gamma_{N,R}(U) = \{ x \in M_N(\mathbb{C})_sa : \lambda_x \in U, \|x\| \leq R \}.
\]

The microstates free entropy of a non-commutative law \( \lambda \) is defined as
\[
\chi_R(\lambda) = \inf_{U \ni \lambda} \limsup_{N \to \infty} \left( \frac{1}{N^2} \log \text{vol} \Gamma_{N,R}(U) + \frac{m}{2} \log N \right)
\]
\[
\chi(\lambda) = \sup_{R > 0} \chi_R(\lambda).
\]

Here \( U \) ranges over all open neighborhoods of \( \mu \) in \( \Sigma_m \). Similarly, we denote
\[
\chi_R(\lambda) = \inf_{U \ni \lambda} \liminf_{N \to \infty} \left( \frac{1}{N^2} \log \text{vol} \Gamma_{N,R}(U) + \frac{m}{2} \log N \right)
\]
\[
\chi(\lambda) = \sup_{R > 0} \chi_R(\lambda).
\]

**Definition 5.2.** A sequence of probability measures \( \mu_N \) on \( M_N(\mathbb{C})_sa \) is said to concentrate around the non-commutative law \( \lambda \) if \( \lambda_x \to \lambda \) in probability when \( x \) is chosen according to \( \mu_N \), that is, for any neighborhood \( \mathcal{U} \) of \( \mu \) in \( \Sigma_m \), we have
\[
\lim_{k \to \infty} \mu_N(x \in \Gamma_N(\mathcal{U})) = 1.
\]

**Proposition 5.3.** Let \( V_N : M_N(\mathbb{C})_sa \to \mathbb{R} \) be a potential with \( \int \exp(-N^2 V_N(x)) \, dx < +\infty \) and let \( \mu_N \) be the associated measure. Assume:

(A) The sequence \( \{\mu_N\} \) concentrates around a non-commutative law \( \lambda \).

(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

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

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

Then \( \lambda \) can be realized as the law of non-commutative random variables \( X = (X_1, \ldots, X_m) \) in a von Neumann algebra \((M, \tau)\) with \( \|X_j\| \leq R_0 \). Moreover, we have

\[
\chi(\lambda) = \chi_{R_0}(\lambda) = \lim_{N \to \infty} \left( \frac{1}{N^2} h(\mu_N) + \frac{m}{2} \log N \right),
\]

(5.1)

and

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

(5.2)

Proof. It follows from assumptions (A) and (D) that for every non-commutative polynomial \( p \),

\[
\lim_{N \to \infty} \int_{\|x\| \leq R_0} \tau_N(p(x)) d\mu_N(x) = \lambda(p).
\]

It follows that \( \lambda(X_j^{2n}) \leq R_0^{2n} \) for any \( n > 0 \). From here it is a standard fact that \( \lambda \) can be realized by self-adjoint random variables in a tracial von Neumann algebra which have norm \( \leq R_0 \).

Now let us evaluate \( \chi_R \) 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

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

Therefore, if we let

\[
d\mu_{N,R}(x) = \frac{1}{Z_{N,R}} 1_{\|x\| \leq R} \exp(-N^2 V_N(x)) \, dx, \quad Z_{N,R} = \int 1_{\|x\| \leq R} \exp(-N^2 V_N(x)) \, dx.
\]

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 scalar-valued trace polynomial \( f \) such that \( |V_N(x) - f(x)| \leq \epsilon/2 \) for \( \|x\| \leq R \) and for sufficiently large \( N \). Now because the trace polynomial \( f \)
is continuous with respect to convergence in non-commutative moments, the set $\mathcal{U} = \{ \lambda' : |\lambda'(f) - \lambda(f)| < \epsilon/2 \}$ is open. Now suppose that $\mathcal{V} \subseteq \mathcal{U}$ is a neighborhood of $\lambda$. Since the topology of $\Sigma_m$ is generated by non-commutative moments, assumption (A) implies that

$$\lim_{N \to \infty} \mu_{N,R}(\Gamma_{N,R}(\mathcal{V})) = 1.$$  

Moreover, by our choice of $f$ and $\mathcal{U}$, we have

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

Therefore,

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

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 $f$ is bounded by some constant $K$ whenever $\|x\| \leq R$. Therefore,

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

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

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

Altogether,

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

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

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

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

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

Because this holds for all sufficiently small neighborhoods $\mathcal{V}$ 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 $\epsilon \to 0$. Since $\chi_R(\lambda)$ is monotone in $R$, this establishes (5.1). The proof of (5.2) is identical.

$\square$
5.2 Classical Fisher Information

The classical Fisher’s information of a measure \(\mu\) describes how the entropy changes when \(\mu\) is convolved with a Gaussian. Suppose \(\mu\) is given by the density \(\rho\), and let \(\gamma_t\) be the multivariable Gaussian measure on \(\mathbb{R}^n\) with covariance matrix \(tI\). Then the density \(\rho_t\) for \(\mu_t = \mu * \gamma_t\) evolves according to the heat equation \(\partial_t \rho_t = (1/2) \Delta \rho_t\). Integration by parts shows that \(\partial_t h(\mu_t) = (1/2) \int |\nabla \rho_t/\rho_t|^2 d\mu\).

The Fisher information of \(\mu\) 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 \(\Xi(x) = -\nabla \rho(x)/\rho(x)\), which is known as the score function. Following Voiculescu’s terminology in the free setting, we will call \(\Xi(x)\) the \((\text{classical})\) conjugate variable for the law \(\mu\). The conjugate variable \(\Xi(x)\) satisfies the following integration by parts formula:

\[
\int \Xi \cdot f \, d\mu = - \int \frac{\nabla \rho(x)}{\rho(x)} f(x) \rho(x) \, dx = \int \rho(x) \nabla f(x) \, dx = \int \nabla f \, d\mu,
\]

for \(f \in C_c^\infty\) and in fact it is characterized by this integration by parts formula.

The entropy of a measure \(\mu\) can be recovered by integrating the Fisher information of \(\mu * \gamma_t\). The following integral formula was the motivation for Voiculescu’s definition of non-microstates free entropy \(\chi^*\). For the reader’s convenience, we include a statement and proof in the random matrix setting with free probabilistic normalizations. See also [8, Lemma 1] and [30, Proposition 7.6].

**Lemma 5.4.** Let \(\mu\) be a probability measure on \(M_N(\mathbb{C})_{sa}\) with finite variance and with a density \(\rho\), and let \(\sigma_{t,N}\) be the law of \(m\) independent GUE’s of variance \(t\). If \(a\) is the variance of \(\mu\) in the normalized norm \(\|\cdot\|_2\), then we have

\[
\frac{m}{a + t} \leq \frac{1}{N^3} \mathcal{I}(\mu * \sigma_{t,N}) \leq \frac{m}{t},
\]

Moreover,

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

and

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

**Proof.** To prove (5.4), let \(X\) and \(Y\) be independent random variables with the laws \(\mu\) and \(\sigma_{t,N}\). As a consequence of the integration by parts formula, we have

\[
\Xi(X + Y) = E[\Xi(Y)|X + Y],
\]

and hence \(\mathcal{I}(X + Y) \leq \mathcal{I}(Y)\). (See [30, Proposition 3.7] for the free case.) But \(\Xi(Y) = (N/t)Y\) which implies \(\mathcal{I}(Y) = N^3m/t\) and proves the upper bound. For the lower bound, note that by Cauchy-Schwarz

\[
E[|\Xi(X + Y)|^2] \geq \frac{|E[\Xi(X + Y)]|^2}{E[X + Y]^2} = \frac{|E[(N/t)Y]|^2}{E[X + Y]^2} = \frac{N^3m}{a + t}.
\]
Next, to prove (5.5), let $\rho_t$ be the probability density of $\mu_t := \mu * \sigma_{t,N}$. Fix $t > \delta > 0$. By basic properties of convolving positive functions with the Gaussian, note that for each $s > 0$, $\rho_s$ is smooth and bounded above, and there are some constants $\alpha$, $\beta$ such that $\rho_s(x) \geq \alpha e^{-\beta\|x\|^2}$ for $s \in [\delta, t]$. This implies that $\log \rho_t$ is bounded below by a constant above by a quadratic. In particular, $h(\mu * \sigma_{s,N})$ is finite for $s \geq \delta$.

To justify integration by parts, consider a smooth compactly supported “cutoff” function $\psi_R : M_N(\mathbb{C})^m \to \mathbb{R}$ such that $0 \leq \psi_R \leq 1$ and $\psi_R(x) = 1$ when $|x| \leq R$ and $\psi_R(x) = 0$ when $|x| \geq 2R$. We can arrange that $\|\nabla \psi_R(x)\|_2 \leq C/R$. Because $\partial_s \rho_s = (1/2N)\Delta \rho_s$, we have

$$
\frac{d}{dt} - \int \psi_R \rho_s \log \rho_s = -\frac{1}{2N} \int \psi_R \cdot (\Delta \rho_s \log \rho_s + \Delta \rho_s)
= \frac{1}{2N} \int \psi_R |\nabla \rho_s|^2 \rho_s + \frac{1}{2N} \int \nabla \psi_R \cdot \nabla \rho_s \cdot (1 + \log \rho_s).
$$

Letting $\Xi_s = -\nabla \rho_s/\rho_s$, this is equal to

$$
\frac{1}{2N} \int \psi_R |\Xi_s|^2 d\mu_s - \frac{1}{2N} \int (\nabla \psi_R \cdot \Xi_s)(1 + \log \rho_s) d\mu_s.
$$

Of course, we have

$$
\lim_{R \to +\infty} \int \psi_R |\Xi_s|^2 d\mu_s = \mathcal{I}(\mu_s).
$$

The other term is an error which can be estimated as follows: Recall that $1 + \log \rho_s$ is sub-quadratic and that $\nabla \psi_R(x)$ is supported when $R \leq |x| \leq 2R$ bounded by $1/R \sim 1/|x|$. Altogether we have $|\nabla \psi_R(1 + \log \rho_s)| \leq \gamma |x|$ for some constant $\gamma$ when $|x|$ is large enough. Thus,

$$
\int |(\nabla \psi_R \cdot \Xi_s)(1 + \log \rho_s)| d\mu_s \leq \int_{|x| \geq R} |x| \Xi_s(x) d\mu_s(x),
$$

which is the tail of a finite integral because $x$ and $\Xi_s(x)$ are both in $L^2(\mu_s)$ with norms bounded respectively by the variance and Fisher information of $\mu_s$. The error thus goes to zero as $R \to +\infty$ by the dominated convergence theorem, and in fact we can say the same thing even after we integrate the error from $\delta$ to $t$ with respect to $s$. The dominated convergence theorem also applies to $\int \psi_R \rho_s \log \rho_s$. The result is that

$$
h(\mu_t) - h(\mu_\delta) = \frac{1}{2N} \int_\delta^t \int |\Xi_s|^2 d\mu_s ds = \frac{1}{2N} \int_\delta^t \mathcal{I}(\mu_s) ds.
$$

Now we take $\delta \to 0$. Note that $h(\mu_\delta) \geq h(\mu)$ by a convexity argument, while $\limsup_{\delta \to 0} h(\mu_\delta) \leq h(\mu)$ follows from Fatou’s lemma because $\rho_\delta \to \rho$ almost everywhere and $\rho_\delta \log \rho_\delta$ is bounded above. This completes the proof of (5.5).

To prove (5.6), we follow [30, Proposition 7.6]. Note that

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

Because $\mu_t = \mu * \sigma_{t,N}$, we have $h(\mu_t) \geq h(\sigma_{t,N})$. On the other hand, it is a standard fact that the Gaussian has the largest entropy out of all measures of a given variance. Therefore, if $a = \int |x|^2 \mu$ is the normalized variance of $\mu$, then $h(\mu_t) \leq h(\sigma_{t+a,N})$. But

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

29
and the same holds for \( t \) replaced by \( t + a \). The upshot is that as \( t \to +\infty \),
\[
    h(\mu_t) - \frac{mN^2}{2} \log(1 + t) \to \frac{mN^2}{2} (\log 2\pi e - \log N).
\]
This proves \( \text{(5.6)} \) modulo showing that the integral from 0 to \( +\infty \) is well-defined. But it follows from \( \text{(5.4)} \) that the integral from 1 to \( +\infty \) is absolutely convergent. On the interval from 0 to 1 the integrand is bounded above and thus the integral is well-defined if we allow the value \( -\infty \).

5.3 Free Fisher Information

The starting point for the definition of free Fisher information is the integration by parts formula \( \text{(5.3)} \). Indeed, if we apply this to a non-commutative polynomial \( p \) and renormalize, we obtain
\[
    \int \tau_N \left( \frac{1}{N} \Xi_j(x)p(x) \right) \, d\mu(x) = \int \tau_N \otimes \tau_N (D_j p(x)) \, d\mu(x). \tag{5.7}
\]
Voiculescu therefore made the following definitions:

**Definition 5.5** ([30, §3]). Let \( X = (X_1, \ldots, X_m) \) be a tuple of self-adjoint random variables in a tracial von Neumann algebra \((M, \tau)\) and assume that \( M \) is generated by \( X \) as a von Neumann algebra. We say that \( \xi = (\xi_1, \ldots, \xi_m) \in L^2(M, \tau) \) is the (free) conjugate variable of \( X \) if
\[
    \tau(\xi_j p(X)) = \tau(\otimes (D_j p(X))) \tag{5.8}
\]
for every non-commutative polynomial \( p \). In this case, we say that \( X \) (or equivalently the law of \( X \)) has finite free Fisher information and define \( \Phi^*(\lambda) := \Phi^*(\lambda_X) := \sum \tau(\xi_j^2) \). We also denote the conjugate variable \( \xi \) by \( J(X) \).

**Definition 5.6** ([30, Definition 7.1]). The non-microstates free entropy of a non-commutative law \( \lambda \) is
\[
    \chi^*(\lambda) := \frac{1}{2} \int_0^{\infty} \left( \frac{\log 1 + t}{t} - \Phi^*(\lambda \boxplus \sigma_t) \right) + \frac{1}{2} \log 2\pi e.
\]

Now we are ready to state conditions under which the classical Fisher information of a sequence of measures \( \mu_N \) converges to the free Fisher information of the law \( \lambda \). 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 \( \Xi_N = N \nabla V_N \). The normalized conjugate variable used in \( \text{(5.7)} \) is \( (1/N)\Xi_N = NV = D N \). The corresponding normalized Fisher information is then
\[
    \int \| D N \|^2 \, d\mu_N = \int \frac{1}{N} \left| \frac{1}{N} \Xi_N \right|^2 \, d\mu = \frac{1}{N^3} \mathcal{I}(\mu_N),
\]
which is the same normalization as in Lemma \( \text{5.4} \).

**Proposition 5.7.** Let \( V_N : M_N(\mathbb{C})_m^+ \to \mathbb{R} \) be a potential with \( \int \exp(-N^2 V_N(x)) \, dx < +\infty \) and let \( \mu_N \) be the associated measure. Assume:

(A) The sequence \( \mu_N \) concentrates around a non-commutative law \( \lambda \).

(B) The sequence \( \{D N\} \) is asymptotically approximable by trace polynomials.

(C) For some \( n \geq 0 \) and \( a, b > 0 \) we have \( \| D N \|^2 \leq a + b \sum_{j=1}^m \tau_N(x_j^{2n}) \).
(D) There exists \( R_0 > 0 \) such that

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

Then

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

(2) There exists a sequence of trace polynomials \( f_k \in (\text{TrP}_m^1)^m \) such that

\[
\lim \lim \sup_{k \to \infty} \sup_{\|x\| \leq R_0} \|DV_N(x) - f_k(x)\|_2 = 0.
\]

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

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

Proof. (1) This follows from the same argument as Proposition 3.3.2 of the text.

(2) This follows from the definition of asymptotic approximability by trace polynomials. (3) Observe that

\[
\lambda([f_j - f_k]^*(f_j - f_k)] = \lim_{N \to \infty} \int_{\|x\| \leq R_0} \tau_N([f_j - f_k]^*(f_j - f_k)](x)) \, d\mu_N(x).
\]

For any \( \epsilon > 0 \), if \( j \) and \( N \) are large enough, then \( \sup_{\|x\| \leq R_0} \|DV_N(x) - p_j(x)\|_2 < \epsilon \). In particular, if \( j \) and \( k \) are sufficiently large, then \( \lambda([p_j - p_k]^*(p_j - p_k))] < (2\epsilon)^2 \). This shows that \( \{p_k(X)\} \) is Cauchy in \( L^2(M, \lambda) \) since \( X \) has the law \( \lambda \).

Let \( \xi = \lim_{k \to \infty} f_k(X) \). We must show that \( \xi \) is the conjugate variable for \( X \). Let \( \psi \in C_0^\infty(\mathbb{R}) \) such that \( \psi(y) = y \) when \( |y| \leq R_0 \). For \( x \in M_N(\mathbb{C})_m^m \), let \( \Psi(x) = (\psi(x_1), \ldots, \psi(x_m)) \).

Note that for any non-commutative polynomial \( p \), we have

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

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

\[
\lim_{N \to \infty} \int_{\|x\| \geq R_0} \|DV_N(x)\|_2^2 \, d\mu_N(x) = 0.
\]

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

\[
\int_{\|x\| \leq R_0} \tau_N(D_j V_N(x)p(\Psi(x))) \, d\mu_N(x) - \int_{\|x\| \leq R_0} D_j[\tau_N(p(\Psi(x)))] \, d\mu_N(x) \to 0
\]

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

\[
\int_{\|x\| \leq R_0} \tau_N(D_j V_N(x)p(x)) \, d\mu_N(x) - \int_{\|x\| < R_0} \tau_N \otimes \tau_N[D_j p(x)] \, d\mu_N(x) \to 0.
\]

31
Now the second term converges to $\lambda \otimes \lambda[D_j p] = \tau \otimes \tau[D_j p(X)]$ by our concentration assumption (A). For the first term, we can replace $D_j V_N(x)$ by $(f_k(x))_j$ with an error bounded by $\sup_{\|x\| \leq R_0} \|f_k(x) - DV_N(x)\|_2$. Then we apply concentration to conclude that $\int \tau_N[(f_k)_j^* p(x)] d\mu_N(x) \to \lambda[(f_k)_j^* p]$. Overall,

$$|\lambda[(f_k)_j p] - \lambda \otimes \lambda[D_j p]| \leq \limsup_{N \to \infty} \sup_{\|x\| \leq R_0} \|f_k(x) - DV_N(x)\|_2.$$ 

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

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

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

By similar arguments as before, we can approximate $DV_N$ by $f_k$ on $\|x\| \leq R_0$, approximate $\int_{\|x\| \leq R_0} \|f_k\|_2^2 d\mu_N$ by $\lambda(f_k^* f_k)$, and then approximate $\lambda(f_k^* f_k)$ by $\tau(\xi \otimes \xi) = \Phi^*(\lambda)$, 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

6.1 Motivation and Statement of the Equation

In the last section, we stated conditions under which the classical entropy and Fisher information of $\mu_N$ converge to their free counterparts for the limiting non-commutative law $\lambda$. In order to prove that $\chi(\lambda) = \chi^*(\lambda)$, we want to take the limit in the integral formula (6.6), and therefore, we want $N^{-3}|\mathcal{I}(\mu_N \ast \sigma_{t,N}) \to \Phi^*(\lambda \ast \sigma_t)$ for all $t > 0$. In order to apply Proposition 6.7 to $\mu_N \ast \sigma_{t,N}$, we need to show that $\{DV_{N,t}\}_N$ is asymptotically approximable by trace polynomials, where $V_{N,t}$ is the potential corresponding to $\mu_N \ast \sigma_{t,N}$.

As in (4) we will analyze the PDE which describes the evolution of $V_{N,t}$. After deriving the equation (below) and reviewing some PDE background (6.2), we will approximate the solution by iterating simpler operations (6.3), and hence argue that if $DV_N$ is asymptotically approximable by trace polynomials, then so is $DV_{N,t}$ (6.4).

From now until (6.4) we will simplify the notation by fixing $N$ and working with a single potential $V : M(C)\to \mathbb{R}$. If the measure $\mu$ is given by $(1/Z) \exp(-N^2V(x)) dx$, we define the potential $V(x,t)$ by

$$\exp(-N^2V(x,t)) = \int \exp(-N^2V(x+y)) d\sigma_{t,N}(y). \quad (6.1)$$

Then $\exp(-N^2V(x,t))$ solves the heat equation

$$\partial_t[\exp(-N^2V(x,t))] = \frac{1}{2} L[\exp(-N^2V(x,t))], \quad (6.2)$$

where $L_N = (1/N)\Delta$ is the normalized Laplacian. After some computation, we see that

$$\partial_t V = \frac{1}{2} L_N V - \frac{1}{2} \|DV\|_2^2. \quad (6.3)$$

where $DV = N\nabla V$ is the normalized gradient.
6.2 Viscosity Solutions; Existence and Uniqueness

In §6.3 we will construct a convergent sequence of functions. To identify the limit as the smooth solution of §6.3, we will show that the limit function is a solution in the viscosity sense (see [8]), and use the fact that there is only one viscosity solution. In preparation, we recall the following definitions and facts concerning viscosity solutions.

**Definition 6.1.** Consider the partial differential equation \( \partial_t u + F(Hu,Du,u,x) = 0 \) for a function \( u : \mathbb{R}^n \times [0, +\infty) \to \mathbb{R} \), where \( F : M^{sa}_n(\mathbb{R}) \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R} \) is continuous.

1. We say that \( u : \mathbb{R}^n \to \mathbb{R} \) is a viscosity subsolution if it is upper semi-continuous, and if for \( A \in L(M_N(C)^m_{sa}, M_N(C)^m_{sa}) \) and \( p \in M_N(C)^m_{sa} \) and \( \alpha \in \mathbb{R} \) and \( x_0 \in M_N(C)^m_{sa} \), the condition
   \[
   u(x,t) \leq u(x_0,t_0) + \alpha(t-t_0) + p \cdot (x-x_0) + \frac{1}{2} A(x-x_0) \cdot (x-x_0) + o(|t-t_0| + |x-x_0|^2) \quad (6.4)
   \]
   implies that \( \alpha + F(A,p,u(x_0),x_0) \leq 0 \).

2. We say that \( u : \mathbb{R}^n \to \mathbb{R} \) is a viscosity supersolution if it is lower semi-continuous, and if the condition
   \[
   u(x,t) \geq u(x_0,t_0) + \alpha(t-t_0) + p \cdot (x-x_0) + \frac{1}{2} A(x-x_0) \cdot (x-x_0) + o(|t-t_0| + |x-x_0|^2) \quad (6.5)
   \]
   implies that \( \alpha + F(A,p,u(x_0),x_0) \geq 0 \).

3. We say that \( u \) is a viscosity solution if it is both a subsolution and a supersolution.

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 the correct inequality.

When we apply the definition of viscosity solution for \( M_N(C)^m_{sa} \cong \mathbb{R}^{mN^2} \), we will write our Taylor expansions using the normalized inner product \( \langle \cdot, \cdot \rangle_2 \) rather than \( \langle \cdot, \cdot \rangle_{\text{Tr}} \). For instance, if \( u \) is twice differentiable at \( 0 \), we will write
   \[
   u(x) = a + \langle p, x \rangle_2 + \langle Ax, x \rangle_2 + o(\|x\|^2),
   \]
   where \( A \) is a linear transformation \( M_N(C)^m \to M_N(C)^m \). In this case, the normalized Laplacian \( Lu(0) \) is given by \( \frac{1}{2N} \text{Tr}(A) = \frac{1}{2N} \text{Tr}(Hu) \), where \( \text{Tr}(A) \) can either be computed either by using \( \langle \cdot, \cdot \rangle_2 \) and an orthonormal basis in this inner product, or by using \( \langle \cdot, \cdot \rangle_{\text{Tr}} \) and an associated orthonormal basis.

**Lemma 6.2.** Let \( u_0 : M_N(C)^m_{sa} \to \mathbb{R} \) be continuous and bounded below. Then there is a unique continuous and bounded below function \( u : M_N(C)^m_{sa} \times [0, +\infty) \to \mathbb{R} \) such that

1. We have the initial condition \( u(x,0) = u_0(x) \).

2. The function \( u \) is a viscosity solution to \( \partial_t u - \frac{1}{2N} \text{Tr}(Hu) + \frac{1}{2} \|Du\|^2 = 0 \).

**Proof.** Note that \( u \) is a viscosity solution of this equation if and only if \( \exp(-N^2u) \) is a viscosity solution of the heat equation. More precisely, \( u \) is a subsolution if and only if \( \exp(-N^2u) \) is a supersolution and vice versa. To see this, note that because \( \exp : \mathbb{R} \to \mathbb{R} \) is monotone and analytic, a Taylor approximation from above for \( u \) will produce a Taylor expansion from below for \( \exp(-N^2u) \), and the inequalities which verify sub / super-solution turn out to have the correct signs. We leave the details to the reader.
Next, note that \( u \) is bounded below if and only if \( \exp(-N^2 u) \) is bounded. We know that the heat equation has a smooth solution for bounded initial data. Furthermore, if \( w_1 \) is a smooth solution of the heat equation and \( w_2 \) is a viscosity solution with the same initial data, and both are bounded and continuous on \( M_N(C)^{m}_a \times [0, +\infty) \), then we must have \( w_1 = w_2 \). This follows from the standard argument for the maximum principle for the heat equation. To summarize the argument, for \( \epsilon > 0 \), we consider the function

\[
w_2(x, t) - w_1(x, t) - \frac{1}{2} \epsilon \|x\|^2_2 - 2\epsilon t.
\]

This function must achieve a maximum. If we suppose the maximum is achieved at some \((x_0, t_0)\) with \( t_0 > 0 \), then because \( w_1(x, t) - \frac{1}{2} \epsilon \|x\|^2_2 - 2\epsilon t \) is smooth, then we have a Taylor expansion from above for \( w_2 \) at \((x_0, t_0)\), so we can apply the subsolution condition for \( w_2 \) to get a contradiction. Taking \( \epsilon \to 0 \) shows that \( w_1 \leq w_2 \) and a symmetrical argument shows \( w_2 \geq w_1 \).

Altogether, we have the existence and uniqueness claims for the heat equation and therefore we have them also for the given equation (6.3). \( \square \)

### 6.3 Approximate Solutions by Iteration

To approximate the solution to (6.3), we view the equation as a hybrid between the heat equation and \( \partial_t u = -(1/2)\|Du\|^2_2 \). The heat equation can be solved by the heat semigroup

\[
P_t u(x) := \int u(x + y) d\sigma_{t,N}(y), \quad (6.6)
\]

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

\[
Q_t u(x) := \inf \left[ u(x + y) + \frac{1}{2t}\|y\|^2_2 \right] \quad (6.7)
\]

as a special case of the Hopf-Lax formula (see [14, 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 \( E[|u(x + B_t + \int_0^t Y_s ds)| + (1/2) \int_0^t \|Y_s\|^2_2 ds \) over a certain class of stochastic processes \( Y_t \) adapted to a standard Brownian motion \( B_t \) (see [3, 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 then identifies the minimizing process \( Y_t \) as a Brownian bridge \([3, \text{Section 5}]\) and analyzes it using forward-backward SDE. Through the Picard iteration solving the SDE, he shows that the solution is well-approximated by non-commutative functions.

We instead give a deterministic proof following the same strategy as in \([3]\). 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 \([3]\) 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 the space

\[
\mathcal{E}_C = \left\{ u : M_N(C)^{m}_a \to \mathbb{R}, u \text{ is convex and } u(x) - \frac{1}{2} C \|x\|^2_2 \text{ is concave} \right\},
\]

34
in which all the functions $u$ satisfy $\|Du\|_{\text{Lip}} \leq C$ a priori. 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. Though the estimates below (especially 1b and 2a) may not be optimal, they are dimension-independent and sufficient for our purposes. Below, $\omega(x) = (1 + \|x\|^2)^{1/2}$ denotes the “Japanese absolute value” and $L^\infty$ denotes $L^\infty(M(N(C)^m_{sa}, \mathbb{R})$ or $L^\infty(M(N(C)^m_{sa}, \mathbb{R})$ with respect to the normalized Hilbert-Schmidt norm $\|\cdot\|_2$ on $M(N(C)^m_{sa})$. We also denote by $Q^+_2 = \bigcup_{n \geq 0} 2^{-n}\mathbb{N}$ the nonnegative dyadic rationals.

**Theorem 6.3.** There exists a semigroup of nonlinear operators $R_t : \mathcal{E}_C \to \mathcal{E}_C$ with the following properties: Here, the $K_j$’s are constants depending only on $C$.

1. For $t \in 2^{-t} \mathbb{N}$, denote $R_{t,e}u = (P_{2^{-t}}Q_{2^{-t}})^{2^{t+1}}u$. If $t \in Q_2^+$, then
   
   (a) $\lim_{t \to \infty} R_{t,e}u$,
   
   (b) $\|R_{t,e}u - R_{t}u\|_{L^\infty} \leq 2^{-t}t \exp(K_1 + K_2t^{3/2} + K_3t\|Du/\omega\|_{L^\infty}).$
   
   (c) $\|D(R_{t,e}u) - D(R_{t}u)\|_{L^\infty} \leq Ce^{Ct}t^{1/2}/(1 - 2^{1/2})$.

2. For $s \leq t \in \mathbb{R}^+$ and $u, v \in \mathcal{E}_C$, we have

   (a) $\|R_{t,u} - R_{t,v}\|_{L^\infty} \leq \|(u - v)/\omega^2\|_{L^\infty} \exp(K_4 + K_5t^{3/2} + K_6t\|Du/\omega\|_{L^\infty}).$
   
   (b) $\|D(R_{t,u}) - D(R_{t,v})\|_{L^\infty} \leq e^{Ct}\|Du - Dv\|_{L^\infty}.$
   
   (c) $\|R_{t,u} - R_{t,v}\|_{L^\infty} \leq |t - s|(K_7 + K_8t^2 + K_9\|Du/\omega\|_{L^\infty}^2)$.
   
   (d) $\|D(R_{t,u}) - D(R_{t,v})\|_{L^\infty} \leq |t - s|^{1/2}K_{10}e^{Ct} + |t - s|(K_{11} + K_{12}t + K_{13}\|Du/\omega\|_{L^\infty}^2)$.

3. For $u_0 \in \mathcal{E}_C$, the function $u(x,t) = R_tu_0(x)$ is a viscosity solution of the equation $\partial_t u - \frac{1}{2}Lu + \frac{1}{2}\|Du\|_2^2 = 0$ and it is equal to the smooth solution obtained from applying the heat equation to $e^{-(N^2u_0)}$. Hence, $P_t[e^{-(N^2u_0)}] = e^{-(N^2R_tu_0)}$.

**Remark 6.4.** Claim (3) in fact gives a direct proof that if $u_0$ is convex and semi-concave, then so is $u(x,t)$, without appealing to the Braskamp-Lieb inequality.

We begin with some basic results about how $P_t$ and $Q_t$ interact with convexity, which in particular show that $P_t$ and $Q_t$ map $\mathcal{E}_C$ into $\mathcal{E}_C$.

**Definition 6.5.** For $C \in \mathbb{R}$, we say that $Hu \geq C$ if $u(x) - (C/2)\|x\|^2$ is convex, and we say that $Hu \leq C$ if $u(x) - (C/2)\|x\|^2$ is concave.

**Lemma 6.6.** Let $c, C \in \mathbb{R}$.

1. If $Hu \geq c$, then $H(P_tu) \geq c$.
2. If $Hu \leq C$, then $H(P_tu) \leq C$.
3. If $Hu \geq 0$, then $H(Q_tu) \geq 0$.
4. If $Hu \leq C$, then $H(Q_tu) \leq C$.

In particular, $P_t$ and $Q_t$ map $\mathcal{E}_C$ into $\mathcal{E}_C$. 

35
Proof. The first two properties are basic facts about convolution with a positive density like the Gaussian. For the third property, assume that \( u \) is convex. Now observe that for \( \alpha \in (0, 1) \) and any \( y_1, y_2 \), we have

\[
Q_t u(\alpha x_1 + (1 - \alpha) x_2) \leq u(\alpha x_1 + (1 - \alpha) x_2 + \alpha y_1 + (1 - \alpha) y_2) + \frac{1}{2t} \|\alpha y_1 + (1 - \alpha) y_2\|_2^2
\]

\[
\leq \alpha \left( u(x_1 + y_1) + \frac{1}{2t} \|y_1\|_2^2 \right) + (1 - \alpha) \left( u(x_2 + y_2) + \frac{1}{2t} \|y_2\|_2^2 \right).
\]

By taking the infimum over \( y_1 \) and \( y_2 \) on the left, we obtain

\[
Q_t u(\alpha x_1 + (1 - \alpha) x_2) \leq \alpha Q_t u(x_1) + (1 - \alpha) Q_t u(x_2).
\]

For the fourth property, note that \( Q_t u(x) = \inf_y g_y(x) \), where \( g_y(x) := u(x + y) + \frac{1}{2t} \|y\|_2^2 \). Since \( H g_y \leq C \), we also have \( H(Q_t u) \leq C \).

Lemma 6.7. If \( u \in \mathcal{E}_C \), then \( u \) is differentiable everywhere and \( Du \) is \( C \)-Lipschitz. Moreover, \( \|Du/\omega\|_{L_\infty} \) and \( \|u/\omega^2\|_{L_\infty} \) are finite.

Proof. Exercise.

The following lemma gives basic properties of \( Q_t \) from the PDE literature; see for instance [20], [8, Lemma A.5], [14, Section 3.3.2].

Lemma 6.8. For \( u \in \mathcal{E}_C \) and \( t \in \mathbb{R}^+ \),

1. \( Q_t u(x) = u(x - t D(Q_t u)(x)) + \frac{1}{2} \|D(Q_t u)(x)\|_2^2 \).
2. \( D(Q_t u)(x) = D u(x - tD(Q_t u)(x)) \).
3. \( \|D(Q_t u)(x)\|_2 \leq \|D u(x)\|_2 \).
4. \( \|D(Q_t u)(x) - D u(x)\|_2 \leq Ct \|D u(x)\|_2 \).
5. \( u(x) - \frac{1}{2} \|D u(x)\|_2^2 \leq Q_t u(x) \leq u(x) - \frac{1}{2} \|D u(x)\|_2^2 + \frac{Ct^2}{2} \|D u(x)\|_2^2 \).

Proof. (1) Fix \( x_0 \) and let us write

\[
Q_t u(x_0) = \inf_y \left[ u(y) + \frac{1}{2t} \|y - x_0\|_2^2 \right].
\]

Let \( y_0 \) be a point which achieves the infimum. Because \( Q_t u(x_0) \) is convex,

\[
Q_t u(x) \geq Q_t u(x_0) + \langle p, x - x_0 \rangle_2.
\]

where \( p = D(Q_t u(x_0)) \). For any \( x \) and \( y \), we have

\[
u(y) + \frac{1}{2t} \|y - x\|_2^2 \geq Q_t u(x)
\]

\[
\geq Q_t u(x_0) + \langle p, x - x_0 \rangle_2
\]

\[
= u(y_0) + \frac{1}{2t} \|x_0 - y_0\|_2^2 + \langle p, x - x_0 \rangle_2
\]

Now we substitute \( y = y_0 \) and \( x = x_0 + \epsilon [t^{-1}(y_0 - x_0) + p] \) for small \( \epsilon \in \mathbb{R} \) to conclude that

\[
\frac{1}{2t} \|y_0 - x_0 - \epsilon [t^{-1}(y_0 - x_0) + p]\|_2^2 \geq \frac{1}{2t} \|y_0 - x_0\|_2^2 + \langle p, \epsilon [t^{-1}(y_0 - x_0) + p] \rangle_2.
\]
and then it follows that
\[ -t^{-1}(y_0 - x_0, \epsilon t^{-1}(y_0 - x_0) + p) + O(\epsilon^2) \geq \langle p, \epsilon t^{-1}(x_0 - y_0) + p \rangle, \]
so that
\[ -t^{-1}\epsilon \|t^{-1}(y_0 - x_0) + p\|_2^2 \geq 0. \]
Taking \( \epsilon \searrow 0 \), we see that \( y_0 = x_0 - tp \), which proves that the minimizer \( y_0 \) is unique and (1) holds.

(2) We substitute \( x = y + x_0 - y_0 \) in (6.10) and get
\[ u(y) \geq u(y_0) + \langle p, y - y_0 \rangle, \]
which proves that \( Du(y_0) = p = D(Q_tu)(x_0) \).

(3) Fix \( x_0, y_0, \) and \( p \) as in the previous parts. By convexity of \( u \), we know that
\[ 0 \leq \frac{d^2}{d\epsilon^2} u(y_0 + \epsilon p) = \frac{d}{d\epsilon} \left[Du(y_0 + \epsilon p), p\right]. \]
Evaluating at \( \epsilon = 0 \) and \( \epsilon = t \), we obtain
\[ \langle Du(y_0), p \rangle \leq \langle Du(x_0), p \rangle, \]
so that \( \|Du(y_0)\|_2^2 \leq \|Du(x_0)\|_2^2 \|Du(y_0)\|_2 \), so that \( \|Du(y_0)\|_2 \leq \|Du(x_0)\|_2 \).

(4) This follows from (3) and (2) because \( Du \) is \( C \)-Lipschitz.

(5) By convexity,
\[ u(x - tD(Q_tu)(x)) \geq u(x) - t\langle Du(x), D(Q_tu)(x) \rangle, \]
Therefore,
\[ Q_tu(x) \geq u(x) - t\langle Du(x), D(Q_tu)(x) \rangle + \frac{t}{2} \|D(Q_tu)(x)\|_2^2 + E(x) \]
\[ = u(x) - \frac{t}{2} \|Du(x)\|_2^2 + \frac{t}{2} \|Du(x) - D(Q_tu)(x)\|_2^2 \]
\[ \geq u(x) - \frac{t}{2} \|Du(x)\|_2^2. \]
Similarly, by semi-concavity,
\[ u(x - tD(Q_tu)(x)) \leq u(x) - t\langle Du(x), D(Q_tu)(x) \rangle + \frac{Ct^2}{2} \|D(Q_tu)(x)\|_2^2, \]
so the second inequality of (5) follows from the same argument, using (3) to estimate \( D(Q_tu) \) by \( Du \).

Our next goal is to prove some estimates on \( P_t \) and \( Q_t \) that will help us control how errors propagate when we apply the operators iteratively. To simplify the statement and proof, we introduce the notation
\[ \mathcal{N}(u) := \frac{1}{2C^2} \sup_x \|Du(x)\|_2^2 - 2C^2\|x\|_2^2. \]
Because \( Du \) is \( C \)-Lipschitz, a straightforward estimate proves that
\[ \mathcal{N}(u) \leq \frac{1}{C^2} \|Du(0)\|_2^2, \]
and on the other hand,
\[ \|Du/\omega\|_{L^\infty}^2 \leq 2C^2 \max(\mathcal{N}(u), 1). \]
Lemma 6.9. Let $s_1, \ldots, s_n$ and $t_1, \ldots, t_n \geq 0$ and write

$$s^* = s_1 + \cdots + s_n$$
$$t^* = t_1 + \cdots + t_n$$
$$T = P_{s_1}Q_{t_1} \cdots P_{s_n}Q_{t_n}$$

Then we have the estimates

1. $\mathcal{N}(Ru) \leq \mathcal{N}(u) + s^*$.
2. $\left\| \frac{Ru - Ru}{\omega} \right\|_{L^\infty} \leq \exp \left( s^* + 2^{3/2}Ct^* \max(\mathcal{N}(u) + s^*, \mathcal{N}(v) + s^*, 1)^{1/2} \right) \left\| \frac{u - v}{\omega} \right\|_{L^\infty}$.
3. $\left\| D(Ru) - D(Rv) \right\|_{L^\infty} \leq e^{Ct^*} \left\| Du - Dv \right\|_{L^\infty}$.

Proof. To prove (1), first consider applying a single operator $P_s$ or $Q_t$. For the operator $P_s$, we have

$$\left\| D(P_su(x)) \right\|^2 \leq \left\| \int Duv(x + y) \, d\sigma_{s,N}(y) \right\|^2 \leq \int \left\| Duv(x + y) \right\|^2 \, d\sigma_{s,N}(y) \leq \int (2C^2\mathcal{N}(u) + 2C^2\|x + y\|_2^2) \, d\sigma_{s,N}(y) = 2C^2\mathcal{N}(u) + 2C^2s + 2C^2\|x\|_2^2,$$

and hence

$$\mathcal{N}(P_su) \leq \mathcal{N}(u) + s.$$  

Moreover, by Lemma 6.8 (3), we have $\mathcal{N}(Q_tv) \leq \mathcal{N}(u)$. The claim (1) follows by iteration.

(2) Similar to part (1), we first consider the effect of each operator individually. The operator $P_s$ satisfies the following estimate derived by integration:

$$|u(x) - v(x)| \leq A + B\|x\|_2^2 \implies |P_su(x) - P_sv(x)| \leq A + Bs + B\|x\|_2^2,$$

and this implies that

$$\left\| (P_su - P_sv)/\omega^2 \right\|_{L^\infty} \leq (1 + s)\left\| (u - v)/\omega^2 \right\|_{L^\infty} \leq e^s\left\| (u - v)/\omega^2 \right\|_{L^\infty}.$$  

Now consider the operator $Q_t$. Let $A = \left\| (u - v)/\omega^2 \right\|_{L^\infty}$ and

$$B = 2^{1/2}C \max(\mathcal{N}(u), \mathcal{N}(v), 1)^{1/2} \geq \max(\|Du/\omega\|_{L^\infty}, \|Dv/\omega\|_{L^\infty}).$$

Then we have

$$u(x) \leq v(x) + A(1 + \|x\|_2^2).$$

So we have

$$Q_tu(x) = \inf_y \left[ u(x + y) + \frac{1}{2t}\|y\|_2^2 \right] \leq \inf_y \left[ v(x + y) + \frac{1}{2t}\|y\|_2^2 + A(1 + \|x + y\|_2^2) \right].$$

Write $p = D(Q_tv)(x)$ and recall that $Q_tv(x) = v(x - tp) - \frac{1}{2t}\|tp\|_2^2$. Thus, taking $y = -tp$ on the right hand side above, we have

$$Q_tu(x) \leq Q_tv(x) + A(1 + \|x - tp\|_2^2).$$

38
We also have
\[ \|p\|_2 = \|D(Q_t u)(x)\|_2 \leq \|Du(x)\|_2 \leq B\omega(x). \]
Therefore,
\[
1 + \|x - tp\|_2^2 \leq 1 + \|x\|_2^2 + 2t\|x\|_2\|p\|_2 + t^2 \|p\|_2^2 \\
\leq \omega(x)^2 + 2t\omega(x) \cdot B\omega(x) + t^2 (B\omega(x))^2 \\
= (1 + tB)^2 \omega(x)^2 \\
\leq e^{2tB} \omega(x)^2.
\]
Therefore, \( u \leq v + Ae^{2tB} \omega^2 \) and the same holds with \( v \) and \( u \) reversed, so that we have
\[
\|(Q_t u - Q_t v)/\omega^2\|_{L^\infty} \leq \exp(2tB)\|(u - v)/\omega^2\|_{L^\infty} \\
= \exp(2^{3/2}Ct \max(\mathcal{R}(u), \mathcal{R}(v), 1))\|(u - v)/\omega^2\|_{L^\infty}.
\]
Finally, to prove estimate (2) for \( R = P_{s_1}Q_{t_1} \ldots P_{s_n}Q_{t_n} \), we apply our estimates for \( P_{s_j} \) and \( Q_{t_j} \) iteratively, using (1) to control the growth of the \( R \) terms at each step of the iteration.

(3). For the single operator \( P_s \), we have the estimate
\[
\|D(P_su) - D(P_tv)\|_{L^\infty} \leq \|Du - Dv\|_{L^\infty}
\] (6.25)
as a basic property of convolutions. Now consider the operator \( Q_t \). Observe that
\[
\|D(Q_t u)(x) - D(Q_t v)(x)\|_2 = \|Du(x - tD(Q_t u)(x)) - Dv(x - tD(Q_t v)(x))\|_2 \\
\leq \|Du(x - tD(Q_t u)(x)) - Dv(x - tD(Q_t v)(x))\|_2 \\
+ \|Du(x - tD(Q_t u)(x)) - Du(x - tD(Q_t v)(x))\|_2 \\
\leq \|Du - Dv\|_{L^\infty} + Ct\|D(Q_t u)(x) - D(Q_t v)(x)\|_2,
\]
where the last inequality follows because \( Du \) is \( C \)-Lipschitz. This implies that for \( t < 1/C \),
\[
\|D(Q_t u) - D(Q_t v)\|_{L^\infty} \leq (1 - Ct)^{-1}\|Du - Dv\|_{L^\infty}.
\]
Pick an integer \( k \). The operators \( Q_t \) form a semigroup and we know that \( Q_{t/k}^j u \) and \( Q_{t/k}^j v \) satisfy the same hypotheses. Thus,
\[
\|D(Q_{t/k}^{j+1} u) - D(Q_{t/k}^{j+1} v)\|_{L^\infty} \leq (1 - Ct/k)^{-j}\|D(Q_{t/k}^j u) - D(Q_{t/k}^j v)\|_{L^\infty}.
\]
Thus, by induction, \( \|D(Q_{t/k}^k u) - D(Q_{t/k}^k v)\|_{L^\infty} \leq (1 - Ct/k)^{-k}\|Du - Dv\|_{L^\infty} \), and taking \( k \to \infty \) yields
\[
\|D(Q_t u) - D(Q_t v)\|_{L^\infty} \leq e^{Ct}\|Du - Dv\|_{L^\infty}.
\] (6.26)
Now by iterating (6.25) and (6.26), we obtain the estimate (3).

The next step is to show that the operators \( P_t \) and \( Q_t \) almost commute when \( t \) is small.

**Lemma 6.10.** Suppose that \( 0 \leq Hu \leq C \). Then

1. \( \|(Q_t P_t u - P_t Q_t u)/\omega^2\|_{L^\infty} \leq C^2C^2[1 + 2C \max(\mathcal{R}(u), 1)]. \)
2. \( \|D(Q_t P_t u) - D(P_t Q_t u)\|_{L^\infty} \leq (1 + C^2)^2/[2^{3/2}]. \)
Proof. (1) First, note that $P_t Q_t u \leq Q_t P_t u$ because the average of the infimum is smaller than the infimum of the average (we leave the details to the reader). For the other inequality, we apply Lemma 6.8 (5) to conclude that

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

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

The last quantity is the variance of $Du(x + y)$ when $y$ is chosen randomly according to $\sigma_{t,N}$. But $Du(x)$ is $C$-Lipschitz and $\int |y|^2 d\sigma_{t,N} = t$ and hence

$$0 \leq \int \left[ \| Du(x + y) \|_2^2 - \| D(P_t u)(x) \|_2^2 \right] d\sigma_{t,N}(y) \leq C^2 t.$$

Meanwhile, by Lemma 6.8 (5) again we have

$$P_t u(x) - \frac{t}{2} \| D(P_t u)(x) \|_2^2 \geq Q_t P_t u(x) - \frac{Ct^2}{2} \| D(P_t u)(x) \|_2^2$$

$$\geq Q_t P_t u(x) - C^2 t^2 \| x \|_2^2$$

Altogether,

$$|P_t Q_t u(x) - Q_t P_t u(x)| \leq \frac{C^2 t^2}{2} + C^3 t^2 \| x \|_2^2,$$

and the asserted estimate (1) follows.

(2) Note that

$$D(Q_t P_t u)(x) = D(P_t u)(x - t D(Q_t P_t u)(x)) = \int Du(x + y - t D(Q_t P_t u)(x)) d\sigma_{t,n}(y).$$

On the other hand,

$$D(P_t Q_t u)(x) = \int D(Q_t u)(x + y) d\sigma_{t,n}(y) = \int Du(x + y - t D(Q_t u)(x + y)) d\sigma_{t,n}(y).$$

Because $Du$ is $C$ Lipschitz, we have

$$\| D(Q_t P_t u)(x) - D(P_t Q_t u)(x) \|_2 \leq C t \int \| D(Q_t u)(x + y) - D(Q_t P_t u)(x) \|_2 d\sigma_{t,n}(y).$$

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 produces

$$\int \| D(Q_t u)(x + y) - D(Q_t u)(x) \|_2 d\sigma_{t,n}(y) \leq C \int |y| d\sigma_{t,n} \leq C t^{1/2}$$

Meanwhile, the second term is independent of $y$, and thus integrating yields

$$\| D(Q_t u)(x) - D(Q_t P_t u)(x) \|_2 \leq e^{C t} \| Du - D(P_t u) \|_{L^\infty} \leq C e^{C t^{1/2}},$$

40
where the last inequality follows because
\[
|D(P_t u)(x) - Du(x)| \leq \int \|Du(x + y) - Du(x)\|_2 d\sigma_{t,n}(y)
\]
\[
\leq \int C\|y\|_2 d\sigma_{t,n}(y) \leq C \left( \int \|y\|^2 d\sigma_{t,N}(y) \right)^{1/2} = Ct^{1/2}.
\]

Altogether, we obtain
\[
\|D(Q_t P_t u)(x) - D(P_t Q_t u)(x)\|_2 \leq (1 + e^{Ct})C^2 t^{3/2}.
\]

Finally, we can construct the semigroup \( R_t \) for dyadic values of \( t \).

**Lemma 6.11.** For \( t \in Q^+ \) and \( u \in E_C \), the limit \( R_t u = \lim_{\ell \to \infty} R_{t,\ell} u \) exists, and we have for \( t \in 2^{-\ell} \mathbb{N} \) that

1. \( \|(R_{t,\ell}u - R_t u)/\omega^2\|_{L^\infty} \leq \frac{C^2 \delta}{2} \exp\left[ t + (2C^{1/2} + 2^{3/2}Ce \max(N(u) + 2t, 1)) \right] \].

2. \( \|D(R_{t,\ell}u) - D(R_t u)\|_{L^\infty} \leq \frac{Ce^{C^2 t^{-\ell/2}}}{1 - 2^{-\ell/2}} \).

**Proof.** We will show that the sequence \( R_{t,\ell} \) is Cauchy by estimating \( \|(R_{t,\ell+1}u - R_{t,\ell}u)/\omega^2\|_{L^\infty} \).

Suppose that \( t = 2^{-\ell} n \). Let us write for shorthand \( \delta = 2^{-\ell - 1} \) and

\[
A = (P_3 Q_\delta)^2 = P_3 Q_\delta P_3 Q_\delta
\]
\[
B = P_{2\delta} Q_{2\delta} = P_3 P_3 Q_\delta Q_\delta.
\]

Recalling the definition of \( R_{t,\ell} \), we write the telescoping series identity

\[
R_{t,\ell+1} u - R_{t,\ell} u = \sum_{j=0}^{n-1} [A^{j+1} B^{n-j-1} u - A^j B^{n-j} u].
\]

Now by Lemma 6.9 (2) and (1),

\[
\|(Av - Bv)/\omega^2\|_{L^\infty} \leq \frac{\epsilon^\delta}{2} \|(Q_3 P_3 Q_\delta v - P_3 Q_\delta Q_\delta v)/\omega^2\|_{L^\infty}
\]
\[
\leq \frac{\epsilon^\delta C^2}{2} (1 + 2C \max(N(v), 1))
\]

We apply this to \( v = B^{n-j-1} u \) together with the fact that \( \mathcal{N}(B^{n-j-1} u) \leq \mathcal{N}(u) + 2\delta(n - j - 1) \) to obtain

\[
\|(AB^{n-j-1} u - B^{n-j} u)/\omega^2\|_{L^\infty} \leq \frac{\epsilon^\delta C^2 \delta^2}{2} (1 + 2C \max(\mathcal{N}(u) + 2\delta(n - j - 1), 1))
\]
\[
\leq \frac{\epsilon^\delta C^2 \delta^2}{2} (1 + 2C \max(\mathcal{N}(u) + t, 1))
\]

Finally, after applying the operator \( A^j \) and using Lemma 6.9 (2), we have

\[
\|(AB^{n-j-1} u - B^{n-j} u)/\omega^2\|_{L^\infty}
\]
\[
\leq \frac{C^2 \delta^2}{2} (1 + 2C \max(\mathcal{N}(u) + t, 1)) \exp(\delta + 2j\delta + 2^{3/2}C(2j\delta) \max(\mathcal{N}(u) + 2\delta j, 1))
\]
\[
\leq \frac{C^2 \delta^2}{2} (1 + 2C \max(\mathcal{N}(u) + t, 1)) \exp(t + 2^{3/2}Ct \max(\mathcal{N}(u) + t, 1))
\]
\[
\leq \frac{C^2 \delta^2}{2} \exp[t + (2C^{1/2} + 2^{3/2}Ct) \max(\mathcal{N}(u) + t, 1)].
\]

41
Now we sum from \( j = 0 \) to \( n - 1 \) and substitute \( t = 2\delta n \) and \( \delta = 2^{-t-1} \) to obtain
\[
\| (R_{t,\ell+1}u - R_{t,\ell}u) / \omega \|_{L^\infty} \leq \frac{C^2 t}{2t+3} \exp[t + (2C^{1/2} + 2^{3/2}Ct) \max(\mathcal{N}(u) + t, 1)^{1/2}].
\]
This implies that the sequence \( R_{t,\ell}u \) is Cauchy in \( \| \cdot / \omega \|_{L^\infty} \) and hence converges to some limit \( Rtu \) with the estimate
\[
\| (R_{t,\ell}u - R_{t,\ell}u) / \omega^2 \|_{L^\infty} \leq \frac{2C^2 t}{2t+3} \exp[t + (2C^{1/2} + 2^{3/2}Ct) \max(\mathcal{N}(u) + t, 1)^{1/2}].
\]

Now we turn to the proof of (2) using the same notation from the proof of (1). Using Lemma 6.9 (3) and Lemma 6.10 (2), we obtain
\[
\| D(Av) - D(Bv) \|_{L^\infty} \leq (1 + e^{C\delta}) C^2 \delta^{3/2}.
\]
We apply this to \( v = B^{n-j-1}u \) together with Lemma 6.9 (3) to conclude that
\[
\| D(A^{j+1}B^{n-j-1}u) - D(A^j B^{n-j}u) \|_{L^\infty} \leq e^{C\delta j} \| D(AB^{n-j-1}u) - D(B^{n-j}u) \|_{L^\infty} \\
\leq (1 + e^{C\delta}) C^2 \delta^{3/2} e^{2C\delta j} \leq 2e^{C\delta} C^2 \delta^{3/2} e^{2C\delta j}.
\]
Then we sum from \( j = 0, \ldots, n - 1 \) and obtain
\[
\| D(R_{t,\ell+1}u) - D(R_{t,\ell}u) \|_{L^\infty} \leq 2C^2 \delta^{3/2} e^{2C\delta n} \| C^2 \delta^{3/2} e^{2C\delta n} - 1 \\
\leq C^2 \delta^{3/2} e^{2C\delta n} = C 2^{-(\ell+1)/2} e^t,
\]
where we have applied the fact that
\[
\frac{e^{C\delta}}{e^{2C\delta} - 1} = \frac{1}{2 \sinh C\delta} \leq \frac{1}{2C\delta}.
\]
As a consequence, the sequence \( D R_{t,\ell}u \) is Cauchy and converges in \( L^\infty \) to some function. We already know that \( R_{t,\ell}u \) converges to \( Rtu \), so the limit of \( D(R_{t,\ell}u) \) must be \( D(Rtu) \). We also have
\[
\| D(R_{t,\ell}u) - D(R_{t,\ell}u) \|_{L^\infty} \leq C \frac{2^{-t/2}}{1 - 2^{-1/2}} e^t.
\]

Now in order to extend the semigroup \( R_t \) to all real values of \( t \), we want estimates to show that it depends continuously on \( t \).

**Lemma 6.12.** Let \( s \leq t \) be two numbers in \( \mathbb{Q}_2 \) and \( u \in \mathcal{E}_C \). Then
1. \( \| (R_t u - R_s u) / \omega^2 \|_{L^\infty} \leq \frac{1}{2} \mathcal{C}(t-s) + C^2(t-s) \max(\mathcal{N}(u) + t, 1)^2 \).
2. We have
\[
\| (D(R_t u) - D(R_s u)) / \omega \|_{L^\infty} \leq C e^{C^2 t^{1/2} (t-s)^{1/2}} + \frac{C^2 t^{1/2} (t-s)^{1/2}}{1 - 2^{-1/2}} + 4C^3(t-s) \max(\mathcal{N}(u) + t, 1).
\]
Proof. (1) First observe that
\[ \|P_t u - u\|_{L^\infty} \leq Ct/2 \] (6.27)
which follows from integrating
\[ \|u(x + y) - [u(x) + (Du(x), y)]\|_2 \leq \frac{C}{2} \|y\|^2 \]
with respect to \( \sigma_t, \kappa(y) \). Moreover, using Lemma 6.8 (5) and the fact that \( Q_t u \leq u \), we have
\[ \|Q_t u - u\|_{L^\infty} \leq \frac{t}{2} \|Du/\omega\|_{L^\infty} \leq C^2 t \max(\mathbb{N}(u), 1)^2. \] (6.28)

Now by iterating (6.27) and (6.28) with a telescoping series argument, we get
\[ \|(R_{t, \ell} u - u)/\omega^2\|_{L^\infty} \leq \frac{Ct}{2} + C^2 t \max(\mathbb{N}(u) + t, 1)^2. \]
Hence, for \( t > s \),
\[ \|(R_{t-s, \ell} R_{s, \ell} u - R_{s, \ell} u)/\omega^2\| \leq \frac{C(t-s)}{2} + C^2 (t-s) \max(\mathbb{N}(u) + t, 1)^2, \]
and (1) follows by taking \( \ell \to \infty \).

(2) From the fact that \( Du \) is \( C \)-Lipschitz, we have
\[ \|D(P_t u) - Du\|_{L^\infty} \leq Ct^{1/2}, \] (6.29)
while from Lemma 6.8 (4), we have
\[ \|(D(Q_t u) - Du)/\omega\|_{L^\infty} \leq Ct \|Du/\omega\|_{L^\infty} \leq 2C^3 t \max(\mathbb{N}(u), 1). \] (6.30)

In order to prove (2) in the case where \( t = 2^{-\ell} \) and \( s = 0 \), we write
\[ D(R_t u) - Du = [D(R_t u) - D(P_t u)] + [D(P_t u) - D(Q_t u)] + [D(Q_t u) - Du]. \]

We estimate the first term using Lemma 6.11 and the last two terms by (6.29) and (6.30) to obtain
\[ \|(D(R_t u) - Du)/\omega\|_{L^\infty} \leq \frac{Ce^{Ct^{1/2}}}{1 - 2^{-1/2}} + Ct^{1/2} + 2C^3 t \max(\mathbb{N}(u), 1). \]

In the general case, let us write \( t > s \) in a binary expansion
\[ t = s + \sum_{j=m+1}^{n} a_j 2^{-j}, \]
where \( a_j \in \{0, 1\} \), and let \( t_k = s + \sum_{j=m+1}^{k} a_j 2^{-j} \). Then we estimate \( \|(D(R_t u) - Du)/\omega\|_{L^\infty} \) by
\[ \leq \sum_{j=m+1}^{n} \|(D(R_{t, \ell} u) - D(R_{t-j, \ell} u))/\omega\|_{L^\infty} \]
\[ \leq \sum_{j=m+1}^{n} \left( \frac{Ce^{Ct_j 2^{-j/2}}}{1 - 2^{-1/2}} + C^2 2^{-j} + 2C^3 \cdot 2^{-j} \max(\mathbb{N}(u) + t_{j-1}, 1) \right) \]
\[ \leq \frac{Ce^{Ct^2/m/2}}{(1 - 2^{-1/2})^2} + \frac{C^2 2^{-m/2}}{1 - 2^{-1/2}} + 2C^3 2^{-m} \max(\mathbb{N}(u) + t, 1) \]
\[ \leq \frac{Ce^{Ct^2/2(t-s)^1/2}}{(1 - 2^{-1/2})^2} + \frac{C^2 2^{1/2}(t-s)^1/2}{1 - 2^{-1/2}} + 4C^3 (t-s) \max(\mathbb{N}(u) + t, 1). \]
\[ \square \]
Proof of Theorem 6.3. Note that $R_{t,\ell} R_{s,\ell} = R_{s+t,\ell}$ for $s, t \in 2^{-\ell} \mathbb{N}$. The semigroup property is preserved when we take the limit as $\ell \to \infty$ because we can control the errors using Lemma 6.9. By Lemma 6.12, the semigroup $R_t$ depends continuously on $t$ with an explicit estimate, and therefore, we can extend the definition of $t$ to all $t \geq 0$. Moreover, all the estimates that we proved for $R_{t,\ell}$ pass in the limit to estimates on $R_t$ for $t \in \mathbb{Q}^+$ and extend to $\mathbb{R}^+$.

Claim (1) of the Theorem therefore follows from Lemma 6.11. The estimates in the Theorem are written in a simpler and less explicit form obtained from estimating $N(u)$ by a constant times $\|Du/\omega\|_{L^\infty}$ and applying sum-max and AM-GM inequalities. Similarly, claims (2a) and (2b) of the Theorem follow from Lemma 6.9 and claims (2c) and (2d) follow from Lemma 6.12.

It remains to show claim (3) of the Theorem, namely, that $R_t u_0$ is a viscosity solution. Suppose that we have a Taylor approximation from below at the point $(x_0, t_0)$ with $t_0 > 0$, so that

$$u(x_0 + h, t_0 + \tau) \geq u(x_0, t_0) + \alpha \tau + \langle p, h \rangle_2 + \frac{1}{2} \langle Ah, h \rangle_2 + o(\|h\|^2_2 + \|\tau\|_\infty).$$

(6.31)

Note that in this case $p = Du(x_0, t_0)$.

We substitute $\tau = -\delta$ where $\delta = 2^{-\ell}$ for some $\ell$. Note that by Lemma 6.11 and Lemma 6.10,

$$u(x_0, t_0) = [P_0 Q_\delta u(\cdot, t_0 - \delta)](x_0) + O(\delta^2) = Q_\delta P_0[u(\cdot, t_0 - \delta)](x_0) + O(\delta^2).$$

Because $u(x) \leq C(1 + \|x\|^2_2)$ and $\sigma_{\delta, N}$ has finite moments, we know that for any $r > 0$,

$$P_0[u(\cdot, t_0 - \delta)](x_0) \geq \int_{\|y\|^2_2 \leq r} u(x_0 + y, t_0 - \delta) \, d\sigma_{\delta, N}(y) + o(\delta).$$

By choosing $r$ small enough and using (6.31), we obtain

$$P_0[u(\cdot, t_0 - \delta)](x_0) \geq \int_{\|y\|^2_2 \leq r} \left[ u(x_0, t_0) - \alpha \delta + \langle p, h + y \rangle_2 + \frac{1}{2} \langle Ah + y, h + y \rangle_2 + o(\|h + y\|^2_2) \right] \, d\sigma_{\delta, N}(y) + o(\delta)$$

$$= u(x_0, t_0) - \alpha \delta + \frac{\delta}{2N} \langle A h + y, h + y \rangle_2 + \frac{1}{2} \langle Ah, h \rangle_2 + o(\|h\|^2_2 + \delta)$$

Letting $q = D(Q_\delta P_0 u(\cdot, t_0 - \delta))(x_0)$, we have by Lemma 6.8 (2),

$$Q_\delta P_0[u(\cdot, t_0 - \delta)](x_0) = P_0[u(x_0 - \delta q, t_0 - \delta)] + \frac{\delta}{2} \|q\|_2^2.$$

On the other hand, by Lemmas 6.10, 6.11 and 6.12

$$q = Du(x_0, t_0 - \delta) + O(\delta^{1/2}) = Du(x_0, t_0) + O(\delta^{1/2}) = p + O(\delta^{1/2}).$$

Therefore, we have

$$Q_\delta P_0[u(\cdot, t_0 - \delta)](x_0) = P_0[u(x_0 - \delta p, t_0 - \delta)] + \frac{\delta}{2} \|p\|_2^2 + O(\delta^{3/2}),$$

where we have applied the fact that $P_0[u(\cdot, t_0 - \delta)](x)$ is locally Lipschitz in $x$ with estimates that are independent of $\delta$ because of Lemma 6.9 (1). In light of the lower Taylor expansion for $P_0 u$, this implies that

$$Q_\delta P_0[u(\cdot, t_0 - \delta)](x_0) \geq u(x_0, t_0) - \alpha \delta + \frac{\delta}{2N} \langle A h, h \rangle_2 + \frac{1}{2} \langle -Ap, -p \rangle_2 + o(\delta)$$

44
We conclude that
\[
u(x_0, t_0) \geq u(x_0, t_0) - \alpha \delta + \frac{\delta}{2N} \text{Tr}(A) - \frac{\delta}{2} ||p||_2^2 + o(\delta)
\]
and hence
\[
\alpha - \frac{1}{2N} \text{Tr}(A) + \frac{1}{2} ||p||_2^2 \geq 0,
\]
which implies that \(u(x, t)\) is a supersolution at the point \((x_0, t_0)\). A completely symmetrical argument shows that \(u(x, t)\) is a subsolution as well.

It follows from Lemma 6.8 (4) that the distance from the fixed point \(D\) from Lemma 3.18. But if \(u\) is asymptotically approximable by trace polynomials. Then for every \(u_0 \in \mathcal{E}_C\), and let \(u(x) = u_0(x) + (\varepsilon/2)||x||_2^2\). Then \(u\) is bounded below and in \(\mathcal{E}_{C+\varepsilon}\). Therefore, we have
\[
\exp(-N^2R_tu_t) = P_t[\exp(-N^2u_t)].
\]
The continuity estimates (1a) of this theorem show that \(R_tu_t \to R_tu\). By monotonicity of \(R_t\), we have \(R_tu_t \searrow R_tu\), and hence by the monotone convergence theorem \(P_t[\exp(-N^2u_t)] \nearrow P_t[\exp(-N^2u_t)]\). Thus, \(\exp(-N^2R_tu_t) = P_t[\exp(-N^2u_t)]\). □

6.4 Approximation by Trace Polynomials

Now we are ready to prove that \(R_t\) preserves approximability by trace polynomials.

**Proposition 6.13.** Let \(\{V_N\}\) be a sequence of functions \(M_N(C)^m_{sa} \to \mathbb{R}\) 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 \(t > 0\), the sequences \(\{D(P_tV_N)\}, \{D(Q_tV_N)\}\), and \(\{D(R_tV_N)\}\) are asymptotically approximable by trace polynomials.

**Proof.** The fact that \(\{D(P_tV_N)\}\) is asymptotically approximable by trace polynomials follows from Lemma 6.18.

Now consider \(D(Q_tV_N)\). Note that by Lemma 6.8 (2), \(D(Q_tV_N)(x)\) is the solution of the fixed point equation
\[
y = DV_N(x - ty).
\]
But if \(t < 1/C\), then \(y \to DV_N(x - ty)\) is a contraction and thus iterates of this function will converge to the fixed point. Let us define \(\phi_{0,N}(x) = 0\) and \(\phi_{N,t+1}(x) = DV_N(x - t\psi_{k,t+1}(x))\).

By Lemma 6.18 (4), the distance from the fixed point \(D(Q_tV_N)(x)\) from 0 is bounded by
\[
t||DV_N(x)||_2,\text{ hence,}
\]
\[
||\phi_{N,t}(x) - D(Q_tV_N)(x)||_2 \leq C't^{t+1}||DV_N(x)||_2.
\]
A straightforward induction shows that \(\phi_{N,t}\) is Lipschitz in \(||\cdot||_2\) and hence using Lemma 3.17
\(\{\phi_{N,t}\}_N\) is asymptotically approximable by trace polynomials.

Now \(||DV_N(0)||_2\) is bounded by some constant \(A\) as \(k \to \infty\) because \(DV_N\) is asymptotically approximable by trace polynomials. Since \(DV_N\) is also C-Lipschitz, \(||DV_N(x)||_2 \leq A + C||x||_2\). In particular, \(||\phi_{N,t}(x) - D(Q_tV_N)(x)||_2 \leq C't^{t+1}(A + C||x||_2)\). Thus, by Lemma 3.16 \(\{D(Q_tV_N)\}\) 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 \(\{Q_tV_N\}\) is asymptotically approximable by trace polynomials.
For the sequence \( \{ D(R_tV_N) \} \), first note that when \( t \in \mathbb{Q}^+_0 \), we know \( \{ D(R_t\ell V_N) \} \) is asymptotically approximable by trace polynomials. By Theorem 6.3 (1c) and Lemma 3.16, the sequence \( \{ D(R_tV_N) \} \) is asymptotically approximable by trace polynomials for all \( t \in \mathbb{Q}^+_0 \). Finally, by Theorem 6.3 (2d) and Lemma 3.16, the sequence \( \{ D(R_tV_N) \} \) is asymptotically approximable by trace polynomials for all \( t \in \mathbb{R}^+ \).

\[
\square
\]

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 \( \mu_N \) be a sequence of probability measures on \( M_N(\mathbb{C})_{sa} \) 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 \( \mu_N \) concentrates around some non-commutative law \( \lambda \).

(C) For some \( R_0 > 0 \), we have \( \lim_{N \to \infty} \int_{\|x\|_2 \geq R_0} (1 + \|x\|_2^2) \, d\mu_N(x) = 0 \).

(D) The sequence \( \{ DV_N \} \) is asymptotically approximable by trace polynomials.

Then

(1) The law \( \lambda \) has finite Fisher information \( \Phi^*(\lambda) \), and for all \( t > 0 \), we have

\[
\lim_{N \to \infty} \frac{1}{N^3} \sum_{\ell=1}^{N} \mathcal{L}(\mu_N * \sigma_{t,N}) \to \Phi^*(\lambda \boxplus \sigma_t).
\]

(2) We have for all \( t > 0 \),

\[
\chi(\lambda \boxplus \sigma_t) = \chi(\lambda \boxplus \sigma_t) = \lim_{N \to \infty} \frac{1}{N^2} \left( h(\mu_N * \sigma_{t,N}) + \frac{m}{2} \log N \right) = \chi^*(\lambda \boxplus \sigma_t).
\]

(3) For each \( T > 0 \), the functions \( t \mapsto \frac{1}{N^3} \mathcal{L}(\mu_N * \sigma_{t,N}) \) and \( t \mapsto \Phi^*(\lambda \boxplus \sigma_t) \) are Lipschitz on \([0, T]\) with a Lipschitz norm that only depends on \( T, C \), sup \( N \| DV_N(0) \|_2 \), and sup \( N \int_0^1 \|x\|_2^2 \, d\mu_N(x) \).

**Remark 7.2.** Note that 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 7.1 implies that \( \mu_N \) satisfies the hypotheses of Theorem 7.1 for some law \( \lambda \).

In preparation for the proof of the theorem, we have already verified that the hypotheses are preserved under Gaussian convolution with the exception of (B). This is straightforward, the only subtlety being that we have not assumed that \( \mu_N \) has finite moments.

**Lemma 7.3.** Suppose that \( \{ \mu_N \} \) concentrates around a non-commutative law \( \lambda \). Then \( \{ \mu_N * \sigma_{t,N} \} \) concentrates around \( \lambda \boxplus \sigma_t \) for every \( t > 0 \).

**Proof.** Fix \( t \). Let \( X_N = (X_{N,1}, \ldots, X_{N,m}) \) and \( Y_N = (Y_{N,1}, \ldots, Y_{N,m}) \) be independent random variables with the laws \( \mu_N \) and \( \sigma_{t,N} \) respectively. Because the topology on \( \Sigma_m \) is generated
by non-commutative moments, it suffices to show that for each non-commutative polynomial \( p \) and \( \delta > 0 \)

\[
\lim_{k \to \infty} P(|\tau_N(p(X_N + Y_N)) - \lambda \oplus \sigma_t(p)| \geq \delta) = 0.
\]

We can choose a neighborhood \( \mathcal{V} \) of \( \lambda \) and a \( C > 0 \) such that \( \lambda_x \in \mathcal{V} \) implies that \( |\tau_N(q(x))| \leq C \) for every monomial \( q \) of degree \( \leq 2 \text{deg}(p) \). In particular, there is some constant \( L \) such that \( x \in \Gamma_N(\mathcal{V}) \) implies that \( \tau_N(p(x + y)) \) is an \( L \)-Lipschitz function of \( y \) with respect to \( \| \cdot \|_2 \) on the region \( \{ y : \|y\| \leq 4t^{1/2} \} \).

Choose a \( C_c^\infty \) 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 \( \Psi : (y_1, \ldots, y_m) \mapsto (\psi(y_1), \ldots, \psi(y_m)) \) is globally Lipschitz in \( \| \cdot \|_2 \) and maps into the operator norm ball of radius \( 4t^{1/2} \). This implies that there is some constant \( L' \) such that \( y \mapsto \tau_N(p(x, \Psi(y))) \) is \( L' \)-Lipschitz for all \( x \in \Gamma_N(\mathcal{V}) \). Let

\[
\begin{align*}
\alpha_N(x) &= E[\tau_N(p(x, \Psi(Y)))] \\
\beta_N(x) &= E[\tau_N(p(x, Y))] = \exp(tL_N/2)[\tau(p)](x) \\
\beta(x) &= \exp(tL/2)[\tau(p)](x)
\end{align*}
\]

Therefore, by Theorem 2.8 applied to \( Y_N \),

\[
x \in \Gamma_N(\mathcal{V}) \implies P(|\tau_N(p(x, \Psi(Y_N))) - \alpha_N(x)| \geq \delta/3) \leq e^{-\delta N^2/6(L')^2}
\]

On the other hand, we know by standard tail estimates on GUE (see Corollary 2.10) that

\[
\lim_{k \to \infty} E[\tau_N(q(Y_N))_1|\|Y_N\| \geq 3t^{1/2}] = 0
\]

for every non-commutative 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 3.13,

\[
\beta_N(x) = \exp(tL_N/2)[\tau(p)](x) \to \exp(tL/2)[\tau(p)](x) = \beta(x)
\]

where the convergence occurs coefficient-wise and therefore uniformly for \( x \in \Gamma_N(\mathcal{V}) \). Thus, if \( N \) is large enough that \( |\alpha_N(x) - \beta_N(x)| \leq \delta/3 \) for \( x \in \Gamma_N(\mathcal{V}) \), then we have

\[
P\left(|\tau_N(p(X_N + Y_N)) - \beta(X_N)| \geq 2\delta/3, X_N \in \Gamma_N(\mathcal{V}), \|Y_N\| \lesssim 3t^{1/2}\right) \leq e^{-\delta N^2/6(L')^2},
\]

where we have applied the Fubini-Tonelli theorem for the product measure \( \mu_N \otimes \sigma_t, N \). By our concentration assumption,

\[
P\left(|\tau_N(\beta(X_N)) - \lambda[\beta]| \geq \delta/3\right) \to 0, \quad P(X_N \in \Gamma_N(\mathcal{V})) \to 1,
\]

and by Corollary 2.10 also \( P(\|Y_k\| \geq 3t^{1/2}) \to 0 \). Altogether, we have

\[
P\left(|\tau_N(p(X_N + Y_N)) - \lambda[\beta]| \geq \delta\right) \to 0.
\]

But note that \( \lambda[\beta] = \lambda[\exp(tL/2)[\tau(p)]] = \lambda \oplus \sigma_t[p] \) by Lemma 3.13.

\[\square\]

\textit{Proof of Theorem 7.1.} Let \( V_{N,t} = R_t V_N \) be the potential associated to \( \mu_N \ast \sigma_t N \). Let us verify that \( V_{N,t} \) satisfies the assumptions (A) - (D) for every \( t > 0 \).

(A) We showed in 6.3 that \( V_{N,t} \) also satisfies (A) because it is in the space \( \mathcal{E}_C \).
(B) This follows from Lemma 7.3.

(C) This follows from tail bounds on the GUE (Corollary 2.10).

(D) This follows from Proposition 6.13.

Now claim (1) of the theorem follows by applying Proposition 5.7 to \( \mu_N \ast \sigma_{t,N} \) with \( n = 2 \).

For claim (2), recall that by Lemma 5.4, equation (5.6),

\[
\frac{1}{N^2} h(\mu) + \frac{m}{2} \log N = \frac{1}{2} \int_0^\infty \left( \frac{m}{1 + t} - \frac{1}{N^4} I(\mu \ast \sigma_{t,N}) \right) ds + \frac{1}{2} \log 2\pi e.
\]

Because of the upper and lower bounds (5.4) for the Fisher information, we can apply the dominated convergence theorem to take the limit as \( N \to \infty \) inside the integral and apply claim (1) to conclude that

\[
\lim_{N \to \infty} \left( \frac{1}{N^2} h(\mu_N) + \frac{m}{2} \log N \right) = \chi^*(\lambda).
\]

On the left hand side, we will apply Proposition 5.3 with \( n = 2 \). We may replace \( V_N \) by \( V_N - V_N(0) \) without changing \( \mu_N \).

Then because \( \{DV_N\} \) is asymptotically approximable by trace polynomials, we know that \( \{V_N\} \) is asymptotically approximable by trace polynomials (Lemma 3.19). Therefore, the hypotheses of Proposition 5.3 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 \( \chi(\lambda) \). Moreover, this holds for \( \mu_N \ast \sigma_{t,N} \) just as well as \( \mu_N \) because \( \mu_N \ast \sigma_{t,N} \) satisfies the same assumptions (A) - (D).

For claim (3), first fix \( N \) and let \( X \) be a random variable with law \( \mu_N \), and let \( Y_t \) be an independent Hermitian Brownian motion (here \( Y_t \sim \sigma_{t,N} \)). Let \( \Xi_t \) be the conjugate variable \( DV_N(X + Y_t) \). Then

\[
\frac{1}{N^2} I(\mu \ast \sigma_{t,N}) = E\|\Xi_t\|_2^2.
\]

Suppose \( 0 \leq s \leq t \leq T \). Then using Theorem 6.3 (2d) and the fact that \( DV_{N,t} \) is \( C \)-Lipschitz,

\[
\|\Xi_s - \Xi_t\|_2 \leq \|DV_{N,s}(X + Y_s) - DV_{N,t}(X + Y_s)\|_2 + \|DV_{N,t}(X + Y_s) + DV_{N,t}(X + Y_t)\|_2 \\
\leq \text{const}(T, \|DV_{N}/\omega\|_{L^\infty})|t - s|^{1/2} \omega(X + Y_s) + C\|Y_s - Y_t\|_2.
\]

Now we can bound \( \|DV_N/\omega\|_{L^\infty} \) in terms of \( \|DV_N(0)\|_2 \) and \( C \). Thus after some easy estimation,

\[
E\|\Xi_s - \Xi_t\|_2^2 \leq \text{const}|s - t|,
\]

where the constant only depends on \( T, C, \sup_N \|DV_N(0)\|_2 \), and \( \sup_N \|x\|_2^2 d\mu_N(x) \). But recall that \( \Xi_t = E[\Xi_s | X + Y_t] \) and hence

\[
E[\|\Xi_s\|_2^2 - \|\Xi_t\|_2^2] = E\|\Xi_s - \Xi_t\|_2^2 \leq \text{const}|s - t|.
\]

\( \square \)
8 Free Gibbs States

In the situation of Theorem 4.1, we want to interpret the law \(\lambda\) as the free Gibbs state for a potential which is the limit of the \(V_N\)'s. To this end, we will define a non-commutative function space where each point is a limit of functions on \(M_N(\mathbb{C})_{sa}^m\). 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.

8.1 Asymptotic Approximation and Function Spaces

Let \(Y_\bullet = \{ Y_N \}\) be a sequence of normed vector spaces. We define a (possibly infinite) semi-norm on sequences \(\phi_\bullet = \{ \phi_N \}\) of functions \(M_N(\mathbb{C})_{sa}^m \rightarrow Y_N\) by

\[
\| \phi_\bullet \|_{R,Y_\bullet} = \limsup_{N \rightarrow \infty} \sup_{\|x\| \leq R} \| \phi_N(x) \|_{Y_N}.
\]

Let \(F_m(Y_\bullet)\) 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 \(F_m(Y_\bullet)\) with the topology generated by the seminorms \(\| \cdot \|_{R,Y_\bullet}\), or equivalently given by the metric

\[
d_{F_m(Y_\bullet)}(\phi_\bullet, \psi_\bullet) = \sum_{n=1}^{\infty} \frac{1}{2^n} \min(\| \phi_\bullet - \psi_\bullet \|_{n,Y}, 1).
\]

(8.1)

Note that \(F_m(Y_\bullet)\) is a complete metric space in this metric and is a locally convex topological vector space.

The vector space of scalar-valued trace polynomials \(\text{TrP}_m^0\) embeds into \(F_m^0 := F_m(\mathbb{C})\) by the map that sends a trace polynomial to the corresponding sequence of functions it defines on \(M_N(\mathbb{C})_{sa}^m\). A sequence \(\phi_\bullet\) is asymptotically approximable by trace polynomials if and only if \([\phi_\bullet]\) is in the closure of \(\text{TrP}_m^0\) in \(F_m^0\), which we will denote by \(T_m^0\).

Similarly, let \(M_m(\mathbb{C})\) be the sequence \(\{ M_N(\mathbb{C})_{sa}^m \}\) equipped with \(\| \cdot \|_{2}\). The vector space \(\text{TrP}_m^1\) embeds into \(F_m^1 := F_m(M_m(\mathbb{C}))\). A sequence \(\phi_\bullet\) of functions \(M_N(\mathbb{C})_{sa}^m \rightarrow M_N(\mathbb{C})_{sa}^m\) is asymptotically approximable by trace polynomials if and only if \([\phi_\bullet]\) is in the closure of \(\text{TrP}_m^1\), which we denote by \(T_m^1\).

The spaces \(T_m^0\) and \(T_m^1\) can be viewed as non-commutative function spaces through the following alternative characterization. Here \(\mathcal{R}\) denotes the hyperfinite \(I_1\) factor and \(\mathcal{R}^\omega\) denotes its ultrapower. For background, see [6, §1.6 and §5.4] or [5, p. 5 - 7].

**Lemma 8.1.** Let \(f \in \text{TrP}_m^0\). Then we have

\[
\limsup_{N \rightarrow \infty} \frac{1}{\|x\| \leq R} \sup_{\|x\| \leq R} |f(x)| = \sup_{\|x\| \leq R} \frac{1}{\|x\| \leq R} \sup_{\|x\| \leq R} |f(x)| = \sup_{\|x\| \leq R} |f(x)|. \quad (8.2)
\]

If we denote the common value by \(\| f \|_{T_m^0(R)}\), then this family of seminorms defines a metrizable topology on \(\text{TrP}_m^0\) with the metric given as in (8.1), and \(T_m^0\) is the completion of \(\text{TrP}_m^0\) in this metric. The same result holds for \(T_m^1\) using the seminorm

\[
\limsup_{N \rightarrow \infty} \frac{1}{\|x\| \leq R} \sup_{\|x\| \leq R} \|f(x)\|_2 = \sup_{\|x\| \leq R} \frac{1}{\|x\| \leq R} \sup_{\|x\| \leq R} \|f(x)\|_2 = \sup_{\|x\| \leq R} \|f(x)\|_2. \quad (8.3)
\]
Passing to the completion with respect to the metric defined as in \((8.1)\), we have a map \(g \in \text{the algebra of trace polynomials}\) is self-adjoint and separates points in \(\Sigma\) by the Stone-Weierstrass theorem, trace polynomials are dense in \(C(\Sigma)\). For a scalar-valued trace polynomial \(f\), and let \(M_N(\mathbb{C})\) be the completion of the trace polynomials with respect to the metric \(\|\cdot\|\). Then there exists \(x_0 \in M_N(\mathbb{C})\) such that \(\|x_0\| \leq R\) and \(\|x_0 - y\| \leq 1/2^n\) and \(\lim_{n \to \infty} N_n = +\infty\). Then \(x = \lim_{n \to \infty} y_n\) and \(|f(x)| = \lim_{n \to \infty} |f(y_n)| \leq A\). 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 \(F_m^0\).

From this point of view, any \(f \in T^0_m\) has a canonical sequence that represents its equivalence class in \(F_m^0\) constructed as follows. If we write \(f\) as the limit of a sequence of trace polynomials \(f_k\), then \(f_k\mid_{M_N(\mathbb{C})^m}\) converges locally uniformly on \(M_N(\mathbb{C})^m\) as \(k \to \infty\) and the limit is independent of the approximating sequence \(f_k\). We can therefore define \(f\mid_{M_N(\mathbb{C})^m}\) to be this limit. Similarly, \(f\) defines a function on \((R^\omega)^m\) and on \(\mathbb{M}_m^\omega\) for any von Neumann algebra \(M\) which embeds into \(R^\omega\). The analogous observations hold for \((\text{TrP}_m^1)^m\).

In the scalar-valued case, we have yet another characterization:

**Lemma 8.2.** Let \(\Sigma_{m,bdd} = \bigcup_{R > 0} \Sigma_{m,R}\). Let \(C(\Sigma_{m,bdd})\) be the space of functions \(g : \Sigma_{m,bdd} \to \mathbb{C}\) such that \(g \in C(\Sigma_{m,R})\) for all \(R\) equipped with the family of seminorms \(\|\cdot\|_{C(\Sigma_{m,R})}\). Then \(T^0_m\) is isomorphic to \(C(\Sigma_{m,bdd})\) 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 \(g \in C(\Sigma_m)\) for all \(R\), and we have \(\|f\|_{T^0_m,R} = \|g\|_{C(\Sigma_m,R)}\).

From this point of view, we have a map \(\iota : T^0_m \to C(\Sigma_{m,bdd})\) which is an isomorphism onto its image. To show that \(\iota\) 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 \(y_n(\lambda_x) = f_n(x)\) such that \(\|g - y_n\|_{C(\Sigma_{m,R})} \leq 1/2^n\). Then \(f_n\) converges to some \(f\) in \(T^0_m\), and we have \(\iota(f) = g\).

### 8.2 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 \(T^0_{m,1}\) be the completion of the trace polynomials with respect to the metric

\[
d(f,g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \min(1, \|f - g\|_{T^0_m,n}) + \min(1, \|Df - Dg\|_{(T^0_m)^m,n})\]

Observe that if \(f \in T^0_{m,1}\) and \(f_n\) is a sequence of trace polynomials converging to \(f\), then \(Df_n\) converges in \((T^1_m)^m\) and the limit is independent of the choice of approximating sequence. We denote this limit by \(Df\).

**Remark 8.3.** The function \(Df\) is the gradient of \(f\) in the following sense: For every \(x, y \in M_N(\mathbb{C})^m\) 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 \(R\). In particular, \(Df\) is uniquely determined by \(f\).
Lemma 8.4. Let $f \in \mathcal{T}_{m}^{0,1}$ be real-valued. The following are equivalent:

(1) The function $f|_{M_N(\mathbb{C})_{sa}^m}$ is convex for every $N$.

(2) The function $f$ is convex as a function on $(\mathcal{R}_0^\omega)^m$.

(3) There exists a sequence of differentiable convex functions $V_N : M_N(\mathbb{C})_{sa}^m \to \mathbb{R}$ such that $[V_*] = f$ and $[DV_*] = Df$.

Proof. The equivalence between (1) and (2) follows from similar argument to the proof of Lemma 8.1.

(1) $\implies$ (3) because we can take $V_N = f|_{M_N(\mathbb{C})_{sa}^m}$.

(3) $\implies$ (1). To prove that $f|_{M_N(\mathbb{C})_{sa}^m}$ is convex, it suffices to show that $\langle Df(x) - Df(y), x - y \rangle \geq 0$ for every $x, y \in M_N(\mathbb{C})_{sa}^m$ for every $n$. Consider $x \otimes 1_k$ and $y \otimes 1_k$ in $M_N(\mathbb{C})_{sa}^m$. Then

$$\langle Df(x) - Df(y), x - y \rangle \geq \langle Df(x \otimes 1_k) - Df(y \otimes 1_k), x \otimes 1_k - y \otimes 1_k \rangle_2,$$

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

$$\langle Df(x \otimes 1_k) - Df(y \otimes 1_k), x \otimes 1_k - y \otimes 1_k \rangle_2 \geq 0.$$ 

Because $V_N$ is convex, the second inner product is $\geq 0$ and therefore $\langle Df(x) - Df(y), x - y \rangle \geq 0$. \hfill $\square$

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 $V \in \mathcal{E}_{m,c,C}^{0,1}$, then the sequence $V|_{M_N(\mathbb{C})_{sa}^m}$ is asymptotically approximable by trace polynomials. If we let $\mu_N$ be the corresponding measure, then Theorem 4.1 (the hypothesis (4.1) being trivially satisfied by unitary invariance) implies that $\mu_N$ concentrates around a non-commutative law $\lambda_V$, which we will call the free Gibbs state for the potential $V$.

Furthermore, the free Gibbs state $\lambda_V$ is independent of the choice of representative sequence $V_N$ in the following sense. Suppose that $V_N$ is a sequence satisfying the hypotheses of Theorem 4.1 such that $[V_*] = V$ in $\mathcal{T}_{m}^{0,1}$. Then $V_N$ concentrates around some non-commutative law $\lambda$ and we claim that $\lambda = \lambda_V$. To prove this, consider the sequence $V_N$ which equals $V_N$ for odd $N$ and $V|_{M_N(\mathbb{C})_{sa}^m}$ for even $N$. Then $[V_*] = V$ in $\mathcal{T}_{m}^{0,1}$, which means that $\{DV_N\}$ is asymptotically approximable by trace polynomials. Therefore,

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

In fact, Lemma 8.4 implies that the non-commutative laws $\lambda$ which occur as limits in Theorem 4.1 are precisely the free Gibbs states for potentials $V \in \mathcal{E}_{m,c,C}^{0,1}$. In particular, Theorem 4.4 implies that $\chi = \underline{\chi} = \underline{\chi}^*$ for every such law.

Remark 8.5. We remark that we have not proved that the law $\lambda_V$ is uniquely characterized by the Schwinger-Dyson equation $\lambda(DV(x) f(X)) = \lambda \otimes \lambda[Df(X)]$, although something like this is implied by [3]. One could hope to prove this by letting the semigroup $T^V_t$ 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 $\mathcal{R}_0^\omega$-embeddable algebras). We would want to show that if $\lambda$ satisfies the Schwinger-Dyson equation, then $\lambda(T^V_t u) = \lambda(u)$, but to justify the computation, we need to show more regularity of $T^V_t u$ than we have done in this paper. In the SDE approach as well, the proof that $\lambda_V$ is characterized by Schwinger-Dyson is subtle when we do not assume more regularity for $V$ (see [10], [11]).
8.3 Examples of Convex Potentials

A natural class of examples of functions in $\mathcal{E}^{0,1}_{m,c,C}$ are those of the form

$$V(x) = \frac{1}{2}\|x\|^2 + \epsilon f(u)$$

where $\epsilon$ is a small positive parameter, $f$ is a trace polynomial, and

$$u = (u_1, \ldots, u_m), \quad u_j = \frac{x_j + 4i}{x_j - 4i}.$$ 

Computations similar to those of \[\text{(8.2)}\] show that the normalized Hessian of $\text{Jac}(Df(u(x)))$ with respect to $x$ is bounded uniformly in $N$. Therefore, $V \in \mathcal{E}^{0,1}_{m,1/2,3/2}$ for sufficiently small $\epsilon$. Similar examples are described in the introduction of \[\text{[3]}\]. 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 $\mathcal{E}^{0,1}_{m,c,C}$ does not include trace polynomials in $x$ because if $g$ is a trace polynomial of degree $\geq 3$, then $g(x) - (C/2)\|x\|^2$ cannot be concave. However, if $V$ is a small perturbation of a quadratic (as considered in \[\text{[12, 17]}\]), we can fix this problem by introducing an operator-norm cut-off as follows. Consider a potential $V$ of the form

$$V(x) = \|x\|^2 + \epsilon f(x), \quad (8.4)$$

where $f$ is a trace polynomial, and note that $Df$ and the normalized Hessian $\text{Jac}(Df)$ are bounded on the operator norm ball $\{x : \|x\| \leq R\}$ for some $R > 2$. Let $0 < R' < R$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a $C^\infty$ function such that $\phi(t) = t$ for $|t| \leq R'$ and $\phi(t) = 0$ for $|t| \geq R$. Let $\Phi(x) = (\phi(x_1), \ldots, \phi(x_m))$ and

$$\tilde{V}(x) = \|x\|^2 + \epsilon f(\Phi(x)). \quad (8.5)$$

We claim that the $D(f \circ \Phi)$ is asymptotically approximable by trace polynomials and the Hessian of $f \circ \Phi$ is bounded uniformly in $N$, and therefore that given $\delta \in (0, 1)$, we have $V \in \mathcal{E}^{0,1}_{m,1-\delta,1+\delta}$ for sufficiently small $\epsilon$. Furthermore, we claim that for $\delta$ small enough, the measure $\mu_N$ associated to $\tilde{V}$ concentrates on $\{x : \|x\| \leq R'\}$, and hence the limiting law only depends on $V$ and is independent of the choice of cut-off function $\phi$. As a consequence, we will show the following.

**Proposition 8.6.** Let $V$ be as in (8.4), let $2 < R' < R$, and let

$$d\mu_N(x) = \frac{1}{Z_N} \exp(-N^2V(x))1_{\|x\| \leq R} \, dx.$$ 

For sufficiently small $\epsilon$, we have the following. The measure $\mu_N$ exhibits exponential concentration around a non-commutative law $\lambda$. If $X \in (M, \tau)$ realizes the law $\lambda$, then $\|X\| \leq R'$ and the conjugate variable $J(X)$ is given by $DV(X)$. Moreover, we have

$$\chi(\lambda) = \chi(\lambda) = \chi^*(\lambda) = \lim_{N \to \infty} \left( \frac{1}{N^2} h(\mu_N) + \frac{m}{2} \log N \right).$$

In order to apply Theorems \[\text{[11]}\] and \[\text{[17]}\] to $\tilde{V}$, we must understand the gradient and Hessian of $f \circ \Phi$. To this end, we recall some results of Peller \[\text{[22]}\] on non-commutative derivatives of $\phi(x)$ where $\phi$ is a smooth function on the real line.
For a polynomial $\phi$, the non-commutative derivative $D\phi \in \mathbb{C}(X) \otimes \mathbb{C}(X)$ defined by Definition 3.3 can be written as

$$D\phi(s, t) = \frac{\phi(s) - \phi(t)}{s - t},$$

where we view $\mathbb{C}(X) \otimes \mathbb{C}(X)$ as a subset of functions on $\mathbb{R}^2$ with the variables $s$ and $t$. Thus, $D\phi$ is defined for every differentiable $\phi : \mathbb{R} \to \mathbb{C}$.

If $\phi$ 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 $D\phi$ in the projective tensor product $L^\infty[-R, R] \bar{\otimes} L^\infty[-R, R]$. Similar estimates for smooth functions $\phi$ can be proved by representing $\phi$ as an integral of simpler functions (e.g., Fourier analysis) whose derivatives are easier to analyze. In this case, it is convenient to write $D\phi$ 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 $\mathcal{B}(\mathbb{R})^\otimes, n$ consists of Borel functions $G$ on $\mathbb{R}^n$ which admit a representation

$$G(x_1, \ldots, x_n) = \int_{\Omega} G_1(x_1, \omega) \ldots G_n(x_n, \omega) \, d\mu(\omega)$$

(8.6)

for some measure space $(\Omega, \mu)$ such that

$$\int_{\Omega} \|G_1(\cdot, \omega)\|_{\mathcal{B}(\mathbb{R})} \ldots \|G_n(\cdot, \omega)\|_{\mathcal{B}(\mathbb{R})} \, d\mu(\omega) < +\infty$$

(8.7)

and we define $\|G\|_{\mathcal{B}(\mathbb{R})^{\otimes}, n}$ to be the infimum of (8.7) over all representations (8.6).

Given $G \in \mathcal{B}(\mathbb{R})^{\otimes, n}$, bounded self-adjoint operators $X_0, \ldots, X_n$ and bounded operators $Y_1, \ldots, Y_n$, we define

$$G(X_0, \ldots, X_n)\#(Y_1 \otimes \cdot \otimes Y_n) = \int_{\Omega} G_0(X_0, \omega)Y_1G_1(X_1, \omega) \ldots Y_nG_n(X_n, \omega) \, d\mu(\omega),$$

(8.8)

where $G_0, \ldots, G_n$ satisfy (8.6). This is well-defined by [23, Lemma 3.1]. If the $X_j$’s and $Y_j$’s are elements of a tracial von Neumann algebra $(M, \tau)$, we have by non-commutative Hölder’s inequality that if $1/p = 1/p_1 + \cdots + 1/p_n$, then

$$\|G(X_0, \ldots, X_n)\#(Y_1 \otimes \cdot \otimes Y_n)\|_p \leq \|G\|_{\mathcal{B}(\mathbb{R})^{\otimes, (n+1)}} \|Y_1\|_{p_1} \ldots \|Y_n\|_{p_n},$$

(8.9)

where $\|a\|_p = \tau((a^*a)^{p/2})^{1/p}$. Moreover, we have the following bounds on the non-commutative derivatives of $\phi$.

**Proposition 8.7.** Let $\phi \in C^\infty_c(\mathbb{R})$. Then for some constant $C_n$,

$$\|D^n\phi\|_{\mathcal{B}(\mathbb{R})^{\otimes, (n+1)}} \leq C_n \int_{\mathbb{R}} |\hat{\phi}(\xi)| \xi^n \, d\xi.$$  

(8.10)

Proof. As in [23, §2], choose $w \in C_c^\infty$ such that $0 \leq w \leq \chi_{[-1/2, 2]}$ and $\sum_{k \in \mathbb{Z}} w(2^{-k}) = 1$ for $\xi > 0$. Let $W_k$ and $W_k^\#$ be given by $\widehat{W_k}(\xi) = w(2^{-k}\xi)$ and $\widehat{W_k^\#}(\xi) = w(-2^{-k}\xi)$ where $\hat{}$ denotes the Fourier transform. It is shown in [23, Theorem 5.5] that

$$\|D^n\phi\|_{\mathcal{B}(\mathbb{R})^{\otimes, (n+1)}} \leq C_n \sum_{k \in \mathbb{Z}} 2^{nk} \left( \|W_k \ast \phi\|_{L^\infty(\mathbb{R})} + \|W_k^\# \ast \phi\|_{L^\infty(\mathbb{R})} \right).$$

This can be estimated by the right hand side of (8.10) by a standard Fourier analysis computation. 

$\blacksquare$
Lemma 8.8. Let $\tilde{V}$ be given by (8.5) and let $\delta \in (0, 1)$. Then for sufficiently small $\epsilon$, we have $\tilde{V} \in C^{1,\frac{1}{2}}_{m,1-\frac{1}{2}+\delta}$.

Proof. Because $\tilde{V}(x) = \frac{1}{2}\|x\|^2 + \epsilon f \circ \Phi$, it suffices to show that $f \circ \Phi \in C^{1,\frac{1}{2}}_{m,-a,b}$ for some $a > 0$. First, let us show that $f \circ \Phi \in T^{1,0}_{m}$ by constructing an approximating sequence of trace polynomials. Fix $r > 0$. By standard approximation techniques, there exist Schwarz functions $\phi_k : \mathbb{R} \to \mathbb{R}$ such that $\phi_k \to \phi$ in the Schwarz space. By Proposition 8.7, we have $D^n \phi_k \to D^n \phi$ in $\mathcal{B}(\mathbb{R}) \hat{\otimes} \mathcal{S}^{(n+1)}$ for every $n$.

Let $\Phi_k(x_1, \ldots, x_m) = (\phi_k(x_1), \ldots, \phi_k(x_m))$. Then $f \circ \Phi_k(x) = f \circ \Phi(x)$ uniformly on $\{\|x\| < r\}$. Next, to bound on the Hessian of $f \circ \Phi$, let us show that $D_j [f \circ \Phi]$. Note that for $y \in M_N(\mathbb{C})_{sa}$

$$\langle D_j [f \circ \Phi_k](x), y \rangle_2 = \langle D_j f(\Phi_k(x)), D\Phi_k(x_j)\# y \rangle_2 = \langle D\Phi_k(x_j)\# D_j f(\Phi_k(x)), y \rangle_2,$$

where the last equality relies on the fact that $\phi_k(x_j)$ only depends on one variable. Therefore, $D_j [f \circ \Phi_k](x) = D\Phi_k(x_j)\# D_j f(\Phi_k(x))$ for every $j = 1, \ldots, m$. This function is given by a trace polynomial on $\{\|x\| \leq r\}$. Moreover, for $\|x\| \leq r$, we have

$$D\phi_k(x_j)\# D_j f(\Phi_k(x)) = D\phi_k(x_j)\# D_j f(\Phi(x)) + D\Phi_k(x_j)\# [D_j f(\Phi_k(x)) - D_j f(\Phi(x))].$$

The first term converges to $D(\phi(x_j))\# D_j f(\Phi(x))$ in $\|\cdot\|_2$ uniformly on $\{\|x\| \leq r\}$ using (8.3). Similarly, because the images of $\Phi_k$ and $\Phi$ are contained in an operator norm ball and $D_j f$ is $L$-Lipschitz in $\|\cdot\|_2$ on this ball for some $L > 0$, 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 $D\phi_k(x_j)$ is uniformly bounded in $\mathcal{B}(\mathbb{R}) \hat{\otimes} \mathcal{B}(\mathbb{R})$. 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 $\{\|x\| \leq r\}$. This means that $f \circ \Phi \in T^{1,0}_{m}$.

Next, to bound on the Hessian of $f \circ \Phi$, it suffices to show that $D_j [f \circ \Phi] = D\phi(x_j)\# D_j f(\Phi(x))$ is Lipschitz in $\|\cdot\|_2$. Because $D^2 \phi$ is bounded in $\mathcal{B}(\mathbb{R}) \hat{\otimes} \mathcal{B}(\mathbb{R}) \hat{\otimes} \mathcal{B}(\mathbb{R})$, we see that

$$\|D(\phi(x_j))\# y - D(\phi(x_j'))\# y\|_2 \leq L\|x_j - x_j'\|_2\|y\|_2$$

for some constant $L$. Together with the fact that $D_j f(\Phi(x))$ is Lipschitz in $\|\cdot\|_2$, this implies that $D_j (f \circ \Phi)$ is Lipschitz in $\|\cdot\|_2$ as desired. □

Proof of Proposition 8.7. Let $\tilde{\mu}_N$ be the measure on $M_N(\mathbb{C})_{sa}$ given by the potential $\tilde{V}$. Let $\delta$ be a number in $\{0,1\}$ to be chosen later. By Lemma 8.8, we have that $\tilde{V} \in C^{1,\frac{1}{2}}_{m,1-\frac{1}{2}+\delta}$ for sufficiently small $\epsilon$. By Theorem 4.13, the laws $\tilde{\mu}_N$ concentrate around a non-commutative law $\lambda$. Furthermore, in Theorem 4.11 (1), we can take $M = 0$ and $e = 1 - \delta$ and $C = 1 + \delta$, so that

$$\limsup_{N \to \infty} R_N \leq \frac{2}{(1 - \delta)^{1/2}} + \frac{\|D\tilde{V}(0)\|_2}{1 - \delta} + \frac{\delta}{(1 - \delta)^{3/2}}.$$

Note that $D\tilde{V}(0) = DV(0) = \epsilon Df(0)$ is a scalar. Because $R' > 2$, we may choose $\delta$ sufficiently small that

$$\frac{2}{(1 - \delta)^{1/2}} + \frac{\delta}{(1 - \delta)^{3/2}} < R'.$$

Then by choosing $\epsilon$ (and hence $\|D\tilde{V}(0)\|_2$) sufficiently small, we can arrange that $\limsup_{N \to \infty} R_N < R'$. This implies that the measures $\tilde{\mu}_N$ are concentrated on the ball $\{\|x\| \leq R'\}$. On this ball, we have $\tilde{V}(x) = V(x)$, and therefore $\tilde{\mu}_N$ is the (normalized) restriction of $\tilde{\mu}_N$ to $\{\|x\| \leq R\}$. It follows that $\mu_N$ concentrates around the law $\lambda$ as well.
If \( X \in (M, \tau) \) realize the law \( \lambda \), then \( \|X\| \leq R' \) and the conjugate variables for \( \lambda \) are given by \( D\tilde{V}(X) = D\tilde{V}(X) \). Moreover, by Theorem 7.1 applied to \( \tilde{\mu}_N \), we have

\[
\chi(\lambda) = \chi(\lambda) = \chi^*(\lambda) = \lim_{N \to \infty} \left( \frac{1}{N^2} h(\tilde{\mu}_N) + \frac{n}{2} \log N \right).
\]

In the last equality, we can replace \( \tilde{\mu}_N \) by \( \mu_N \) as in the proof of Proposition 5.3 because \( \tilde{\mu}_N \) is concentrated on \( \{\|x\| \leq R'\} \).

**Remark 8.9.** The approach given here probably does not give the optimal range of \( \epsilon \) for Proposition 8.6. To get the best result, one would want a more direct way to extend the potential \( V : \{\|x\| \leq R\} \to \mathbb{R} \) to a potential \( \tilde{V} \) 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 \( \|f\|_{T_{m, R}} + \|Df\|_{T_{m, R}} \), and hence \( V \) defines a function \( \|x\| \leq R \to \mathbb{R} \) for \( x \in M_N(\mathbb{C})^m_{sa} \). If \( V(x) - (c/2)\|x\|_2^2 \) is convex and \( V(x) - (C/2)\|x\|_2^2 \) is concave on \( \{\|x\| \leq R\} \), then does \( V \) extend to a potential \( \tilde{V} \in \mathcal{E}_{m,c,C}^{0,1} \)? What if we allow \( \tilde{V} \) 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.

**References**

1. G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2009.
2. C. Antharaman and S. Popa, *An Introduction to III Factors*, 2017. preprint available at http://www.math.ucla.edu/~popa/Books/IIun-v10.pdf.
3. A. R. Barron, *Entropy and the central limit theorem*, Ann. Prob., 14 (1996), pp. 336–342.
4. P. Biane, *Segal-Bargmann transform, functional calculus on matrix spaces and the theory of semi-circular and circular systems*, Journal of Functional Analysis, 144 (1997), pp. 232 – 286.
5. P. Biane, M. Capitaine, and A. Guionnet, *Large deviation bounds for matrix Brownian motion*, Invent. Math., 152 (2003), p. 433459.
6. V. Capraro, *A survey on Connes’ embedding conjecture*, arXiv e-prints, (2010).
7. G. Cébron, *Free convolution operators and free hall transform*, Journal of Functional Analysis, 265 (2013), pp. 2645 – 2708.
8. 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., 27 (1992), pp. 1–67.
9. Y. Dabrowski, *A Laplace principle for Hermitian Brownian motion and free entropy I: the convex functional case*, arXiv e-prints, (2017).
10. Y. Dabrowski, *A non-commutative path space approach to stationary free stochastic differential equations*, arXiv e-prints, (2010).
[11] Y. Dabrowski, A. Guionnet, and D. Shlyakhtenko, *Free transport for convex potentials*, arXiv e-prints, (2016).

[12] B. K. Driver, B. C. Hall, and T. Kemp, *The large-n limit of the segalbargmann transform on un*, Journal of Functional Analysis, 265 (2013), pp. 2585 – 2644.

[13] T. C. Duvillard and A. Guionnet, *Large deviations upper bounds for the laws of matrix-valued processes and non-communicative entropies*, Ann. Probab., 29 (2001), pp. 1205–1261.

[14] L. C. Evans, *Partial Differential Equations*, vol. 19 of Graduate Studies in Mathematics, American Mathematical Society, Providence, Rhode Island, 2 ed., 2010.

[15] A. Guionnet and E. Maurel-Segala, *Combinatorial aspects of random matrix models*, Alea, 1 (2006), pp. 241–279.

[16] A. Guionnet and D. Shlyakhtenko, *Free diffusions and matrix models with strictly convex interaction*, Geometric and Functional Analysis, 18 (2009), pp. 1875–1916.

[17] ——, *Free monotone transport*, Inventiones mathematicae, 197 (2014), pp. 613–661.

[18] G. Hargé, *A convex/log-concave correlation inequality for gaussian measure and an application to abstract wiener spaces*, Probab. Theory Relat. Fields, 130 (2004), pp. 415–440.

[19] T. Kato, *Trotter’s product formula for an arbitrary pair of self-adjoint contraction semigroups*, in Topics in Functional Analysis, Essays Dedicated to M.G. Krein, I. Gohberg and M. Kac, eds., vol. 3 of Advances in Mathematics, Supplementary Studies, Academic Press, 1978.

[20] J. M. Lasry and P. L. Lions, *A remark on regularization in hilbert spaces*, Israel Journal of Mathematics, 55 (1986), pp. 257–266.

[21] M. Ledoux, *A heat semigroup approach to concentration on the sphere and on a compact Riemannian manifold*, Geometric & Functional Analysis, 2 (1992), pp. 221–224.

[22] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*, vol. 335 of London Mathematical Society Lecture Note Series, Cambridge University Press, 2006.

[23] V. V. Peller, *Multiple operator integrals and higher operator derivatives*, Journal of Functional Analysis, 233 (2006), pp. 515–544.

[24] B. Simon, *Functional Integration and Quantum Physics*, Academic Press, 1979.

[25] H. F. Trotter, *On the product of semi-groups of operators*, Proc. Amer. Soc., 10 (1959), pp. 545–551.

[26] D. Voiculescu, *Addition of certain non-commuting random variables*, Journal of Functional Analysis, 66 (1986), pp. 323 – 346.

[27] ——, *Limit laws for random matrices and free products*, Inventiones mathematicae, 104 (1991), pp. 201–220.

[28] ——, *The analogues of entropy and of Fisher’s information in free probability, I*, Comm. Math. Phys., 155 (1993), pp. 71–92.
[29] ______, *The analogues of entropy and of Fisher’s information in free probability, II*, Invent. Math., 118 (1994), pp. 411–440.

[30] ______, *The analogues of entropy and of Fisher’s information in free probability V*, Invent. Math., 132 (1998), pp. 189–227.

[31] ______, *Free entropy*, Bulletin of the London Mathematical Society, 34 (2002), p. 257278.

[32] D. VOICULESCU, K. J. DYKEMA, AND A. NICA, *Free Random Variables*, vol. 1 of CRM Monograph Series, American Mathematical Society, 1992.