1.1. Motivation.

Free probability initiated a fruitful exchange between random matrix theory and operator algebras. In many situations, tuples of $N \times N$ random matrices $(X_1^{(N)}, \ldots, X_m^{(N)})$ can be described in the large $N$ limit by non-commutative random variables $X_1, \ldots, X_m$ which are operators in a tracial $W^*$-algebra. Conversely, many properties of non-commutative random variables (and the $W^*$-algebras that they generate) are easier to understand when they can be simulated by finite-dimensional random matrix models. For instance, Voiculescu used free entropy, defined in terms of matricial microstates, to prove the absence of Cartan subalgebras and asymptotically central sequences for free group factors, which arises as the large $N$ limit of transport maps for probability measures on $M_N(C)^{sa}$. This transport can be made “upper-triangular” in the sense that $W^*(X_1, \ldots, X_m)$ is mapped to $W^*(S_1, \ldots, S_m)$ generated by a free semicircular family (i.e. a free group factor), which arises as the large $N$ limit of transport maps for probability measures on $M_N(C)^{sa}$. We show that for a function $f$ that is well approximated by trace polynomials, the classical conditional expectation of $f(X^{(N)})$ given $X_1^{(N)}, \ldots, X_k^{(N)}$ for $k < m$ converges to the $W^*$-algebraic conditional expectation of $f(X)$ given $X_1, \ldots, X_k$. We also construct an isomorphism from $W^*(X_1, \ldots, X_m)$ to the algebra $W^*(S_1, \ldots, S_m)$ generated by a free semicircular family (i.e. a free group factor), which arises as the large $N$ limit of transport maps for probability measures on $M_N(C)^{sa}$. This transport can be made “upper-triangular” in the sense that $W^*(X_1, \ldots, X_m)$ is mapped to $W^*(S_1, \ldots, S_m)$ for each $k = 1, \ldots, m$. At the same time, this transport map witnesses the Talagrand entropy-cost inequality for $\mu$ relative to the law of a semicircular family.

1. Introduction

1.1. Motivation. Free probability initiated a fruitful exchange between random matrix theory and operator algebras. In many situations, tuples of $N \times N$ random matrices $(X_1^{(N)}, \ldots, X_m^{(N)})$ can be described in the large $N$ limit by non-commutative random variables $X_1, \ldots, X_m$ which are operators in a tracial $W^*$-algebra. Conversely, many properties of non-commutative random variables (and the $W^*$-algebras that they generate) are easier to understand when they can be simulated by finite-dimensional random matrix models. For instance, Voiculescu used free entropy, defined in terms of matricial microstates, to prove the absence of Cartan subalgebras and asymptotically central sequences for free group $W^*$-algebras $L(F_n)$ [29]. Further applications of random matrices to the properties of $C^*$- and $W^*$-algebras can be found in [13, §4].

Free Gibbs laws are a prototypical example of the connection between random matrices and $W^*$-algebras. Free Gibbs laws describe the large $N$ behavior of self-adjoint tuples of random matrices $X^{(N)} = (X_1^{(N)}, \ldots, X_m^{(N)})$ given by a probability density $\mu^{(N)}$ of the form

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

where $x \in M_N(C)^{sa}$ is a self-adjoint tuple, $dx$ denotes Lebesgue measure, $Z^{(N)}$ is normalizing constant to $\mu^{(N)}$ a probability measure, and $V^{(N)} : M_N(C)^{sa} \to \mathbb{R}$ is a function known as a potential. Here $V_N(x)$ could be given by $V_N(x) = \tau_N(p(x_1, \ldots, x_m))$, where $\tau_N = (1/N)\text{Tr}$ and $p(x_1, \ldots, x_m)$ is a polynomial of $x_1, \ldots, x_m$.

1991 Mathematics Subject Classification. Primary: 46L53, Secondary: 35K10, 37A35, 46L52, 46L54, 60B20.

Key words and phrases. free entropy, free Fisher information, free Gibbs state, trace polynomials, invariant ensembles.
\( p \) is a non-commutative polynomial; for instance, taking

\[
V^{(N)}(x) = \sum_{j=1}^{m} \tau_N(x_j^2)
\]

produces the Gaussian unitary ensemble (GUE). Under certain assumptions on \( V \) (e.g. convexity and good asymptotic behavior as \( N \rightarrow \infty \)), there will be a non-commutative random variables \( X_1, \ldots, X_m \) in a tracial \( W^\ast \)-algebra \( (\mathcal{M}, \tau) \) such that

\[
\tau_N(p(X_1^{(N)}, \ldots, X_m^{(N)})) \rightarrow \tau(p(X_1, \ldots, X_m))
\]

in probability for every non-commutative polynomial \( p \); see [12, Thm. 3.3 and 3.4], [10, Prop. 50 and Thm. 51], [17, Theorem 4.1]. The random models satisfy the relation, derived from integration by parts, that

\[
\tau(DV(X)p(X)) = \tau \otimes \tau(D_{x_j} p(X));
\]

see [12, §2.2 - 2.3]. The non-commutative law of \( X \) satisfying such an equation is known as a \textit{free Gibbs law} for the potential \( V \).

Many of the classical quantities associated to \( X^{(N)} \) will converge in the large \( N \) limit to their free counterparts (given sufficient assumptions on \( V_N \)), besides obviously the convergence of the non-commutative moments \( \tau_N(p(X^{(N)})) \). For instance, the normalized classical entropy will converge to the microstates free entropy (see [28, §2], [13, Thm. 5.1], [17, §5.2]), and the normalized classical Fisher information will converge to the free Fisher information (see [17, §5.3]). The monotone transport maps of Guionnet and Shlyakhtenko are well-approximated by classical transport maps for the random matrix models [14, Theorem 4.7]. The solutions of classical SDE associated to the random matrix models approximate the solutions of free SDE; see for instance [13, §2], [3, §4].

This paper will further develop the connection between classical and free probability for convex free Gibbs states, by studying conditional expectation (§5), conditional entropy and Fisher information (§6), and conditional transport (§7). After reviewing notation, we will motivate and explain our main results.

1.2. Notation and Background. We will continue to use the same notation and background as in §17. The one major change is that we will write superscript \((N)\) rather than subscript \( N \) for measures and functions defined on \( N \times N \) matrices. We will also assume greater background in von Neumann algebras.

We assume familiarity with the setup of tracial \( W^\ast \)-algebras (or tracial von Neumann algebras); see for instance [2]. In particular, a tracial \( W^\ast \)-algebra is a finite \( W^\ast \)-algebra \( \mathcal{M} \) with a specified trace \( \tau : \mathcal{M} \rightarrow \mathbb{C} \). If \( \mathcal{N} \subseteq \mathcal{M} \) is a \( W^\ast \)-algebra, then there is a unique normal trace-preserving conditional expectation \( E_{\mathcal{N}} : \mathcal{M} \rightarrow \mathcal{N} \). If \( x = (x_1, \ldots, x_m) \) is a tuple of operators in \( \mathcal{M} \), then we denote by \( W^\ast(x) \) the \( W^\ast \)-subalgebra which they generate.

There is an inner product on \( \mathcal{M} \) defined by \( \langle x, y \rangle_2 = \tau(x^* y) \), and the completion of \( \mathcal{M} \) in this inner product is a Hilbert space known as \( L^2(\mathcal{M}, \tau) \). We denote the self-adjoint elements of \( \mathcal{M} \) by \( \mathcal{M}_{sa} \) and recall that if \( x \) and \( y \) are self-adjoint, then \( \langle x, y \rangle_2 \) is real. If \( x = (x_1, \ldots, x_m) \) and \( y = (y_1, \ldots, y_m) \) are tuples, we denote \( \langle x, y \rangle_2 = \sum_{j=1}^{m} \langle x_j, y_j \rangle_2 \). We define \( \|x\|_\infty = \max_j \|x_j\|_\infty \), that is the maximum of the operator norms of \( x_j \).

We denote by \( \text{NCP}_m = \mathbb{C}(X_1, \ldots, X_m) \) the *-algebra of non-commutative polynomials in \( m \) self-adjoint variables. A \textit{non-commutative law} is a linear map \( \lambda : \mathbb{C}(X_1, \ldots, X_m) \rightarrow \mathbb{C} \) satisfying
(A) \(\lambda(1) = 1\).
(B) \(\lambda(p^*)p \geq 0\) for all \(p \in \text{NCP}_m\).
(C) \(\lambda(pq) = \lambda(qp)\) for all \(p, q \in \text{NCP}_m\).
(D) \(|\lambda(X_1, \ldots, X_n)| \leq R^k\) for some constant \(R\).

The set of non-commutative laws that satisfy (D) for a fixed value of \(R\) is denoted \(\Sigma_{m,R}\), and it is equipped with the topology of pointwise convergence on \(\text{NCP}_m\). Likewise, the space of all laws, equipped with the topology of pointwise convergence will be denoted by \(\Sigma_m\).

If \(x = (x_1, \ldots, x_m)\) is a tuple of self-adjoint elements of \((\mathcal{M}, \tau)\), then we may define a non-commutative law \(\lambda_x\) by \(\lambda_x(p) = \tau(p(x))\). Conversely, every non-commutative law can be realized in this way through the GNS construction. In particular, a Gibbs law can be realized by a tuple \((x_1, \ldots, x_m)\) of self-adjoint operators, and thus the free Gibbs law has a corresponding \(W^*\)-algebra \(W^*(x)\), that is unique up to isomorphism.

We always consider \(M_N(\mathbb{C})\) as a tracial \(W^*\)-algebra with the normalized trace \(\tau_N = (1/N)\text{Tr}\), and in particular, we use the notation \(\|x\|_2\), \(\|x\|_\infty\), and \(\lambda_x\) as defined above when \(x\) is an \(m\)-tuple of matrices. The notation \(\|\cdot\|_2\) and \(\|\cdot\|_\infty\) will never be used for the \(L^2\) or \(L^\infty\) norms of functions on matrices, but if we write an \(L^p\) norm it will be expressed \(\|\cdot\|_{L^p}\).

For a smooth function \(u : M_N(\mathbb{C})_{sa} \to \mathbb{R}\), we denote by \(Du\) and \(Hu\) the gradient and Hessian with respect to the normalized inner product \(\langle \cdot, \cdot \rangle_2\). In other words, \(Du(x_0)\) is the vector in \(M_N(\mathbb{C})_{sa}\) and \(Hu(x_0)\) is the \(\mathbb{R}\)-linear transformation of \(M_N(\mathbb{C})_{sa}\) satisfying
\[
u(x) = u(x_0) + \langle Du(x_0), x - x_0 \rangle_2 + \frac{1}{2} \langle Hu(x_0)(x - x_0), x - x_0 \rangle_2 + o(\|x-x_0\|_2^2).
\]

Note that \(M_N(\mathbb{C})_{sa}\) can also be equipped with the real inner product \(\langle x, y \rangle_{\text{Tr}} = \sum_{j=1}^m \text{Tr}(x_jy_j)\). Being a real inner-product space, \(M_N(\mathbb{C})_{sa}\) may be identified with \(\mathbb{R}^{mN^2}\) by choosing an orthonormal basis in \(\langle \cdot, \cdot \rangle_{\text{Tr}}\). Lebesgue measure on \(M_N(\mathbb{C})_{sa}\) should be understood with respect to this identification. Moreover, the gradient \(\nabla\), Jacobian matrix \(J\), divergence \(\text{Div}\), and Laplacian \(\Delta\) for functions on \(M_N(\mathbb{C})_{sa}\) should also be understood with respect to this identification. Beware that this is not equivalent to using entrywise coordinates for \(M_N(\mathbb{C})_{sa}\) since the off-diagonal entries are complex and conjugate-symmetric, while the diagonal entries are real. For further discussion see [17] §2.1.

1.3. Main Results on Conditional Expectation. Consider a tuple
\[
(X^{(N)}, Y^{(N)}) = (X_1^{(N)}, \ldots, X_m^{(N)}, Y_1^{(N)}, \ldots, Y_n^{(N)})
\]
of random self-adjoint matrices given by a probability density \((1/Z^{(N)})e^{-N^2V^{(N)}(x,y)}\,dx\,dy\). We assume that \(V^{(N)}\) is uniformly convex and semi-concave and that the normalized gradient \(DV^{(N)}\) is asymptotically approximable by trace polynomials (a certain notion of good asymptotic behavior as \(N \to \infty\), explained below). The precise hypotheses are listed in Assumption 5.1. We showed in [17, Theorem 4.1] that in this case, there exists an \((m+n)\)-tuple \((X,Y)\) of non-commutative random variables such that \(\tau_N(p(X,Y)) \to \tau(p(X,Y))\) in probability.

Our first main result (Theorem 5.9) says roughly that the classical conditional expectation given \(Y^{(N)}\) well approximates the \(W^*\)-algebraic conditional expectation \(E_{W^*(Y)} : W^*(X,Y) \to W^*(Y)\). This is motivated in general by the importance of conditional expectation in free probability (e.g. its relationship to free independence with amalgamation and to free conjugate variables). The relationship between classical and free conditional expectation also has implications for the study of relative matricial microstate spaces, such as the "external averaging property" introduced in the upcoming work with Hayes, Nelson, and Sinclair.

Applications of conditional expectation within this paper include our results on free Fisher information and entropy (see Theorem 6.6 and Remark 6.8), as well as our proof that Assumption 5.1 is preserved under marginals (see Proposition 8.2).
The statement and proof of Theorem 5.9 rely on a notion of asymptotic approximation for functions on \( M_N(\mathbb{C})_{sa}^m \) as explained in [3]. We define a class of non-commutative functions \( \text{TfP}_m \) as a certain Fréchet space completion of trace polynomials, such that if \( f \in \text{TfP}_m \) and \( x_1, \ldots, x_m \) are self-adjoint elements in an \( \mathcal{R}^\omega \)-embeddable tracial \( W^* \)-algebra \((\mathcal{M}, \tau)\), then \( f(x_1, \ldots, x_m) \) is a well-defined element of \( L^2(\mathcal{M}) \). In particular, every \( f \in \text{TfP}_m \) can be evaluated on a tuple of self-adjoint matrices. Now if \( f^{(N)} : M_N(\mathbb{C})_{sa}^m \rightarrow M_N(\mathbb{C}) \), we say that \( f^{(N)} \sim f \) if for every \( R > 0 \),
\[
\lim_{N \to \infty} \sup_{x \in M_N(\mathbb{C})_{sa}^m} \frac{\|f^{(N)}(x) - f(x)\|_2}{\|x\|_\infty \leq R} = 0,
\]
Moreover, if there exists such an \( f \), then we say that \( f^{(N)} \) is asymptotically approximable by trace polynomials.

Consider the random matrices \( (X^{(N)}, Y^{(N)}) \) and non-commutative random variables \((X, Y)\) as above, and suppose that \( f^{(N)} : M_N(\mathbb{C})_{sa}^{m+n} \rightarrow M_N(\mathbb{C}) \) is uniformly Lipschitz in \( \|\cdot\|_2 \) and that \( f^{(N)} \sim f \in \text{TfP}_{m+n} \). Then we show that \( E[f^{(N)}(X^{(N)}, Y^{(N)})|Y^{(N)}] \) is given by a function \( g^{(N)}(Y^{(N)}) \) such that \( g^{(N)} \sim g \in \text{TfP}_{n} \), and moreover \( E_{W^*}(f(X, Y)) = g(Y) \).

A curious feature of this result is that the function \( g \) is defined for all self-adjoint \( n \)-tuples of non-commutative random variables, not only for the specific \( n \)-tuple \( Y \) that we are concerned with. Similarly, the claim that \( g^{(N)} \sim g \) describes the asymptotic behavior of \( g^{(N)}(y) \) for all \( y \in M_N(\mathbb{C})_{sa}^n \), even though the distribution of the random matrix \( Y^{(N)} \) is highly concentrated as \( N \to \infty \) on much smaller sets, namely the “matricial microstate spaces” consisting of tuples \( y \in M_N(\mathbb{C})_{sa}^n \) with non-commutative moments close to those of \( Y \). Thus, the statement we prove about the functions \( g^{(N)} \) is much stronger than an asymptotic result about \( L^2 \) approximation.

1.4. Main Results on Entropy. Voiculescu defined two types of free entropy. The first \( \chi(X) \) is based on measuring the size of matricial microstate spaces, which is closely related to the classical entropy of the random matrix models (see [17 §5.2]). The second \( \chi^*(X) \) is defined in terms of free Fisher information, which is based on classical Fisher information. Either one should heuristically be the large \( N \) limit of the classical entropy of random matrix models, but there were many technical obstacles to proving this. The inequality \( \chi \leq \chi^* \) is known in general thanks to [3]. However, even for non-commutative laws as well-behaved and explicit as free Gibbs states given by convex potentials, the equality of \( \chi \) and \( \chi^* \) when \( m > 1 \) was not proved until Dabrowski’s paper [8], and the problem is still open for non-convex Gibbs states.

Our previous work [17] gave a proof of this equality in the convex case based on the asymptotic analysis of functions and PDE related to the random matrix models. Here we will use similar techniques for the conditional setting. We will show (Theorem 6.9) that for \( (X^{(N)}, Y^{(N)}) \) given by a convex potential as above, the classical conditional entropy \( N^{-2}h(X^{(N)}|Y^{(N)}) + (m/2) \log N \) converges to the conditional free entropy \( \chi^*(X : W^*(Y)) \). Actually, the proof here is shorter than those of [3] and [17] (see Remark 6.8).

We focus here only on the non-microstates entropy (defined using Fisher information). It is not yet resolved in the literature what the correct definition of conditional microstates free entropy should be. In light of [17 §5.2], the conditional classical entropy for the random matrix models seems to be a reasonable substitute for microstates entropy, and in the convex setting we expect this to agree with any plausible definition of conditional microstates entropy due to the exponential concentration of measure.

1.5. Main Results on Transport. A transport map between two probability measures \( \mu \) and \( \nu \) is a function \( f \) such that \( f_\mu = \nu \). In probabilistic language, if \( X \sim \mu \) and \( Y \sim \nu \) are random variables, then \( f_\mu = \nu \) means that \( f(X) \sim Y \) in distribution. The theory of transport (and in
particular optimal transport) has numerous and significant applications in the classical setting. For instance, if we have a function $f$ such that $f(X) \sim Y$ and we can numerically simulate the random variable $X$, then we can also simulate $Y$.

In the non-commutative world, transport is even more significant. As remarked in §1.1, there is no known analogue of a probability density in free probability. However, the existence of transport maps that would express our given random variables as functions of a free semicircular family (for instance) would serve a similar purpose to a density, namely to provide a fairly explicit and analytically tractable model for many non-commutative laws.

Moreover, in contrast to the classical setting, the very existence of transport maps is a nontrivial condition. In the classical setting, any two diffuse (non-atomic) Borel probability spaces are isomorphic. On the other hand, there are many non-isomorphic diffuse tracial $W^*$-algebras, even if we demand that they are factors. Being able to express a non-commutative random variable $X$ as a function of another non-commutative random variable $Y$ implies that $W^*(X)$ embeds into $W^*(Y)$, and having a transport map in the other direction as well implies that $W^*(Y) \cong W^*(X)$.

The papers [12] and [10] showed the existence of monotone transport maps between certain free Gibbs laws given by convex potentials and the law of a free semicircular family, and thus concluded that each of the corresponding $W^*$-algebras was isomorphic to a free group factor $L(F_n)$. In particular, this result applies to the $q$-Gaussian variables for sufficiently small $q$. These transport techniques have been extended to type III von Neumann algebras [21], to planar algebras [22], and to interpolated free group factors [15]. We will focus on “conditional transport” in the tracial setting.

Our first main result about transport is contained in Theorems 7.11 and 7.13. Let $(X^{(N)}, Y^{(N)})$ be an $(m+n)$-tuple of random matrices arising from a sequence of convex potentials satisfying Assumption 5.1. Let $(X, Y)$ be an $(m+n)$-tuple of non-commutative self-adjoint variables realizing the limiting free Gibbs law. Then we construct functions $F^{(N)} : M_N(C)^{m+n} \to M_N(C)^m$ such that $(F^{(N)}(X^{(N)}, Y^{(N)})) \sim (S^{(N)}, Y^{(N)})$ in distribution, where $S^{(N)}$ is a GUE $m$-tuple independent of $Y^{(N)}$. We think of this as a conditional transport, which transports the law of $X^{(N)}$ to the law of $S^{(N)}$ conditioned on $Y^{(N)}$.

Moreover, we show that the transport maps satisfy $F^{(N)} \sim F \in (M_N(F_m))^m$. In the large $N$ limit, we obtain $(F^{(N)}(X, Y), Y) \sim (S, Y)$ in non-commutative law, where $S$ is a free semicircular $m$-tuple freely independent of $Y$. In particular, this means that $W^*(S, Y) = W^*(S) \ast W^*(Y)$ (where $\ast$ denotes free product). In other words, $W^*(Y)$ is freely complemented in $W^*(X, Y)$.

By iterating this result, we can show that if $X = (X_1, \ldots, X_m)$ is a tuple of non-commutative random variables given by a convex free Gibbs state as above, then there is an isomorphism $W^*(X) \to W^*(S)$ such that $W^*(X_k, \ldots, X_m)$ is mapped onto $W^*(S_k, \ldots, S_m)$ for each $k = 1, \ldots, m$. In other words, there is an “upper-triangular transport.” See Theorem 8.11. This implies in particular that $W^*(X_m)$ is a maximal abelian subalgebra and in fact maximal amenable, and the same holds for each $W^*(X_j)$ by symmetry. For context on maximal amenable subalgebras, see for instance [28] [13] [14].

Denote by $F$ the transport map from the law of $X$ to the law of $S$ in our construction, so that $F(X) \sim S$. We can also arrange that $F$ witnesses the Talagrand entropy-cost inequality relative to the semicircular law, that is,

$$
\|F(X) - X\|_2^2 \leq \|X\|_2^2 + \frac{m}{2} \log 2\pi - 2\chi^*(X),
$$

where the left hand side is twice the entropy relative to semicircular (see §8.3). This is not surprising because our construction of the transport maps is a direct application of the same method that Otto and Villani used to prove the entropy-cost inequality under the assumption
of the log-Sobolev inequality [24, §4]. Our main contribution is to study the large \(N\) limit of the transport maps using asymptotic approximation by trace polynomials. We also show that \(F\) is \(\|\cdot\|_2\)-Lipschitz and estimate \(\|F(X) - X\|_\infty\) in terms of the constants \(c\) and \(C\) specifying the uniform convexity and semi-concavity of \(V^{(\mathbb{N})}\).

1.6. Outline. The paper is organized as follows.

§2 gives background on convex and semi-concave functions and on log-concave random matrix models, primarily as a reference.

§3 sets up the algebra of trace polynomials, and the spaces \(\text{TrP}^0_m\) and \(\text{TrP}^1_m\) of functions that can be approximated by trace polynomials. These spaces provide a framework for functional calculus in multiple self-adjoint variables \(X_1, \ldots, X_m\) that can realize every element of \(L^2(W^*(X_1, \ldots, X_m))\). They are a convenient tool to describe the large \(N\) behavior of functions of several matrices, and thus will be used in the statements of our main theorems.

§4 describes solving ODE’s and the heat equation over \(\text{TrP}^1_m\). These are the technical lemmas used in the rest of the paper to show that the solutions of certain PDE’s have well-defined large \(N\) limits.

§5 explains the setup of our random matrix models given by convex potentials, and then proves our main result on conditional expectation (Theorem 5.9).

§6 shows that the conditional entropy for random matrix models converges to the conditional non-microstates entropy (Theorem 6.6).

§7 proves the existence of transport maps from a free Gibbs law to the law of a free semicircular \(m\)-tuple which arise as the large \(N\) limit of transport maps for the random matrix models (Theorem 7.11 and 7.13).

§8 discusses applications of our results. We show that our standard set of assumptions for log-concave random matrix models is preserved under marginals, independent joins, linear change of variables, and convolution (§8.1). We show that the transport maps constructed above witness (the conditional version of) Talagrand’s entropy-cost inequality relative to Gaussian measure (Theorem 8.10). Then by iterating our conditional transport results, we show the existence of triangular transport (Theorem 8.11).

Acknowledgements. I thank Dima Shlyakhtenko, Ben Hayes, Brent Nelson, and Yoann Dabrowski for various useful conversations and suggesting many references. The results of this paper were motivated in part by discussions with Ben Hayes regarding free entropy and maximal amenable subalgebras. Dima Shlyakhtenko suggested the name “triangular transport.” Yoann Dabrowski pointed out several typos. I acknowledge support of the NSF grant DMS-1762360.

2. Multi-matrix Models from Convex Potentials

This section is a review and reference for basic results we will use throughout the paper. We will be concerned with probability measures on \(M_N(\mathbb{C})^m_{sa}\) of the form

\[d\mu(x) = \frac{1}{Z} e^{-N^2V(x)} \, dx,\]

where \(x = (x_1, \ldots, x_m)\) is a tuple of self-adjoint matrices, \(V : M_N(\mathbb{C})^m_{sa} \to \mathbb{R}\), and \(Z = \int e^{-N^2V(x)} \, dx\) is the normalizing constant. Here \(dx\) denotes Lebesgue measure where we identify \(M_N(\mathbb{C})^m_{sa}\) with \(\mathbb{R}^{mnN^2}\) using the inner product associated to the trace (the normalization of Lebesgue measure is irrelevant here because if we multiply it by a constant, the normalizing constant \(Z\) for \(\mu\) will change to compensate). In this case, we will say that \(\mu\) is the measure given by the potential \(V\). We will often assume \(V\) is convex. Note that \(\mu\) only determines \(V\) up
to an additive constant, but we will still say that “\( V \) is the potential corresponding to \( \mu \)” with a slight abuse of terminology.

A primary motivating example is \( V(x) = \tau_N(f(x)) \), where \( \tau_N = (1/N) \text{Tr} \) is the normalized trace and \( f \) is a non-commutative polynomial in \( x_1, \ldots, x_m \). Unlike the notation in many random matrix papers, we prefer to write \( N^2 \tau_N(f) \) rather than \( N \text{Tr}(f) \). This seems natural because \( \tau_N(f) \) is a function with dimension-independent normalization and it would make sense for self-adjoint elements of a tracial \( W^\star \)-algebra. Meanwhile, \( N^2 \) is the dimension of \( M_N(\mathbb{C})^N \) and also the “rate” (in the sense of large deviations) for the standard concentration estimates that hold when \( V \) is uniformly convex (see for instance [3] or §2.3 below).

2.1. Semi-convex and Semi-concave Functions.

**Definition 2.1.** Let \( A : M_N(\mathbb{C})_{sa}^m \to M_N(\mathbb{C})_{sa}^m \) be a linear transformation and let \( u : M_N(\mathbb{C})_{sa}^m \to \mathbb{R} \). We say that \( Hu \leq A \) if \( u(x) - (1/2)\langle Ax, x \rangle_2 \) is concave. We say that \( Hu \geq A \) if \( u(x) - (1/2)\langle Ax, x \rangle_2 \) is convex.

We will also regularly use the following observation.

**Lemma 2.2.** Suppose that \( u : M_N(\mathbb{C})_{sa}^m \to \mathbb{R} \) is continuously differentiable, and let \( A \) and \( B \) be self-adjoint linear transformations. Then the following are equivalent:

1. \( A \leq Hu \leq B \).
2. For each \( x \in M_N(\mathbb{C})_{sa}^m \), there exists \( p \in M_N(\mathbb{C})_{sa}^m \) such that

\[
\frac{1}{2} (A(x' - x), x' - x) \leq u(x') - u(x) - \langle p, x' - x \rangle_2 \leq \frac{1}{2} (B(x' - x), x' - x)_2
\]

for all \( x' \in M_N(\mathbb{C})_{sa}^m \).

3. \( u \) is continuously differentiable and we have

\[
\langle A(x' - x), x' - x \rangle_2 \leq \langle Du(x') - Du(x), x' - x \rangle_2 \leq \langle B(x' - x), x' - x \rangle_2
\]

for all \( x, x' \in M_N(\mathbb{C})_{sa}^m \).

Moreover, in this case, \( Du \) is max(\( ||A||, ||B|| \))-Lipschitz with respect to \( ||\cdot||_2 \).

**Sketch of proof.** (1) \( \implies \) (3). Suppose (1) holds. If \( C = \max(||A||, ||B||) \), then for each \( x \) there exists \( p \) such that

\[
\frac{1}{2} C \|x' - x\|^2_2 \leq u(x') - u(x) - \langle p, x' - x \rangle_2 \leq \frac{1}{2} \|x' - x\|^2_2.
\]

We showed in [17, Proposition 2.13] that \( u \) must be continuously differentiable and \( Du \) is \( C \)-Lipschitz (which proves the last claim of our lemma as well). Now we prove the inequality asserted by (3). Using a similar argument as in [17, Proposition 2.13]), we can reduce the proof of the inequality to the case when \( u \) is smooth. But in the smooth case, the claim follows by estimating from above and below the formula

\[
\langle Du(x') - Du(x), x' - x \rangle_2 = \int_0^1 \langle Hu(x + t(x' - x))(x' - x), x' - x \rangle_2 dt.
\]

(3) \( \implies \) (2). Recall the formula

\[
u(x') - u(x) = \int_0^1 \langle Du(x + t(x' - x)), x' - x \rangle dt.
\]
This implies that
\[
(u'(x') - u(x) - \langle Du(x), x' - x \rangle)_2 = \int_0^1 \langle Du(x + t(x' - x)) - Du(x), x' - x \rangle_2 dt \\
= \int_0^1 \frac{1}{t} \langle Du(x + t(x' - x)) - Du(x), [x + t(x' - x)] - x \rangle_2 dt \\
\leq \int_0^1 \frac{1}{t} \| B[t(x' - x)], t(x' - x) \|_2 dt \\
= \frac{1}{2} \langle B(x' - x), x' - x \rangle_2.
\]
This proves the upper bound, and the lower bound is symmetrical. □

**Lemma 2.3.** Suppose that \( 0 \leq Hu \leq A \). Then we have
\[
\| (Du(x) - Du(x'))_2 \| \leq \langle A(x - x'), x - x' \rangle_{2} (Ay, Ay)_{2}^{1/2}.
\]

**Proof.** As in [17, Proposition 2.13], it suffices to prove the claim for smooth functions \( u \). In this case,
\[
\langle Du(x) - Du(x'), y \rangle = \int_0^1 \langle Hu(tx + (1 - t)x'), (x - x'), y \rangle dt \\
\leq \int_0^1 \langle Hu(tx + (1 - t)x'), (x - x') \rangle^{1/2} (Ay, y)_{2}^{1/2} dt \\
\leq \int_0^1 \langle A(x - x'), (x - x') \rangle^{1/2} (Ay, y)_{2}^{1/2} dt \\
= \langle A(x - x'), x - x' \rangle^{1/2} (Ay, y)_{2}^{1/2}.
\]
□

Consider a function \( u(x, y) \) on \( M_N(\mathbb{C})_{sa} \times M_N(\mathbb{C})_{sa} \). Let us write \( Hu \geq cI_m \oplus c'I_n \) to mean that
\[
u(x, y) - \frac{c}{2} \| x \|_2^2 - \frac{c'}{2} \| y \|_2^2
\]
is convex
and similarly let us write \( Hu \leq CI_m \oplus C'I_n \) to mean that
\[
u(x, y) - \frac{C}{2} \| x \|_2^2 - \frac{C'}{2} \| y \|_2^2
\]
is concave.

## 2.2. Some Basic Lemmas.

**Lemma 2.4.** Let \( V : M_N(\mathbb{C})_{sa} \to \mathbb{R} \) satisfy \( c \leq HV \leq C \) for some \( 0 < c \leq C \). Let \( \mu \) be the probability measure given by \( d\mu(x) = e^{-N^2V(x)} dx \) is the corresponding probability measure and \( X \sim \mu \), then
\[
E[DV(X)] = 0
\]
and
\[
\frac{m}{C} \leq E[X - E(X)]_2^2 \leq \frac{m}{c}.
\]

**Proof.** \( E[DV(X)] = 0 \) follows from integration by parts (see [0,2] for further context on this integration by parts). Also, by integration by parts, \( E(D_jV(X), X_j - E(X)_j)_2 = 1 \), so that
\[
E(DV(X) - DV(E(X)_j), X - E(X)_j)_2 = E(DV(X), X - E(X)_2) = m.
\]
On the other hand, by Lemma 2.3
\[
cE[X - E(X)]_2^2 \leq E(DV(X) - DV(E(X)_j), X - E(X)_j)_2 \leq CE[X - E(X)]_2^2.
\]
Since the middle term evaluates to \( m \), the proof is complete. □
Lemma 2.5. Let $X$ be a random variable in $M_N(\mathbb{C})_{sa}^m$ and let $G : M_N(\mathbb{C})_{sa}^m \to M_N(\mathbb{C})_{sa}^n$ be Lipschitz with respect to $\| \cdot \|_2$. Then
\[
\|G(x)\|_2 \leq \|E(G(X))\|_2 + \|G\|_{\text{Lip}} \left( \|x - E(X)\|_2 + (E \|X - E(X)\|_2^2)^{1/2} \right).
\]

Proof. Note that
\[
\|G(x) - E(G(X))\|_2 \leq \|G\|_{\text{Lip}} \|x - X\|_2 \\
\leq \|G\|_{\text{Lip}} (\|x - E(X)\|_2 + E \|X - E(X)\|_2) \\
\leq \|G\|_{\text{Lip}} (\|x - E(X)\|_2^2 + (E \|X - E(X)\|_2^2)^{1/2}).
\]

\[\square\]

Corollary 2.6. Let $V : M_N(\mathbb{C})_{sa}^m \to \mathbb{R}$ satisfies $e \leq HV \leq C$, let $\mu$ be the corresponding measure, and let $X \sim \mu$. Then
\[
\|DV(x)\|_2 \leq C \left( \|x - E(X)\|_2 + \frac{m^{1/2}}{c^{1/2}} \right).
\]

Proof. We apply Lemma 2.5 to $DV(X)$. Also, $DV$ is $C$-Lipschitz by Lemma 2.2. By Lemma 2.4 $E(DV(X)) = 0$ and $E \|X - E(X)\|_2^2 \leq m/c$. \[\square\]

Lemma 2.7. Let $A$ and $B$ be positive definite linear transformations $M_N(\mathbb{C})_{sa}^m \to M_N(\mathbb{C})_{sa}^n$. Let $\{V_k\}_{k \in \mathbb{N}}$ be a sequence of functions such that $A \leq HV_k \leq B$. Let $d\mu_k(x) = \frac{1}{Z_k} e^{-N^2V_k(x)} dx$ be the associated probability measure. Let $\mu$ be another measure with finite mean. Suppose $\mu_k$ converges weakly to $\mu$ and the mean of $\mu_k$ is bounded in $\| \cdot \|_2$ as $k \to \infty$. Then there exists $V$ such that $d\mu(x) = \frac{1}{Z} e^{-N^2V(x)} dx$ and $A \leq HV \leq B$.

Proof. Since adding a constant to $V_k$ does not change $\mu_k$, we can assume without loss of generality that $V_k(0) = 0$. Now $DV_k$ is $C$-Lipschitz, hence the sequence is equicontinuous. It is also pointwise bounded in light of the previous lemma, since we assumed the mean of $\mu_k$ is bounded as $k \to \infty$. Thus, by the Arzelà-Ascoli theorem, by passing to a subsequence, we can assume that $DV_k$ converges locally uniformly to some $F$ as $k \to \infty$. Since $V_k(0) = 0$, this also implies that $V_k$ converges locally uniformly to some $V$, which must satisfy $A \leq HV_k \leq B$ since the family of such functions is closed under pointwise limits. Moreover, $DV = F$.

Let $\nu$ be the probability measure given by $d\nu(x) = \frac{1}{Z} e^{-N^2V(x)} dx$. Since $A$ is positive definite, we have $A \geq c$ for some constant $c > 0$. Because $DV_k$ is bounded in $\| \cdot \|_2$ and $V_k(x) \geq (x, DV_k(0))_2 + c|x|^2_2$, we can see using the dominated convergence theorem that $Z_k \to Z$ as $k \to \infty$. It follows again from dominated convergence that $\int \phi d\mu_k \to \int \phi d\nu$ for every continuous compactly supported $\phi$. Hence, $\nu = \mu$, so $\mu$ is given by the potential $V$. \[\square\]

2.3. Log-Sobolev Inequality and Concentration. Log-concave matrix models exhibit concentration of measure as $N \to \infty$ as a consequence of the following classical inequalities.

Definition 2.8. We say that a measure $\mu$ on $\mathbb{R}^m$ satisfies the log-Sobolev inequality with constant $c$ if for all sufficiently smooth $f$,
\[
\int f^2 \log \frac{f^2}{\int f^2 d\mu} d\mu \leq 2c \int |\nabla f|^2 d\mu.
\]

Definition 2.9. We say that a measure $\mu$ on $\mathbb{R}^m$ satisfies Herbst’s concentration inequality with constant $c$ if for all Lipschitz functions $f : \mathbb{R}^m \to \mathbb{R}$ and $\delta > 0$, we have
\[
P(f(X) - E[f(X)] \geq \delta) \leq e^{-c\delta^2/2\|f\|_{\text{Lip}}^2},
\]
where $X$ is a random variable distributed according to $\mu$. Note that by symmetry this implies
\[
P(|f(X) - E[f(X)]| \geq \delta) \leq 2e^{-c\delta^2/2\|f\|_{\text{Lip}}^2}.
\]
The following theorem is now standard. See for instance [4, §2.3.3 and 4.4.2] and [4]. Below (1) is due to Bakry and Emery and (2) is due to Herbst. The application to random matrices was introduced by Guionnet.

**Theorem 2.10.**

(1) Suppose that $\mu$ is a probability measure on $\mathbb{R}^m$ satisfying $d\mu(x) = (1/Z)e^{-V(x)}dx$ and suppose that $V(x) - (c/2)|x|^2$ is convex. Then $\mu$ satisfies the log-Sobolev inequality with constant $1/c$.

(2) If $\mu$ satisfies the log-Sobolev inequality with constant $1/c$, then it satisfies Herbst’s concentration inequality with constant $c$.

In particular, we have the following consequences for random matrices. Here we use the gradient $Df$ and Hessian $Hf$ with respect to the normalized inner product $\langle \cdot, \cdot \rangle_2$.

**Corollary 2.11.** Suppose that $V : M_N(\mathbb{C})^m_{sa} \to \mathbb{R}$ satisfies $HV \succeq c > 0$ and let $d\mu(x) = (1/Z)e^{-N^2V(x)}dx$. Then $\mu$ satisfies the normalized log-Sobolev inequality

$$\int f^2 \log \frac{f^2}{\int f^2 d\mu} d\mu \leq \frac{2}{N^2c} \int \|Df\|_2^2 d\mu,$$

and hence also satisfies the normalized Herbst concentration inequality

$$P(f(X) - E[f(X)] \geq \delta) \leq e^{-cN\delta^2/2}\|f\|_{\text{Lip}},$$

where $f : M_N(\mathbb{C})^m_{sa} \to \mathbb{R}$ is Lipschitz and $\|f\|_{\text{Lip}}$ denotes the Lipschitz norm with respect to $\|\cdot\|_2$.

**Lemma 2.12.** Suppose that $\mu$ is a probability measure on $M_N(\mathbb{C})^m_{sa}$ satisfying (2.5) for some constant $c$ (and for this to make sense, we also assume that Lipschitz functions of $X^{(N)}$ are integrable). Let $f : M_N(\mathbb{C})^m_{sa} \to M_N(\mathbb{C})_{sa}$ be Lipschitz with respect to $\|\cdot\|_2$. Then we have

$$P\left(\|f(X) - E[f(X)]\|_{\infty} \geq c^{-1/2}\|f\|_{\text{Lip}}(\Theta + \delta)\right) \leq e^{-N\delta^2/2},$$

where $X \sim \mu$ and where $\Theta$ is a universal constant (independent of $N$ and $c$).

**Proof.** First, observe that $\|x\|_{\infty} \leq N^{1/2}\|x\|_2$ for $x \in M_N(\mathbb{C})^m_{sa}$. In particular, $g(X) = \|f(X) - E[f(X)]\|_{\infty}$ is $N^{1/2}\|f\|_{\text{Lip}}$-Lipschitz with respect to $\|\cdot\|_2$, and thus

$$P(g(X) \geq E[g(X)] + \delta) \leq e^{-cN\delta^2/2}\|f\|^2_{\text{Lip}},$$

which implies after a change of variables for $\delta$ that

$$P(g(X) \geq E[g(X)] + c^{-1/2}\|f\|_{\text{Lip}}\delta) \leq e^{-N\delta^2/2}.$$

Therefore, it suffices to show that for some constant $\Theta$, we have

$$E[g(X)] = E[\|f(X) - E[f(X)]\|_{\infty}] \leq \Theta c^{-1/2}\|f\|_{\text{Lip}}.$$

We may assume without loss of generality that $f$ is self-adjoint since in the general case, $f = (1/2)(f + f^*) + i(1/2)(f - f^*)$, and each of the terms on the right hand side is Lipschitz. Thus, the self-adjoint case would imply the non-self-adjoint case at the cost of doubling the constant $\Theta$. Now to prove self-adjoint case, we use an “$e$-net argument” that is well-known in random matrix theory (see [27, §2.3.1]). Fix $N$. Let $\{\eta_j\}_{j=1}^J$ be a maximal collection of unit vectors in $\mathbb{C}^N$ such that $|\eta_i - \eta_j| \geq 1/3$ for all $i \neq j$. Since this collection is maximal, for every
unit vector \( \eta \), there exists some \( \eta_j \) with \(|\eta - \eta_j| < 1/3\). Now if \( a \in M_N(\mathbb{C})_{sa} \), then there is a unit vector with \( \|a\|_\infty = \langle \eta, a\eta \rangle \). We may then choose \( \eta_j \) with \(|\eta - \eta_j| < 1/3\) and

\[
\|a\|_\infty = \langle \eta, a\eta \rangle = \langle \eta_j, a\eta_j \rangle + \langle \eta_j, a(\eta - \eta_j) \rangle + \langle \eta - \eta_j, a\eta \rangle \\
\leq \langle \eta_j, a\eta_j \rangle + \frac{1}{3}\|a\|_\infty + \frac{1}{3}\|a\|_\infty,
\]

so that

\[
\|a\|_\infty \leq 3 \max_j \langle \eta_j, a\eta_j \rangle.
\]

Note that the balls \( \{B(\eta_j, 1/6)\}_{j=1}^J \) in \( \mathbb{C}^N \) are disjoint and contained in \( B(0, 7/6) \). Hence, we can estimate the number of vectors by

\[
J \leq \left| \frac{B(0, 7/6)}{B(0, 1/6)} \right| = 7^2 N.
\]

Let \( K = \|f\|_{\text{Lip}} \). For a matrix \( a \in M_N(\mathbb{C})_{sa} \), we have

\[
|\langle \eta_i, a\eta_j \rangle| \leq \|a\|_\infty \leq N^{1/2}\|a\|_2.
\]

This implies that \( x \mapsto \langle \eta_j, f(x)\eta_j \rangle \) is \( KN^{1/2} \)-Lipschitz with respect to \( \|\cdot\|_2 \) and hence

\[
P(\langle \eta_j, f(X) - E[f(X)]\rangle \eta_j \geq \delta) \leq e^{-cN\delta^2/2K^2}
\]

Since \( \|a\|_\infty \leq 3 \max_j \langle \eta_j, a\eta_j \rangle \), we have

\[
P(\|f(X) - E[f(X)]\|_\infty \geq 3\delta) \leq Je^{-cN\delta^2/2} \leq 7^2 Ne^{-cN\delta^2/2K^2}.
\]

Thus, for any \( t_0 > 0 \), we have

\[
E(\|f(X) - E[f(X)]\|_\infty) = \int_0^\infty P(\|f(X) - E[f(X)]\|_\infty \geq t) \, dt \\
\leq \int_0^{t_0} 1 \, dt + \int_{t_0}^\infty 7^2 Ne^{-cNt^2/18K^2} \, dt \\
\leq t_0 + \int_{t_0}^\infty 7^2 Ne^{-cNt^2/18K^2} \, dt \\
= t_0 + 7^2 N \frac{9K^2}{cNt_0} e^{-cNt_0^2/18K^2}.
\]

Now substitute \( t_0 = 6c^{-1/2}K(\log 7)^{1/2} \) and obtain (2.7) with

\[
\Theta = 6(\log 7)^{1/2} + \frac{9}{6(\log 7)^{1/2}}.
\]

(In fact, for a fixed \( N \), we may use \( \Theta_N = 6(\log 7)^{1/2} + 9/6N(\log 7)^{1/2} \) in the self-adjoint case.) □
3. Functional Calculus and Asymptotic Approximation

In this section, we introduce the algebra \( \text{TrP}^1_m \) of trace polynomials in self-adjoint variables \( X_1, \ldots, X_m \), as well as a certain completed quotient \( \overline{\text{TrP}}^1_m \) of this algebra. The elements of \( \text{TrP}^1_m \) represent functions that can be applied to \( \text{any} \) tuple of self-adjoint non-commutative random variables \( (X_1, \ldots, X_m) \) in an \( \mathcal{R}^\omega \)-embeddable tracial \( W^* \)-algebra, and application of these functions will produce every element of \( L^2(W^*(X_1, \ldots, X_m)) \). These functions are closed under certain algebraic and composition operations. Moreover, they are a natural tool to describe the large \( N \) limit of functions on \( M_N(\mathbb{C})_s^m \), which we will apply in the rest of the paper.

3.1. The Algebra of Trace Polynomials. We denote by \( \text{NCP}_m = \mathbb{C}(X_1, \ldots, X_m) \) the *-algebra of polynomials in \( m \) self-adjoint non-commuting variables \( X_1, \ldots, X_m \).

We denote by \( \text{TrP}_m^0 \) the *-algebra of scalar-valued trace polynomials. A formal definition is given in [17]. Informally, this is the commutative *-algebra generated by functions of the form \( \tau(p(X_1, \ldots, X_m)) \), where \( p \) is a non-commutative polynomial in \( X = (X_1, \ldots, X_m) \) and \( \tau \) is a formal symbol (which stands in for a normalized trace on a von Neumann algebra), where \( \tau(p(X))^* = \tau(p(X)^*) \), and where we identify \( \tau(p(X)q(X)) \) with \( \tau(q(X)p(X)) \) for all polynomials \( p \) and \( q \). Thus, \( \text{TrP}_m^0 \) is spanned as a vector space by elements of the form \( \tau(p_1(X)) \cdots \tau(p_n(X)) \) where \( p_1, \ldots, p_n \in \text{NCP}_m \).

We denote by \( \text{TrP}_m^k \) the *-algebra of operator-valued trace polynomials. This is the *-algebra generated formally as \( \text{TrP}_m^0 \otimes \text{NCP}_m \). As a vector space, it is spanned by elements of the form \( \tau(p_1(X)) \cdots \tau(p_n(X))q(X), \) where \( p_1, \ldots, p_n \) and \( q \) are in \( \text{NCP}_m \). More generally, we would denote \( \text{TrP}_m^k = \text{TrP}_m^0 \otimes (\text{NCP}_m)^\otimes k \), but these spaces will not be needed in this paper.

The degree of a trace polynomial is defined as one would expect; see [17, §3.1] for precise explanation.

Suppose that \( x_1, \ldots, x_m \) are self-adjoint elements of a tracial von Neumann algebra \( (\mathcal{M}, \tau_0) \). Then elements of \( \text{NCP}_m, \text{TrP}_m^0, \) and \( \text{TrP}_m^1 \) can be evaluated on \( (x_1, \ldots, x_m) \) and \( \tau_0 \) by substituting the operator \( x_j \) and the trace \( \tau_0 \) in place of the formal symbols \( X_j \) and \( \tau \). More precisely, the evaluation map \( \varepsilon_{(x_1, \ldots, x_m)} : \text{NCP}_m \to \mathcal{M} \) is the unique *-algebra homomorphism that sends \( X_j \) to \( x_j \). Similarly, the evaluation map \( \varepsilon^0_{(x_1, \ldots, x_m)} : \text{TrP}_m^0 \to \mathbb{C} \) is the unique *-algebra homomorphism that sends \( \tau(p(X)) \) to \( \tau_0(\varepsilon_{(x_1, \ldots, x_m)}(p)) \). Finally, the evaluation map \( \varepsilon^1_{(x_1, \ldots, x_m)} : \text{TrP}_m^1 \to \mathcal{M} \) is \( \varepsilon^0_{(x_1, \ldots, x_m)} \otimes \varepsilon_{(x_1, \ldots, x_m)} \), that is, \( \tau(p_1(X)) \cdots \tau(p_n(X))q(X) \to \tau_0(\varepsilon^0_{(x_1, \ldots, x_m)}(p_1)) \cdots \tau_0(\varepsilon^0_{(x_1, \ldots, x_m)}(p_n))\varepsilon_{(x_1, \ldots, x_m)}(q) \).

For the most part, we will abuse notation and denote \( f(x) = \varepsilon_{(x_1, \ldots, x_m)}(f) \) when \( f \in \text{NCP}_m \), and similarly for \( f \in \text{TrP}_m^0 \) or \( f \in \text{TrP}_m^1 \). Note in particular that we can consider \( (\mathcal{M}, \tau_0) = (\mathcal{M}_N(\mathbb{C})_s^m, \tau_N) \) and thus \( f \) is defined for \( x \in \mathcal{M}_N(\mathbb{C})_s^m \) and \( f \in \text{TrP}_m^0 \) or \( \text{TrP}_m^1 \).

These evaluation maps thus allow us to view \( f \in \text{TrP}_m^0 \) as a function (or rather a family of functions) \( \mathcal{M}_N^m \to \mathbb{C} \) for every tracial \( W^* \)-algebra \( (\mathcal{M}, \tau) \) and in particular \( \mathcal{M}_N(\mathbb{C})_s^m \to \mathbb{C} \) for every \( m \). Similarly, every \( f \in \text{TrP}_m^1 \) defines a function \( \mathcal{M}_N^m \to \mathcal{M} \) for every tracial \( W^* \)-algebra and in particular a function \( \mathcal{M}_N(\mathbb{C})_s^m \to \mathcal{M}_N(\mathbb{C}) \) for every \( N \).

3.2. Functions Approximable by Trace Polynomials. From an analytic viewpoint, we prefer to work with certain “completions” of (the functions defined by) \( \text{TrP}_m^0 \) and \( \text{TrP}_m^1 \). In [17, §8.1], we sketched several equivalent ways of defining these completions. Here we emphasize their description as functions that can be evaluated on any self-adjoint tuple in \( \mathcal{R}^\omega \) (or, as we will see, any \( \mathcal{R}^\omega \)-embeddable \( W^* \)-algebra).
Let $\mathcal{R}$ denote the hyperfinite $\mathbb{II}_1$ factor (tracial $W^*$-algebra with trivial center) and let $\mathcal{R}^\omega$ be its (tracial $W^*$-algebra) ultrapower with respect to some free ultrafilter $\omega \in \mathcal{N} \setminus \mathbb{N}$.

Consider the case of $\operatorname{TrP}^0_m$ first. Let $\mathcal{F}^0_m$ denote the space of functions $(\mathcal{R}^\omega)^m_{sa} \to \mathbb{C}$ that are bounded on operator norm balls, equipped with the family of semi-norms

$$
\| \phi \|_{u,R} = \sup \{|\phi(x)| : \|x\|_\infty \leq R \}.
$$

(Here “$u$” stands for uniform.) This is clearly a Fréchet space since the topology is given by the countable family of semi-norms given by taking $R \in \mathbb{N}$ (for background on Fréchet spaces, see e.g. [11, §5.4]). Every $f \in \operatorname{TrP}^0_m$ defines a function $(\mathcal{R}^\omega)^m_{sa} \to \mathbb{C}$ that is a bounded an operator norm ball. In other words, evaluation produces a map $\operatorname{TrP}^0_m \to \mathcal{F}^0_m$. We denote by $\overline{\operatorname{TrP}}^0_m$ the closure of the image of this map in $\mathcal{F}^1_m$. In other words, $\overline{\operatorname{TrP}}^0_m$ is the space of functions $(\mathcal{R}^\omega)^m_{sa} \to \mathbb{C}$ that can be approximated uniformly on operator-norm balls by trace polynomials.

Remark 3.1. The notation $\overline{\operatorname{TrP}}^m$ is slightly abusive since we have not shown that the map $\operatorname{TrP}^m \to \mathcal{F}^m$ is injective (and perhaps it is not). However, we will still use the notation $\overline{\operatorname{TrP}}^m$ since it indicates the connection with trace polynomials.

Earlier, we saw that it makes sense to evaluate a trace polynomial $f$ on any self-adjoint tuple $(x_1, \ldots, x_m)$ in a tracial von Neumann algebra. In fact, $f(x_1, \ldots, x_m)$ makes sense for every $f \in \overline{\operatorname{TrP}}^m$ when $x_1, \ldots, x_m$ come from a tracial von Neumann algebra that embeds into $\mathcal{R}^\omega$. To see this, suppose $(\mathcal{M}, \tau)$ admits a normal trace-preserving embedding $\iota : \mathcal{M} \to \mathcal{R}^\omega$. Then we define $f(x_1, \ldots, x_m) = f(\iota(x_1), \ldots, \iota(x_m))$. This is independent of the choice of trace-preserving embedding if $f$ is a trace polynomial, and hence it must also be independent of the choice of embedding when $f$ is in $\overline{\operatorname{TrP}}^m$.

The same completion can be defined for $\operatorname{TrP}^1_m$. Indeed, let $\mathcal{F}^1_m$ be the set of functions $\phi : (\mathcal{R}^\omega)^m_{sa} \to L^2(\mathcal{R}^\omega)$ such that

$$
\| \phi \|_{u,R} := \sup \{|\phi(x)| : \|x\|_\infty \leq R \}
$$

is finite for each $R$. Again, this is a Fréchet space. Through the evaluation map, every trace polynomial defines an element of $\mathcal{F}^1_m$ and hence there is a linear map $\operatorname{TrP}^1_m \to \mathcal{F}^1_m$. We define $\overline{\operatorname{TrP}}^1_m$ to be the closure of the image of this map in $\mathcal{F}^1_m$.

Similar to the scalar-valued case, we can define evaluation of $f \in \overline{\operatorname{TrP}}^m$ for tuples in an $\mathcal{R}^\omega$-embeddable tracial $W^*$-algebra $(\mathcal{M}, \tau)$ by using any trace-preserving embedding $\iota : \mathcal{M} \to \mathcal{R}^\omega$. Indeed, let $x_1, \ldots, x_m \in \mathcal{M}_{sa}$. Clearly, for $f \in \operatorname{TrP}^1_m$, we have $f(\iota(x_1), \ldots, \iota(x_m)) \in \iota(L^2(\mathcal{M}))$ where the latter is defined by extending $\iota$ to a map $L^2(\mathcal{M}) \to L^2(\mathcal{R}^\omega)$. Since this holds for $f \in \operatorname{TrP}^1_m$, then by taking limits, we have $f(\iota(x_1), \ldots, \iota(x_m)) \in \iota(L^2(\mathcal{M}))$ for all $f \in \overline{\operatorname{TrP}}^1_m$. Therefore, we may define $f(x_1, \ldots, x_m)$ by $f(\iota(x_1), \ldots, \iota(x_m))$.

3.3. Asymptotic Approximation for Functions of Matrices. Our earlier work introduced asymptotic approximability by trace polynomials for a sequence of functions on $M_N(\mathbb{C})^m_{sa}$, which is a precise description of good asymptotic behavior as $N \to \infty$ suitable for free probabilistic analysis in the limit.

Definition 3.2. Let $\phi^{(N)} : M_N(\mathbb{C})^m_{sa} \to \mathbb{C}$. We say that $\{\phi^{(N)}\}$ is asymptotically approximable by trace polynomials if for every $R > 0$ and $\epsilon > 0$, there exists $f \in \operatorname{TrP}^0_m$ such that

$$
\limsup_{N \to \infty} \sup_{x \in M_N(\mathbb{C})^m_{sa}} \{\|\phi^{(N)}(x) - f(x)\|_\infty \leq R \} = 0.
$$
Similarly, for matrix-valued functions $\phi^{(N)} : M_N(\mathbb{C})^m \to M_N(\mathbb{C})$, we say that $\{\phi^{(N)}\}$ is asymptotically approximable by trace polynomials if for every $R > 0$ and $\epsilon > 0$, there exists $f \in \text{TrP}_m$ such that

$$\lim_{N \to \infty} \sup_{x \in M_N(\mathbb{C})^m, \|x\|_\infty \leq R} \|\phi^{(N)}(x) - f(x)\|_2 \leq \epsilon.$$ 

It will be convenient to denote

$$\|\phi^{(N)}\|_{u,R}^{(N)} = \sup_{x \in M_N(\mathbb{C})^m, \|x\|_\infty \leq R} |\phi(x)|$$

in the scalar-valued case and similarly for the matrix-valued case with $\|\phi(x)\|_2$ rather than $|\phi(x)|$. Thus, for instance, the preceding definition says that there exists a trace polynomial $f$ with

$$\lim_{N \to \infty} \|\phi^{(N)} - f\|_{u,R}^{(N)} < \epsilon.$$ 

Moreover, it is implicit from our discussion in [17, §8.1] that if $\phi^{(N)}$ is asymptotically approximable by trace polynomials, then it will be asymptotic to some $f \in \text{TrP}_m^0$ or $\text{TrP}_m^1$ in the following sense.

**Definition 3.3.** Let $\phi^{(N)} : M_N(\mathbb{C})^m \to \mathbb{C}$ or $M_N(\mathbb{C})$ respectively, and let $f \in \text{TrP}_m^0$ or $\text{TrP}_m^1$ respectively. Then we say that $\{\phi^{(N)}\}$ is asymptotic to $f$, or $\phi^{(N)} \sim f$ if for every $R > 0$,

$$\lim_{N \to \infty} \|\phi^{(N)} - f\|_{u,R}^{(N)} = 0.$$ 

Similarly, if $\phi^{(N)} : M_N(\mathbb{C})^m \to M_N(\mathbb{C})$ and $f \in \text{TrP}_m^1$, we make the same definitions with $|\phi^{(N)}(x) - f(x)|$ replaced by $\|\phi^{(N)}(x) - f(x)\|_2$.

**Lemma 3.4.** Let $\phi^{(N)} : M_N(\mathbb{C})^m \to \mathbb{C}$ (respectively, $\to M_N(\mathbb{C})$). Then $\phi^{(N)}$ is asymptotically approximable by trace polynomials if and only if there exists $f \in \text{TrP}_m^0$ (respectively, $f \in \text{TrP}_m^1$) such that $\phi^{(N)} \sim f$. Moreover, $\|f\|_{u,R} = \lim_{N \to \infty} \|\phi^{(N)}\|_{u,R}^{(N)}$ for each $R$.

Proof. We record the proof only for the case of scalar-valued functions, since the proof for operator-valued case is identical with minor changes of notation. Suppose that $\{\phi^{(N)}\}$ is asymptotically approximable by trace polynomials. Then there exists a sequence $\{f_k\}$ of trace polynomials such that for every $R > 0$,

$$\lim_{k \to \infty} \lim_{N \to \infty} \sup_{x \in M_N(\mathbb{C})^m, \|x\|_\infty \leq R} |\phi^{(N)}(x) - f_k(x)| = 0.$$ 

As in [17, Lemma 8.1], if $g \in \text{TrP}_m^0$, then

$$\sup_{x \in (\mathbb{R}^m)^m, \|x\|_\infty \leq R} |g(x)| = \sup_{N \in \mathbb{N}} \sup_{x \in M_N(\mathbb{C})^m, \|x\|_\infty \leq R} |g(x)| = \lim_{N \to \infty} \sup_{x \in M_N(\mathbb{C})^m, \|x\|_\infty \leq R} |g(x)|$$

which implies that

$$\|g\|_{u,R} = \lim_{N \to \infty} \|g^{(N)}\|_{u,R}.$$ 

Applying this to $g = f_j - f_k$, we obtain from the triangle inequality

$$\|f_j - f_k\|_{u,R} \leq \lim_{N \to \infty} \|f_j(x) - \phi^{(N)}\|_{u,R} + \lim_{N \to \infty} \|\phi^{(N)} - f_k\|_{u,R}.$$
and hence $f_k$ is Cauchy with respect to $\|\cdot\|_{u,R}$ for each $R > 0$. Hence, $f_k$ converges to some $f \in \overline{\text{TrP}}^0_m$. By similar use of the triangle inequality,

$$\limsup_{N \to \infty} \left\| \phi^{(N)}(x) - f \right\|_{u,R} \leq \|f - f_k\|_{u,R} + \limsup_{N \to \infty} \left\| \phi^{(N)} - f_k \right\|_{u,R}.$$ 

Hence, $\phi^{(N)} \rightharpoonup f$.

Conversely, suppose that $\phi^{(N)} \rightharpoonup f \in \overline{\text{TrP}}^0_m$. Choose $f_k \in \text{TrP}_m^0$ such that $\|f_k - f\|_{u,R} \to 0$ for every $R$. Then

$$\limsup_{N \to \infty} \left\| \phi^{(N)}(x) - f_k \right\|_{u,R} \leq \|f - f_k\|_{u,R} + \limsup_{N \to \infty} \left\| \phi^{(N)} - f_k \right\|_{u,R}.$$ 

Hence, it follows that $\{\phi^{(N)}\}$ is asymptotically approximable by trace polynomials, namely the polynomials $\{f_k\}$.

We leave the proof of the last claim that $\|f\|_{u,R} = \lim_{N \to \infty} \|\phi^{(N)}\|_{u,R}$ to the reader. □

**Remark 3.5.** If $\phi^{(N)} : M_N(\mathbb{C})_{sa} \to M_N(\mathbb{C})_{sa}$ and $\{\phi^{(N)}\}$ is asymptotically approximable by trace polynomials, then we can asymptotically approximate it using self-adjoint trace polynomials. Indeed, if

$$\limsup_{N \to \infty} \left\| \phi^{(N)} - f \right\|_{u,R} \leq \epsilon,$$

then the same holds with $f$ replaced by $(1/2)\left(f + f^*\right)$. Similarly, if $\phi^{(N)}(x)$ is self-adjoint and $\phi^{(N)} \rightharpoonup f \in \overline{\text{TrP}}^1_m$, then $f$ must be self-adjoint.

**Remark 3.6.** Definitions 3.2 and 3.3 and Lemma 3.4 extend naturally to tuples $f = (f_1, \ldots, f_n) \in (\overline{\text{TrP}}^1_m)^n$ and $\phi^{(N)} = (\phi^{(N)}_1, \ldots, \phi^{(N)}_n) : M_N(\mathbb{C})_{sa} \to M_N(\mathbb{C})^n$. We shall apply them to tuples without further comment in the rest of the paper.

### 3.4. Algebra, Composition, and Limits.

**Lemma 3.7.** $\overline{\text{TrP}}^0_m$ is an algebra and $\overline{\text{TrP}}^1_m$ is a module over $\overline{\text{TrP}}^0_m$. Also, if $f, g \in \overline{\text{TrP}}^0_m$, then $\tau(fg) \in \overline{\text{TrP}}^0_m$. Moreover, suppose that $\phi^{(N)}, \psi^{(N)} : M_N(\mathbb{C})_{sa} \to \mathbb{C}$ and $f^{(N)}, g^{(N)} : M_N(\mathbb{C})_{sa} \to M_N(\mathbb{C})$ are asymptotically approximable, and $\phi^{(N)} \rightharpoonup \phi$, $\psi^{(N)} \rightharpoonup \psi$, $f^{(N)} \rightharpoonup f$, and $g^{(N)} \rightharpoonup g$. Then we have

$$\phi^{(N)} + \psi^{(N)} \rightharpoonup \phi + \psi,$$

$$\phi^{(N)}\psi^{(N)} \rightharpoonup \phi\psi,$$

$$f^{(N)} + g^{(N)} \rightharpoonup f + g,$$

$$\phi^{(N)}f^{(N)} \rightharpoonup \phi f,$$

$$\tau_N(f^{(N)}g^{(N)}) \rightharpoonup \tau(fg).$$

**Proof.** Since the proofs of all the statements are straightforward and similar to each other, we will only explain how to show that if $\phi \in \overline{\text{TrP}}^0_m$ and $f \in \overline{\text{TrP}}^1_m$, then $\phi f \in \overline{\text{TrP}}^1_m$ and that if $\phi^{(N)} \rightharpoonup \phi$ and $f^{(N)} \rightharpoonup f$, then $\phi^{(N)} f^{(N)} \rightharpoonup \phi f$.

First, note that $\phi f$ is well-defined as a function on $(\mathbb{R}^\omega)_{sa}$ by multiplying the scalar $\phi(x)$ times the vector $f(x)$ for each $x \in (\mathbb{R}^\omega)_{sa}$, and also clearly $\|\phi f\|_{u,R} \leq \|\phi\|_{u,R} \|f\|_{u,R}$. To show that $\phi f \in \overline{\text{TrP}}^1_m$, it suffices to show that for every $\epsilon > 0$ and $R > 0$, the function $\phi f$ can be
approximated by an element of $\text{TrP}_m^1$ with respect to $\|\cdot\|_{u,R}$ with error less than $\epsilon$. We first choose $h \in \text{TrP}_m^1$ such that

$$\|h - f\|_{u,R}\|\phi\|_{u,R} < \frac{\epsilon}{2}.$$ 

Then we choose $\theta \in \text{TrP}_m^0$ such that

$$\|\theta - \phi\|_{u,R}\|h\|_{u,R} < \frac{\epsilon}{2},$$

and we conclude with the routine observation that

$$\|\theta h - \phi f\|_{u,R} \leq \|(\theta - \phi) \cdot h\|_{u,R} + \|\phi \cdot (h - f)\|_{u,R}$$

$$\quad \leq \|\theta - \phi\|_{u,R}\|h\|_{u,R} + \|\phi\|_{u,R}\|h - f\|_{u,R} < \epsilon.$$

Next, to show $\phi^{(N)} f^{(N)} \rightsquigarrow \phi f$, first observe that

$$M := \sup_N \|\phi^{(N)}\|_{u,R} < +\infty.$$ 

Then

$$\left\|\phi^{(N)} f^{(N)} - \phi f\right\|_{u,R} \leq \left\|\phi^{(N)} \cdot (f^{(N)} - f)\right\|_{u,R} + \left\|(\phi^{(N)} - \phi) \cdot f\right\|_{u,R}$$

$$\quad \leq M \left\|f^{(N)} - f\right\|_{u,R} + \left\|\phi^{(N)} - \phi\right\|_{u,R}\|f\|_{u,R},$$

which implies that $\|\phi^{(N)} f^{(N)} - \phi f\|_{u,R} \to 0$. \qed

In addition to their algebraic structure, functions $(\mathcal{R}^\omega)^m_{sa} \to (\mathcal{R}^\omega)^n_{sa}$ given by trace polynomials are closed under composition. It turns out that self-adjoint tuples from $\text{TrP}_m^1$ are closed under composition under the assumption of $\|\cdot\|_2$-uniform continuity of the “outside” function (Lemma 5.10 below).

We say that $f \in \text{TrP}_m^1$ is $\|\cdot\|_2$-uniformly continuous if for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$\forall x, y \in (\mathcal{R}^\omega)^m_{sa}, \quad \|x - y\|_2 < \delta \implies \|f(x) - f(y)\|_2 < \epsilon.$$ 

Furthermore, we say $f \in \text{TrP}_m^1$ is $\|\cdot\|_2$-Lipschitz if $\|f(x) - f(y)\|_2 \leq K\|x - y\|_2$ for some constant $K$, which is an important special case of uniform continuity. We denote the minimum such constant by $\|f\|_{\text{Lip}}$. We make the analogous definitions for $f \in \text{TrP}_m^0$.

**Observation 3.8.** If $f$ is a function from $(\mathcal{R}^\omega)^m_{sa}$ to $\mathcal{R}^\omega$ or $\mathbb{C}$ that is $\|\cdot\|_2$-uniformly continuous, then it has a unique continuous extension to $L^2(\mathcal{R}^\omega)^m_{sa}$, which is also $\|\cdot\|_2$-uniformly continuous. Similarly, if $f$ is Lipschitz on $(\mathcal{R}^\omega)^m_{sa}$, then the extension is also Lipschitz.

**Lemma 3.9.** Suppose that $f^{(N)} : M_N(\mathbb{C})^m_{sa} \to \mathbb{C}$ or $M_N(\mathbb{C})$ and that $f^{(N)} \rightsquigarrow f$. If $f^{(N)}$ is $\|\cdot\|_2$-uniformly continuous with respect to some modulus of continuity independent of $N$, then $f$ is $\|\cdot\|_2$-uniformly continuous on $(\mathcal{R}^\omega)^m_{sa}$ with the same modulus of continuity.

**Proof.** Let us only explain the operator-valued case where $f^{(N)}$ is $M_N(\mathbb{C})$-valued and $f \in \text{TrP}_m^1$, since the scalar-valued case is easier. We define scalar-valued functions of $2m$ variables by $F^{(N)}(x, y) = \|f^{(N)}(x) - g^{(N)}(y)\|_2^2$ and $F(x, y) = \|f(x) - f(y)\|_2^2$. By Lemma 5.7, we have $F^{(N)} \rightsquigarrow F \in \text{TrP}_m^0$.

Let $\epsilon(\delta)$ be a common modulus of continuity for $f^{(N)}$. Let $x$ and $y \in (\mathcal{R}^\omega)^m_{sa}$. Then we may embed $W^*(x, y)$ into $(\mathcal{M}, \tau) := \prod_{N \to \omega} M_N(\mathbb{C}), \tau_N)$, that is the tracial $W^*$-ultraproduct of matrices. There exist tuples $x^{(N)}$ and $y^{(N)}$ of $N \times N$ matrices such that $x = \{x^{(N)}\}_{N \in \mathbb{N}}$ and
\[ y = \{ y^{(N)} \}_{N \in \mathbb{N}} \text{ in the ultraproduct and also } \| x^{(N)} \|_{\infty} \leq \| x \|_{\infty} \text{ and } \| y^{(N)} \|_{\infty} \leq \| y \|_{\infty}. \] Observe that
\[ F(x, y) = \lim_{N \to \omega} F(x^{(N)}, y^{(N)}). \]

(This equality holds for trace polynomials and hence holds for all functions in \( \text{TrP}_{\omega}^{m} \) by approximation.) On the other hand, we also have for \( R > \max(\| x \|_{\infty}, \| y \|_{\infty}) \) that
\[ |F^{(N)}(x^{(N)}, y^{(N)}) - F(x^{(N)}, y^{(N)})| \leq \| F - F^{(N)} \|_{u, R} \to 0. \]

Therefore,
\[ \| f(x) - f(y) \|_{2} = \lim_{N \to \omega} \| f^{(N)}(x^{(N)}) - f^{(N)}(y^{(N)}) \|_{2} \leq \lim_{N \to \omega} \epsilon \| \| x^{(N)} - y^{(N)} \|_{2} \| \leq \epsilon(\| x - y \|_{2}), \]
since \( \| x^{(N)} - y^{(N)} \|_{2} \to \| x - y \|_{2}. \]

**Lemma 3.10.** Let \( j = 0 \) or 1. Let \( f \in \text{TrP}_{\omega}^{m} \) be \( \| \cdot \|_{2} \)-uniformly continuous and let \( g = (g_{1}, \ldots, g_{m}) \in (\text{TrP}_{\omega}^{m})_{sa}. \)

1. Then \( f \circ g \) is a well-defined function on \((\mathbb{R}^{\omega})_{sa}^{m}, \) and it is in \( \text{TrP}_{\omega}^{m}. \)
2. If \( g \) is also \( \| \cdot \|_{2} \)-uniformly continuous, then so is \( f \circ g. \)
3. Suppose \( f^{(N)} \) is a function on \( M_{N}(\mathbb{C})_{sa}^{m} \) and \( g^{(N)} : M_{N}(\mathbb{C})_{sa}^{m} \to M_{N}(\mathbb{C})_{sa}^{m} \) such that \( f^{(N)} \rightsquigarrow f \) and \( g^{(N)} \rightsquigarrow g. \) Also, suppose that \( f^{(N)} \) is \( \| \cdot \|_{2} \)-uniformly continuous with the modulus of continuity also uniform in \( N. \) Then \( f^{(N)} \circ g^{(N)} \rightsquigarrow f \circ g. \)

**Proof.** (1) Because \( f \) extends to a function on \( L^{2}(\mathbb{R}^{\omega})_{sa}^{m}, \) we can define \( f \circ g. \) Now let us show \( f \circ g \in \text{TrP}_{\omega}^{m}. \) Choose \( \epsilon > 0 \) and \( R > 0. \) By uniform continuity of \( f, \) there exists a \( \delta > 0 \) such that \( \| x - y \|_{2} < \delta \) implies \( \| f(x) - f(y) \|_{2} < \epsilon/2 \) (for \( j = 0 \) or 1 respectively). Now choose \( \tilde{g} \in (\text{TrP}_{\omega}^{m})_{sa}^{m} \) such that \( \| \tilde{g} - g \|_{u, R} < \delta, \) and hence
\[ \| f \circ g - f \circ \tilde{g} \|_{u, R} < \frac{\epsilon}{2}. \]

Because \( \tilde{g} \) is a trace polynomial, there is some \( R' \) such that \( \| x \|_{\infty} \leq R \) implies \( \| \tilde{g} \|_{\infty} \leq R'. \)
Choose \( \tilde{f} \in \text{TrP}_{\omega}^{m} \) with \( \| \tilde{f} - f \|_{u, R'} < \epsilon/2, \) and hence
\[ \| f \circ \tilde{g} - \tilde{f} \circ \tilde{g} \|_{u, R} < \frac{\epsilon}{2}. \]
Then altogether we have
\[ \| f \circ g - \tilde{f} \circ \tilde{g} \|_{u, R} < \epsilon. \]

(2) This is immediate.

(3) This is similar to the proof of (1). Fix \( R > 0 \) and \( \epsilon > 0. \) Choose \( \delta > 0 \) such that \( \| x - y \|_{2} < \delta \) implies \( \| f(x) - f(y) \|_{2} < \epsilon/2 \) and such that the same holds for \( f^{(N)} \) as well. Let \( \tilde{g} \in (\text{TrP}_{\omega}^{m})_{sa}^{m} \) such that \( \| \tilde{g} - g \|_{u, R} < \delta. \) Note that for sufficiently large \( N, \) we have \( \| g^{(N)} - \tilde{g}^{(N)} \|_{u, R} < \delta \) and hence
\[ \| f^{(N)} \circ g^{(N)} - f^{(N)} \circ \tilde{g}^{(N)} \|_{u, R} < \epsilon/2. \]
Then let \( R' \) and \( \tilde{f} \) be as in (1). Then for sufficiently large \( N, \) we have
\[ \| f^{(N)} \circ \tilde{g} - \tilde{f} \circ \tilde{g} \|_{u, R} < \frac{\epsilon}{2}, \]
so overall
\[ \| f^{(N)} \circ g^{(N)} - \tilde{f} \circ \tilde{g} \|_{u, R} < \epsilon, \quad \| f \circ g - \tilde{f} \circ \tilde{g} \|_{u, R} < \epsilon, \]
so that \( \| f^{(N)} \circ g^{(N)} - f \circ g \|_{u,R} < 2 \epsilon \) for large enough \( N \). \( \square \)

Moreover, asymptotically approximable sequences are closed under limits in an appropriate sense.

**Lemma 3.11.** Let \( f_k^{(N)} : M_N(\mathbb{C})^m_{sa} \to \mathbb{C} \) or to \( M_N(\mathbb{C}) \) for \( k \) and \( N \in \mathbb{N} \). Suppose that \( f_k^{(N)} \sim f_k \) in \( \mathbb{TP}_m^1 \) for each \( k \), and that

\[
\lim_{k \to \infty} \limsup_{N \to \infty} \| f_k^{(N)} - f^{(N)} \|_{u,R} = 0 \quad \text{for all } R.
\]

Then \( f_k \) converges in \( \mathbb{TP}_m^1 \) to some \( f \), and we have \( f^{(N)} \sim f \).

**Proof.** Note that

\[
\| f_k - f_j \|_{u,R} = \lim_{N \to \infty} \| f_k^{(N)} - f_j^{(N)} \|_{u,R}
\]

Then because of our assumption (3.1), we see that \( \{ f_k \}_{k \in \mathbb{N}} \) is Cauchy with respect to \( \| \cdot \|_{u,R} \) for each \( R \). Thus, \( f_k \) converges to some \( f \). Then to show that \( f^{(N)} \sim f \) is a routine argument. \( \square \)

### 3.5. Functional Calculus and Operator Norm Bounds

Now we will show that every element of \( L^2(W^*(x_1, \ldots, x_m)) \) can be expressed as \( f(x_1, \ldots, x_m) \) for some \( f \in \mathbb{TP}_m^1 \). In fact, we can arrange that \( f \) can be approximated uniformly by Lipschitz functions. It will be convenient to define the uniform norm

\[
\| f \|_u = \sup_{R > 0} \| f \|_{u,R} = \sup_{x \in (\mathbb{R}^\omega)^m} \| f(x) \|_2,
\]

and we make the same definition for \( \| f \|^{(N)}_u \) where the supremum is instead taken over \( x \in M_N(\mathbb{C})^m_{sa} \).

**Proposition 3.12.** Let \( x_1, \ldots, x_m \) be self-adjoint variables which generate a tracial \( W^* \)-algebra \( (M, \tau) \) that is embeddable into \( \mathcal{R}^\omega \). Let \( z \in L^2(M, \tau) \).

1. There exists a \( \| \cdot \|_2 \)-uniformly continuous \( f \in \mathbb{TP}_m^1 \) such that \( Z = f(x_1, \ldots, x_m) \).
2. The \( f \) in (1) can be chosen so that there are \( \| \cdot \|_2 \)-Lipschitz functions \( f_k \in \mathbb{TP}_m^1 \) such that \( \| f_k - f \|_u \to 0 \).
3. If \( z \in C(x_1, \ldots, x_m) \), then \( f \) can be chosen to be \( \| \cdot \|_2 \)-Lipschitz.

We use the following auxiliary observation. Here \( \Sigma_{m,R} \) will denote the space of non-commutative laws for an \( m \)-tuple of operators with operator norm \( \leq R \). We equip \( \Sigma_{m,R} \) with the topology of convergence in moments. Recall that \( \Sigma_{m,R} \) is compact, separable, and metrizable. In [17, Lemma 8.2], we noted the relationship between \( \mathbb{TP}_m \) and continuous functions on \( \Sigma_{m,R} \) for each \( R \). This same idea motivates the proof of the next lemma.

**Lemma 3.13.** Let \( \mu \in \Sigma_{m,R} \) and let \( \mathcal{U} \) be a neighborhood of \( \mu \), and let \( \epsilon > 0 \). Then there exists a trace polynomial \( f \) such that

\[
\mu(f) = 1 \quad 0 \leq \nu(f) \leq 1_{\mathcal{U}} + \epsilon \quad \text{for } \nu \in \Sigma_{m,R}.
\]

**Proof.** By Urysohn’s lemma, there exists a continuous function \( F : \Sigma_{m,R} \to [0, 1] \) such that \( F(\mu) = 1 \) and \( F(\nu) = 0 \) for \( \nu \not\in \mathcal{U} \). The functions \( \Sigma_{m,R} \to \mathbb{C} \) of the form \( \mu \mapsto \mu(f) \) for \( f \in \mathbb{TP}_m^0 \) form a self-adjoint algebra in \( C(\Sigma_{m,R}) \), and they separate points because by definition two laws are the same if they agree on every non-commutative polynomial. So by the Stone-Weierstrass theorem, this algebra is dense in \( C(\Sigma_{m,R}) \). In particular, there exists a trace polynomial \( g \) such that \( |\nu(g) - F(\nu)| < \epsilon/2 \) for all \( \nu \in \Sigma_{m,R} \). Then let \( f = (g + \epsilon/2)/(g(\mu) + \epsilon/2) \). \( \square \)
We will also use the following smooth cut-off trick.

**Lemma 3.14.** Let \(0 < R' \leq R\). Let \(\phi \in C^\infty_c(\mathbb{R}; \mathbb{R})\) such that \(\phi(t) = t\) for \(t \leq R'\) and \(|\phi(t)| \leq R\).

For \(y \in (\mathcal{R}^\omega)^{sa}\), define \(\Phi(y) = \phi(y)\) where \(\phi\) is applied through functional calculus. Then

1. \(\Phi(y) = y\) if \(\|y\|_\infty \leq R'\).
2. \(\|\Phi(y)\|_\infty \leq R\) for all \(y\).
3. \(\Phi \in \mathcal{T}P^1_m\).
4. \(\Phi\) is globally \(\|\cdot\|_2\)-Lipschitz.

**Proof.** (1) and (2) follow from the properties of functional calculus. To prove (3), note by the Weierstrass approximation theorem that for every \(r > 0\), there is a polynomial \(p\) such that \(|p(t) - \phi(t)| < \epsilon\) for \(|t| \leq r\). This implies as with (1) that \(|p(y) - \phi(y)| < \epsilon\) for all \(y\) with \(\|y\|_\infty \leq r\). Claim (4) follows from the results of [13] (8.9) and Proposition 8.8.

**Proof of Proposition 3.12.** Let \(\mu\) be the law of \(x = (x_1, \ldots, x_m)\), and let \(R > \|X\|_\infty\). Since \(z \in L^2(M, \tau)\), there exist non-commutative polynomials \(\{p_k\}_{k=1}^\infty\) such that \(\|p_k(x) - z\|_2 < 1/2^{k+1}\) and hence for \(k \geq 1\),

\[
\|p_{k+1}(x) - p_k(x)\|_2 = \mu[(p_{k+1} - p_k)^2]^{1/2} \leq \frac{1}{2^{k+1}} + \frac{1}{2^k} < \frac{1}{2^k}.
\]

By scaling, we may assume without loss of generality that \(\|z\|_2 < 1\) and set \(p_0 = 0\), and then the above statement also holds for \(k = 0\). Now let

\[
U_k = \{\nu \in \Sigma_{m,R} : \nu((p_{k+1} - p_k)^2)^{1/2} < 1/2^k\},
\]

which is a neighborhood of \(\mu\) in \(\Sigma_{m,R}\). By the previous lemma, there exists a scalar-valued trace polynomial \(u_k\) such that \(\mu(u_k) = 1\) and

\[
0 \leq \nu(u_k) \leq \frac{1}{2^k\|p_{k+1} - p_k\|_{u,R}}.
\]

(We can assume without loss of generality that \(\|p_{k+1} - p_k\|_{u,R} \neq 0\).) Now the function \(u_k(p_{k+1} - p_k)\) will evaluate at the point \(X\) to \(p_{k+1}(x) - p_k(x)\). If \(y \in (\mathcal{R}^\omega)^{sa}\) with \(\|y\|_\infty \leq R\) and if the law of \(y\) is in \(U_k\), then we will have

\[
\|u_k(y)(p_{k+1}(y) - p_k(y))\|_2 \leq \|p_{k+1}(y) - p_k(y)\|_2 + \frac{1}{2^k\|p_{k+1} - p_k\|_{u,R}} \|p_{k+1}(y) - p_k(y)\|_2 \leq \frac{1}{2^k} + \frac{1}{2^k}.
\]

On the other hand, if the law of \(y\) is not in \(U_k\), then \(\|u_k(y)(p_{k+1}(y) - p_k(y))\|_2 \leq 1/2^k\). Overall, we have

\[
\|u_k \cdot (p_{k+1} - p_k)\|_{u,R} \leq \frac{2}{2^k}.
\]

This implies that \(\sum_{k=0}^\infty u_k \cdot (p_{k+1} - p_k)\) converges with respect to \(\|\cdot\|_{u,R}\) for our given choice of \(R\), and of course evaluating this function on \(X\) it produces the desired operator \(Z\) since \(u_k(x) = 1\).

To extend the function to be globally defined on \((\mathcal{R}^\omega)^{sa}\), we use the smooth cut-off trick. Let \(\Phi \in C^\infty_c(\mathbb{R}; \mathbb{R})\) such that \(\Phi(t) = t\) for \(|t| \leq \|X\|_\infty\) and \(|\Phi| \leq R\). For \(y = (y_1, \ldots, y_m) \in (\mathcal{R}^\omega)^{sa}\), let \(\Phi(y) = (\phi(y_1), \ldots, \phi(y_m))\). Then \(u_k \cdot (p_{k+1} - p_k) \circ \Phi \in \mathcal{T}P^1_m\) because it is the composition of a trace polynomial with a function \(\Phi \in (\mathcal{T}P^1_m)^{sa}\) that is uniformly bounded in operator norm.

Also, since \(\Phi\) is globally \(\|\cdot\|_2\)-Lipschitz and since \(u_k \cdot (p_{k+1} - p_k)\) is \(\|\cdot\|_2\)-Lipschitz on the operator norm ball of radius \(R\), we see that \(u_k \cdot (p_{k+1} - p_k) \circ \Phi\) is globally Lipschitz in \(\|\cdot\|_2\).
For all \( y \in (\mathcal{R}^n)_{sa}^m \),
\[
\|u_k(\Phi(y))(p_{k+1}(\Phi(y)) - p_k(\Phi(y)))\|_2 \leq \frac{2}{2^c}.
\]

Therefore,
\[
f(y) := \sum_{k=0}^{\infty} u_k(\Phi(y))(p_{k+1}(\Phi(y)) - p_k(\Phi(y)))
\]
converges, and clearly \( f \in \mathcal{TTP}_m^1 \) since each of the individual terms is. Furthermore, \( \|\cdot\|_2 \)-uniform continuity of each term and the uniform convergence of the series implies uniform continuity of \( f \). Since \( \|x\|_\infty \leq R \), we have \( \Phi(x) = x \) and \( u_k(x) = 1 \), so that
\[
f(x) = \sum_{k=0}^{\infty} [p_{k+1}(x) - p_k(x)] = \lim_{k \to \infty} p_{k+1}(x) = z.
\]
This concludes the proof of (1).

To verify (2), we take \( f_n \) to be the \( n \)th partial sum of the series defining \( f \); we have shown that the individual terms are \( \|\cdot\|_2 \)-Lipschitz, hence so are the partial sums. Finally, to prove (3), note that if \( z = p(x_1, \ldots, x_m) \), then \( z \) also equals \( f(z_1, \ldots, z_m) \) where \( f = p \circ \Phi \), and by the same reasoning as above \( p \circ \Phi \) is globally \( \|\cdot\|_2 \)-Lipschitz.

We have shown that every element of \( L^2(W^*(x_1, \ldots, x_m)) \) has the form \( f(x_1, \ldots, x_m) \) for some \( f \in \mathcal{TTP}_m \). On the other hand, we will prove that if \( f \) is Lipschitz, then \( f(x) \) is actually bounded in operator norm. We state our estimate in terms unitarily invariant random matrix models which satisfy concentration (2.5), but as explained in Remark 3.16 such models exist whenever \( L^2(W^*(x_1, \ldots, x_m)) \) is embeddable into \( \mathcal{R}^\omega \).

**Proposition 3.15.** Let \( x = (x_1, \ldots, x_m) \) be a tuple of self-adjoint variables in a \( W^* \)-algebra \((M, \tau)\) whose non-commutative law is \( \lambda \). Suppose there is a sequence \( \{\mu^{(N)}\} \) of probability measures on \( M_N(\mathbb{C})_{sa}^m \), invariant under unitary conjugation, that satisfies the concentration estimate (2.5) for some constant \( c \), and such that the corresponding random variables \( X^{(N)} = (X^{(N)}_1, \ldots, X^{(N)}_m) \) satisfy \( \lambda x^{(N)} \to \lambda \) in probability. Then \( W^*(x) \) is embeddable into \( \mathcal{R}^\omega \). Moreover, if \( f \in \mathcal{TTP}_m \) is \( \|\cdot\|_2 \)-Lipschitz, then \( f(x) \) is a bounded operator and
\[
\|f(x) - \tau(f(x))\|_\infty \leq \Theta c^{-1/2}\|f\|_{\text{Lip}},
\]
where \( \Theta \) is a universal constant.

**Proof.** In light of Lemma 2.7.2
\[
P \left( \|X^{(N)} - E(X^{(N)})\|_\infty \leq c^{-1/2}(\Theta + N^{-1/3}) \right) \to 1
\]
and
\[
P \left( \|f(X^{(N)}) - E(f(X^{(N)}))\|_\infty \leq c^{-1/2}\|f\|_{\text{Lip}}(\Theta + N^{-1/3}) \right) \to 1.
\]
Also, the non-commutative law of $X^{(N)}$ converges in probability to that of $x$ and finally $\tau_N(f(X^{(N)})) = E[\tau_N(f(X^{(N)}))] \to 0$ in probability as a consequence of concentration. Therefore, we may choose a sequence of elements $y^{(N)} \in M_N(\mathbb{C})_{sa}$ such that
\[
\limsup_{N \to \infty} \| y^{(N)}_j - E(y^{(N)}_j) \|_\infty \leq c^{-1/2}\Theta,
\]
\[
\limsup_{N \to \infty} \| f(y^{(N)}) - E(f(X^{(N)})) \|_\infty \leq c^{-1/2}\|f\|_{\text{Lip}}\Theta,
\]
\[
\|\tau_N(f(y^{(N)})) - E[\tau_N(f(X^{(N)}))]\|_\infty \to 0,
\]
\[
\lambda_{y^{(N)}} \to \lambda_x.
\]
Because $E(X^{(N)}_j) = E(\tau_N(X^{(N)}_j))$ by unitary invariance and because of concentration, $E(\tau_N(X^{(N)}_j))$ must converge to $\tau(x_j)$ because $\tau_N(X^{(N)}_j)$ converges to the $\tau(x_j)$ in probability. So overall $E(X^{(N)}_j) - \tau_N(x_j) \to 0$ in operator norm. In particular,
\[
\limsup_{N \to \infty} \| y^{(N)}_j - \tau(x_j) \|_\infty \leq c^{-1/2}\Theta,
\]
and hence $\|y^{(N)}\|_\infty$ is bounded as $N \to \infty$. Moreover, our choice of $y^{(N)}$ also satisfies
\[
\limsup_{N \to \infty} \| f(y^{(N)}_j) - \tau_N(f(y^{(N)})) \|_\infty \leq c^{-1/2}\|f\|_{\text{Lip}}\Theta,
\]
since $E[f(X^{(N)}_j)] = E[\tau_N(f(X^{(N)}_j))]$ again by unitary invariance.

Fix a free ultrafilter $\omega$ and let $(M, \tau) = \prod_{N \to \omega}(M_N(\mathbb{C}), \tau_N)$ be the tracial $W^*$-ultraproduct of the sequence of matrix algebras. Now $\tau(y^{(N)})$ is Since $\{y^{(N)}\}$ is bounded in operator norm, $y = \{y^{(N)}\}_{N \in \mathbb{N}}$ defines an element of $(M, \tau)$. By definition of ultraproducts, $\tau(p(y)) = \lim_{N \to \omega} \tau_N(p(y^{(N)}))$ for every non-commutative polynomial $p$ and therefore the non-commutative law of $y$ is $\lambda$ (which is the same as that of $x$). In particular, $W^*(x) \cong W^*(y)$ embeds into $(M, \tau)$ and hence also into $\mathcal{R}^\omega$. (Compare [13, Thm. 4.4].)

Since $W^*(x)$ is $\mathcal{R}^\omega$-embeddable, $f(x)$ is well-defined, and clearly $\|f(x) - \tau(f(x))\|_\infty = \|f(y) - \tau(f(y))\|_\infty$. Now we claim that $f(y)$ is given by the sequence $\{f(y^{(N)})\}_{N \in \mathbb{N}}$ as an element of $(M, \tau)$ (that is, application of $f$ commutes with ultralimits). It is easy to check that $g(y) = \{g(y^{(N)})\}_{N \in \mathbb{N}}$ when $g \in \text{TrP}_m^1$. But for any $\epsilon > 0$, there exists $g \in \text{TrP}_m$ with $\|g - f\|_{\text{lip}} < \epsilon$. Thus, $\|f(y) - g(y)\|_2 < \epsilon$ and also $\|f(y^{(N)}) - g(y^{(N)})\|_2 < \epsilon$ for sufficiently large $N$. This implies that $\|f(y) - \{f(y^{(N)})\}_{N \in \mathbb{N}}\|_2 < 2\epsilon$. Thus, $f(y) = \{f(y^{(N)})\}_{N \in \mathbb{N}}$ as claimed. The same holds with $f$ replaced by $f - \tau(f)$. This implies
\[
\|f(y) - \tau(f(y))\|_\infty \leq \limsup_{N \to \infty} \|f(y^{(N)}) - \tau(f(y^{(N)}))\|_\infty \leq c^{-1/2}\|f\|_{\text{Lip}}\Theta.
\]

**Remark 3.16.** Suppose that $W^*(x_1, \ldots, x_m)$ is embeddable into $\mathcal{R}^\omega$. Then there exist tuples $x^{(N)} = (x^{(N)}_1, \ldots, x^{(N)}_m)$ in $M_N(\mathbb{C})_{sa}$ such that $\|x^{(N)}\|_\infty \leq \|x\|_\infty$ and $\lambda_{x^{(N)}} \to \lambda_x$. Let $U^{(N)}$ be an $N \times N$ random Haar unitary matrix and let $X^{(N)} = U^{(N)}x^{(N)}(U^{(N)})^*$. Clearly, the probability distribution of $X^{(N)}$ is unitarily invariant and also $\lambda_{X^{(N)}} \to \lambda_x$ in probability.

To check concentration, observe that $u \mapsto u_{x^{(N)}}u^*$ is a $2m^{1/2}\|x\|_\infty$-Lipschitz function from the unitary group to $M_N(\mathbb{C})_{sa}$ with respect to $\|\|_2$. Therefore, if $f : M_N(\mathbb{C})_{sa} \to \mathbb{R}$ is Lipschitz, then $u \mapsto f(u_{x^{(N)}}u^*)$ is also Lipschitz, with the Lipschitz constant $2m^{1/2}\|x\|_\infty\|f\|_{\text{lip}}$. It was proved in [24, Theorem 15], [13] that the Haar measure on the unitary group satisfies the (non-normalized) log-Sobolev inequality with constant $6/N$ and the corresponding concentration of measure for Lipschitz functions with respect to the the Hilbert-Schmidt metric $N^{1/2}\|\|_2$. After
renormalization this implies that the Haar measure on the unitary group satisfies \((2.3)\) with \(c = 1/6\). Hence, \(X^{(N)}\) satisfies \((2.5)\) with \(c = 1/12m\|x\|_\infty^2\).

4. Tools for Differential Equations in \(\mathbb{T}^m\)

This section describes two analytic operations — solution of ODE and convolution with the Gaussian law — that can be performed on tuples in \(\mathbb{T}^m\) and on asymptotically approximable sequences of functions on \(N \times N\) matrices. These operations were applied in \([17, \text{Lemma 4.10}]\), and will be applied in the remainder of this paper, to analyze the large \(N\) limit of certain PDE associated to random matrix models, and hence to understand the behavior of convex matrix models in the large \(N\) limit.

4.1. Flows Along Vector Fields. Several times in our study of partial differential equations, we will use flows along vector fields given by functions in \(\mathbb{T}^m\) and by asymptotically approximable sequences of functions on matrices. For instance, this idea was used in \([17, \text{Lemma 4.10}]\), and in this paper, it will be used in the proof of Lemma \([5.13, \text{and Theorem 7.11}]\).

The setup is roughly speaking as follows. Consider a time interval \([0, T] \subseteq \mathbb{R}\). Let \(H : (\mathbb{R}^\omega)^m_{sa} \times [0, T] \to L^2(\mathbb{R}^\omega)^m_{sa}\) be a function such that \(H(\cdot, t)\) is a tuple of functions in \(\mathbb{T}^m\) for each \(t\) (satisfying certain uniform continuity assumptions). Also, let \(F_0 : (\mathbb{R}^\omega)^m_{sa} \to L^2(\mathbb{R}^\omega)^m_{sa}\) be a function such that

\[
F(x, 0) = F_0(x)\partial_t F(x, t) = H(F(x, t), t).
\]

Moreover, we would like to show that if \(H(\cdot)\) is a function on \(M_N(\mathbb{C})^m_{sa} \times [0, T]\) that is asymptotic to \(H\) and \(F_0(\cdot) \sim F_0\), then the solutions \(F(\cdot)\) are asymptotic to the solution \(F\).

Such a proof was essentially carried out in \([17, \text{Lemma 4.10}]\), but now we introduce the added complexity that \(H\) will depend on \(x, t\), and an auxiliary parameter \(y \in (\mathbb{R}^\omega)^m_{sa}\), and we must solve the initial value problem

\[
\begin{align*}
F(x, y, 0) &= F_0(x, y) \\
\partial_t F(x, y, t) &= H(F(x, y, t), y, t).
\end{align*}
\]

The added parameter \(y\) arises naturally in our analysis of conditional expectation, entropy, and transport since it represents the variables we are conditioning upon (see for instance \([4,5,3]\)).

For the sake of future reference, let us state the set of assumptions we make about the vector field \(H(x, y, t)\). These assumptions are framed for a convenient and applicable level of generality rather than maximum generality.

**Assumption 4.1.** We are given \(T > 0\) and a function \(H : (\mathbb{R}^\omega)^m_{sa} \times (\mathbb{R}^\omega)^m_{sa} \times [0, T] \to L^2(\mathbb{R}^\omega)^m_{sa}\) satisfying:

1. For each \(t\), we have \(H(\cdot, \cdot, t) \in (\mathbb{T}^m)_{sa}\).
2. \(H\) is \(\|\cdot\|_2\)-Lipschitz in \((x, y)\), that is, for some constant \(K\) independent of \(t\), we have

\[
\|H(x, y, t) - H(x', y', t)\|_2 \leq K\|x - x', y - y'\|_2.
\]
3. The map \(t \mapsto H(\cdot, \cdot, t)\) is a continuous function \([0, T] \to (\mathbb{T}^m)_{sa}\) with respect to the Fréchet topology on \(\mathbb{T}^m\). This implies that for every \(R > 0\) and for every \(\epsilon > 0\), there exists \(\delta > 0\), such that

\[
|t - t'| < \delta \implies \|H(\cdot, \cdot, t) - H(\cdot, \cdot, t')\|_{sa,R} < \epsilon \text{ for all } t, t' \in [0, T].
\]

(where we have upgraded from continuity to uniform continuity because of compactness of \([0, T]\)).
Observation 4.2. Under this assumption, as in Observation 3.8 we see that $H(\cdot, t)$ has a unique continuous extension to $L^2(R^\omega)_{sa+}$. Furthermore, for each $(x, y) \in L^2(R^\omega)_{sa+}$, the function $t \mapsto H(x, y, t)$ is continuous (though the modulus of continuity cannot be chosen independent of $(x, y)$). Continuity follows because there exists a sequence $(x_n, y_n) \in (R^\omega)_{sa+}$ such that $(x_n, y_n) \to (x, y)$ in $\|\cdot\|_2$. Now $H(x_n, y_n, \cdot)$ is continuous by assumption (3), but assumption (2) implies that $H(x_n, y_n, \cdot) \to H(x, y, \cdot)$ uniformly on $[0, T]$.

Under these assumptions, (1.1) can be solved by the standard method of Picard iteration. We first verify that Assumption 4.1 is preserved under the composition and integration operations used to define Picard iterates.

Lemma 4.3. Suppose that $H(x, y, t)$ satisfies Assumption 4.1 and suppose that $G_0 \in (\text{TrP}_m^1)_{sa+}$ is globally $\|\cdot\|_2$-Lipschitz. Then the function

$$G(x, y, t) = G_0(x, y) + \int_0^t H(x, y, s) \, ds$$

is well-defined by Riemann integration and it also satisfies Assumption 4.1.

Proof. The Riemann integral is defined because $t \mapsto H(x, y, t)$ is continuous with respect to $\|\cdot\|_2$ for each $(x, y) \in (R^\omega)_{sa+}$ (and in fact, each $(x, y) \in L^2(R^\omega)_{sa+}$). Now let us check that $G$ satisfies Assumption 4.1.

(1) Fix $R > 0$ and $\epsilon > 0$. By assumption (2) for $H$, there exists $\delta > 0$ such that

$$|t - t'| < \delta \implies \|H(\cdot, t) - H(\cdot, t')\|_{a, R} < \frac{\epsilon}{2T}.$$ 

Fix $t$, then choose a partition $0 = t_0, \ldots, t_n = t$ of $[0, t]$ such that $|t_j - t_{j-1}| < \delta$. Then let $h_j \in (\text{TrP}_m^1)_{sa+}$ such that

$$\|h_j - H(\cdot, t_j)\|_{a, R} < \frac{\epsilon}{2T}.$$ 

Then

$$\|h_j - H(\cdot, s)\|_{a, R} < \frac{\epsilon}{T}$$

for all $s \in [t_{j-1}, t_j]$.

Therefore,

$$\left\|\int_0^t H(\cdot, s) \, ds - \sum_{j=1}^n (t_j - t_{j-1}) h_j\right\|_{a, R} < \sum_{j=1}^n (t_j - t_{j-1}) \frac{\epsilon}{T} = \frac{\epsilon T}{T} \leq \epsilon.$$ 

This shows that $\int_0^t H(\cdot, s) \, ds$ is in $(\text{TrP}_m^1)_{sa+}$. Because $G_0$ is in this space as well, this implies that $G(\cdot, t)$ is in $(\text{TrP}_m^1)_{sa+}$ as desired.

(2) If $H(\cdot, t)$ is $K$-Lipschitz for all $t$, then $\|G(\cdot, t)\|_{\text{Lip}} \leq \|G_0\|_{\text{Lip}} + tK$.

(3) Since $t \mapsto H(\cdot, t)$ is continuous with respect to $\|\cdot\|_{a, R}$, we must have $\|H(\cdot, t)\|_{a, R} \leq M$ for some constant $M$. Then $\|G(\cdot, t) - G(\cdot, t')\|_{a, R} \leq M|t - t'|$. □

Lemma 4.4. Suppose that $H(x, y, t)$ and $G(x, y, t)$ satisfy Assumption 4.1. Then $H(G(x, y, t), y, t)$ also satisfies Assumption 4.1.

Proof. The composition makes sense because $H(x, y, t)$ extends to be defined for $(x, y) \in L^2(R^\omega)_{sa+}$. It follows from Lemma 3.10 that $H(G(x, y, t), y, t)$ satisfies (1). The Lipschitz estimate (2) is straightforward and left to the reader. To prove (3), let $K$ be a Lipschitz constant for $H$ as a function of $(x, y)$ that works for all $t$. Fix $\epsilon > 0$. Proceeding as in the proof of Lemma 4.3, we can choose a partition $\{t_0, \ldots, t_n\}$ of $[0, T]$ and $g_j \in (\text{TrP}_m^1)_{sa+}$ such that

$$\|g_j - G(\cdot, t)\|_{a, R} < \frac{\epsilon}{4K}$$

for $t \in [t_{j-1}, t_j]$. 

Then there exists some $R'$ such that $\|(x, y)\|_{\infty} \leq R$ implies $\|(g_j(x, y), y)\|_{\infty} \leq R'$ for all $j$. Then by applying assumption (3) to $H$, there exists $\delta$ such that

$$|t - t'| < \delta \implies \|H(\cdot, t) - H(\cdot, t')\|_{u,R} < \frac{\epsilon}{4}.\]$$

We also choose $\delta'$ such that

$$|t - t'| < \delta' \implies \|G(\cdot, t) - G(\cdot, t')\|_{u,R} < \frac{\epsilon}{4K}.\]$$

Supposing that $\|(x, y)\|_{\infty} \leq R$ and $|t - t'| < \min(\delta, \delta')$, we have

$$\|H(G(x, y, t), y, t) - H(G(x, y, t'), y, t')\|_2 \leq \|H(G(x, y, t), y, t) - H(G(x, y, t'), y, t')\|_2 + \|H(G(x, y, t'), y, t) - H(G(x, y, t'), y, t')\|_2 + \|H(G(x, y, t'), y, t) - H(G(x, y, t'), y, t')\|_2 \leq \frac{\epsilon}{4} + \|H(G(x, y), y, t) - H(G(x, y, t'), y, t')\|_2.\]$$

Meanwhile, after we pick $j$ such that $t' \in [t_j - 1, t_j]$, then

$$\|H(G(x, y, t'), y, t) - H(G(x, y, t'), y, t')\|_2 \leq \|H(G(x, y, t'), y, t) - H(g_j(x, y), y, t)\|_2 + \|H(g_j(x, y), y, t) - H(g_j(x, y), y, t')\|_2 + \|H(g_j(x, y), y, t) - H(G(x, y, t'), y, t')\|_2.$$}

The middle term can be estimated by $\epsilon/4$ because $\|(g_j(x, y), y)\|_{\infty} \leq R'$. Meanwhile, the first and third terms can each be estimated by $K(\epsilon/4K) = \epsilon/4$ using the Lipschitz property of $H$ and our choice of $g_j$. Altogether, $|t - t'| < \min(\delta, \delta')$ implies that $\|H(G(x, y, t), y, t) - H(G(x, y, t'), y, t')\|_2 < \epsilon$ whenever $\|(x, y)\|_{\infty} \leq R$. \hfill \qed

**Proposition 4.5.** Let $H(x, y, t)$ satisfy Assumption [4.1] and let $G_0 \in (\text{TriP}_{m+n})^m_{sa}$. Then there exists a unique continuous $F : L^2(\mathbb{R}^\omega)_{sa} \times [0, T] \rightarrow L^2(\mathbb{R}^\omega)_{sa}$ satisfying

$$F(x, y, t) = G_0(x, y) + \int_0^t H(F(x, y, s), y, s) \, ds.$$\]

Moreover, $F(x, y, t)$ also satisfies Assumption [4.1].

**Proof.** We define the Picard iterates $F_\ell$ inductively by

$$F_0(x, y, t) = G_0(x, y)$$

$$F_{\ell+1}(x, y, t) = G_0(x, y) + \int_0^t H(F_\ell(x, y, s), y, s) \, ds.$$\]

The previous two lemmas imply that $F_\ell$ is well-defined and satisfies Assumption [4.1]. Convergence of the Picard iterates follows from the standard proof of Picard-Lindelöf. Briefly, given that $H$ is $K$-Lipschitz in $(x, y)$ with respect to $\|\cdot\|_2$, we have

$$\|F_{\ell+1}(x, y, t) - F_\ell(x, y, t)\|_2 \leq K \int_0^t \|F_\ell(x, y, s) - F_{\ell-1}(x, y, s)\|_2 \, ds.$$\]

Also, we have

$$\|F_1(x, y, t) - F_0(x, y, t)\|_2 \leq t M(x, y),$$

where $M(x, y) = \sup_{s \in [0, T]} \|H(G_0(x, y), y, s)\|_2$, which is finite because of continuity of $H(G_0(x, y), y, t)$ in $t$. From here a straightforward induction on $\ell$ shows that

$$\|F_{\ell+1}(x, y, t) - F_\ell(x, y, t)\|_2 \leq \frac{K^{\ell-1} t^\ell}{\ell!}.$$
because $K \int_0^t K^{t-1}s^j \ell! \, ds = K^{t-1}s^j \ell!/\ell!$. Now because $\sum_{\ell=1}^\infty K^{t-1}s^j \ell!$ converges, we know that

$$F(x, y, t) := \lim_{\ell \to \infty} F_\ell(x, y, t)$$

exists, and

$$\|F_\ell(x, y, t) - F(x, y, t)\|_2 \leq M(x, y) \sum_{j=\ell+1}^\infty K^{j-1}s^j j!.$$  

The fact that $F(x, y, t)$ satisfies the integral equation is straightforward, and the proof of the uniqueness of this $F$ is also standard.

It remains to show that $F$ satisfies Assumption 4.6. First, recall that $H(G_0(x, y), y, t)$ is Lipschitz in $(x, y)$ uniformly for all $t$. If $K'$ is a Lipschitz constant for this function, then

$$M(x, y) \leq M(0, 0) + K'\|(x, y)\|_2.$$  

In particular,

$$\|F_\ell(x, y, t) - F(x, y, t)\|_2 \leq (M(0, 0) + K')\|(x, y)\|_2 \sum_{j=\ell+1}^\infty K^{j-1}s^j j!.$$  

This implies that the convergence of $F_\ell$ to $F$ occurs uniformly for $(x, y)$ with $\|(x, y)\|_\infty \leq R$ and all $t \in [0, T]$. Then because $F_\ell(\cdot, \cdot, t)$ can be approximated in $\|\cdot\|_{u, R}$ by trace polynomials, the same must be true for $F(\cdot, \cdot, t)$ for each $t$, which shows that $F$ satisfies (1). Similarly, because of the uniform convergence of $F_\ell$ to $F$ for $\|(x, y)\|_\infty \leq R$ and $t \in [0, T]$, the uniform continuity property (3) for $F$ follows from property (3) for $F_\ell$.

Finally, we must show (2) that $F$ is Lipschitz in $(x, y)$. More precisely, we claim that

$$\|F(x, y, t) - F(x, y, t')\|_2 \leq e^{K't} \|G_0(x, y) - G_0(x', y')\|_2 + (e^{K't} - 1)\|y - y'\|_2.$$  

Now it suffices to check that each Picard iterate $F_\ell$ satisfies this estimate. This can be verified by induction on $\ell$. The base case $F_0(x, y, t) = G_0(x, y)$ is immediate. For the induction step, we observe that

$$\|F_{\ell+1}(x, y, t) - F_\ell(x', y', t)\|_2 \leq \|G_0(x, y) - G_0(x', y')\|_2 + \int \|H(F_\ell(x, y, s), y, s) - H(F_\ell(x', y', s), y', s)\|_2 \, ds$$

$$\leq \|G_0(x, y) - G_0(x', y')\|_2 + \int K(\|F_\ell(x, y, s) - F_\ell(x', y', s)\|_2 + \|y - y'\|_2) \, ds,$$

using the fact that $H$ is $K$-Lipschitz. Then we plug in our induction hypothesis that $\|F_\ell(x, y, s) - F_\ell(x', y', s)\|_2$ is bounded by $e^{K't}\|G_0(x, y) - G_0(x', y')\|_2 + (e^{K't} - 1)\|y - y'\|_2$, and then directly evaluate the integral to close the induction. \hfill \square

We have now shown that it makes sense to solve ODE for tuples in $(\mathbb{T}^m_n)^n \times [0, T]$. There is a parallel list of results which instead deal with functions on $N \times N$ matrices that are asymptotically approximable as $N \to \infty$. We use the following assumptions.

**Assumption 4.6.** We are given $T > 0$ and for each $N \in \mathbb{N}$ a function $H^{(N)} : M_N(\mathbb{C})^n \times M_N(\mathbb{C})^n \times [0, T] \to M_N(\mathbb{C})^n$ such that

1. For each $t$, there exists $H(\cdot, \cdot, t) \in (\mathbb{T}_m+n)^n$ such that $H^{(N)}(\cdot, \cdot, t) \to H(\cdot, \cdot, t)$.
2. $H^{(N)}$ is $\|\cdot\|_\infty$-Lipschitz in $(x, y)$ with some Lipschitz constant $K$ independent of $t$ and $N$.
3. For every $\bar{R} > 0$ and for every $\epsilon > 0$, there exists $\delta > 0$, such that

$$|t - t'| < \delta \implies \left\|H^{(N)}(\cdot, \cdot, t) - H^{(N)}(\cdot, \cdot, t')\right\|_{u, R} < \epsilon \text{ for all } t, t' \in [0, T] \text{ for all } N.$$
Proposition 4.7. Let \( \{H^{(N)}\} \) satisfy Assumption 4.6, and let \( G^{(N)}_0 : M_N(\mathbb{C})_{sa}^{m+n} \to M_N(\mathbb{C})_{sa}^m \) be asymptotically approximable such that \( G^{(N)}_0 \rightsquigarrow G_0 \) and \( G_0^{(N)} \) is \( \|\cdot\|_2 \) Lipschitz uniformly in \( N \). Then for each \( N \) there is a unique \( F^{(N)} : M_N(\mathbb{C})_{sa}^{m+n} \times [0,T] \to M_N(\mathbb{C})_{sa}^m \) satisfying

\[
F^{(N)}(x,y,t) = G^{(N)}_0(x,y) + \int_0^t H^{(N)}(F^{(N)}(x,y,s),y,s) \, ds.
\]

Moreover, \( \{F^{(N)}\} \) also satisfies Assumption 4.4. Furthermore, the vector field \( H \) such that \( H^{(N)}(\cdot,\cdot,t) \rightsquigarrow H(\cdot,\cdot,t) \) satisfies Assumption 4.1 and we have \( F^{(N)} \rightsquigarrow F \) where \( F \) is the solution given by Proposition 4.5.

Proof. The proof of existence and uniqueness of the solution is almost identical to that of Proposition 4.5. First, one shows that Assumption 4.6 is preserved under integration and composition (analogous to Lemma 4.3 and 4.4). Then exactly as in the proof of Proposition 4.5, one defines Picard iterates, proves they converge, establishes Lipschitz bounds, and checks they satisfy Assumption 4.1. The one additional feature in these proofs is to make all the estimates uniform in \( N \). For instance, the quantity \( M(x,y) \) in the proof of Proposition 4.5 is replaced by

\[
M^{(N)}(x,y) = \sup_{s \in [0,T]} \left\| H^{(N)}(G^{(N)}_0(x,y),y,s) \right\|_2.
\]

Then \( H^{(N)}(G^{(N)}_0(x,y),y,t) \) has some Lipschitz constant \( K' \) independent of \( N \), and

\[
M^{(N)}(x,y) \leq M^{(N)}(0,0) + K'\|\|x,y\|_2.
\]

But then we can show that \( \sup_N M^{(N)}(0,0) \) is finite. This is because if \( \Phi^{(N)}(x,y,t) = H^{(N)}(G^{(N)}_0(x,y),y,t) \), then \( \sup_N \sup_x\|\Phi^{(N)}(\cdot,\cdot,t)\|_{u,R}^{(N)} \) is finite because of Assumption 4.4 (3) and the fact that \( \Phi^{(N)}(x,y,0) \) is asymptotically approximable and hence bounded in \( \|\| \|^{(N)} \) as \( N \to \infty \).

Now the fact that \( H \) satisfies Assumption 4.1 is a straightforward limiting argument. The key ingredient is that if \( f^{(N)} \rightsquigarrow f \), then \( \|f\|_{u,R} = \lim_{N \to \infty} \|f^{(N)}\|_{u,R}^{(N)} \).

Finally, to show that \( F^{(N)} \rightsquigarrow F \), it suffices to show that for each of the Picard iterates \( F^{(N)}_t \rightsquigarrow F_t \) because of the uniform convergence of \( F^{(N)}_t \to F^{(N)} \) as \( t \to \infty \) for \( \|\|x,y\|_\infty \leq R \), where the rate of convergence is also independent of \( N \). Furthermore, since the Picard iterates are defined inductively by composition and integration, it suffices to show that the asymptotic approximation relation \( \rightsquigarrow \) is preserved by these operations. Preservation under integration follows because the integrals can be approximated by Riemann sums and this approximation is uniformly good for \( \|\|x,y\|_\infty \leq R \) and for all \( N \) because of the uniform continuity Assumption 4.1 (3). Preservation under composition follows from Lemma 3.10. \( \square \)

4.2. The Heat Semigroup. Recall that the solution to the classical heat equation is given by convolution the heat kernel (which is given by a Gaussian probability density). In particular, let \( \sigma_{m,t}^{(N)} \) be the probability distribution of an \( m \)-tuple of independent GUE matrices \( (S_1^{(N)}, \ldots, S_m^{(N)}) \) such that \( \mathbb{E}[\tau_N[(S_j^{(N)})^2]] = t \), which is given by density \( (1/Z^{(N)})e^{-\|x\|^2/2t} \, dx \).

If \( u_0 : M_N(\mathbb{C})_{sa}^n \to \mathbb{C} \), then \( u_t := u_0 * \sigma_{m,t}^{(N)} \) solves the normalized heat equation

\[
\partial_t u_t = \frac{1}{2N} \Delta u_t.
\]

Here \( u_0 * \sigma_{m,t}^{(N)} \) is meant in the sense of convolving a function with a measure, and this is the same as convolving of \( u_0 \) with the density function for \( \sigma_{m,t}^{(N)} \). The meaning of \( \Delta \) is to be interpreted using coordinates with respect to some orthonormal basis of \( M_N(\mathbb{C})_{sa} \) in the inner product
Define \( \langle x, y \rangle = \text{Tr}(xy) \); this is not the same as differentiating entrywise since some of the entries are real and some are complex.

Our goal is to describe the large \( N \) behavior of \( u^{(N)} \ast \sigma_{m,t}^{(N)} \) when \( \{u^{(N)}\} \) is asymptotically approximable by trace polynomials, and to define \( u \oplus \sigma_{m,t} \) when \( u \in \text{TrP}_m \).

In [17, §3.2 and 3.3], we explained the computation of \( (1/N)\Delta f \) as a function on \( M_N(\mathbb{C})_{sa}^m \) when \( f \in \text{TrP}_m^0 \) or \( \text{TrP}_m^1 \). More precisely, let \( \Delta_j f(x_1, \ldots, x_m) \) denote the Laplacian with respect to the coordinates of the matrix \( x_j \). We found that for \( j = 1, \ldots, m \) there are linear maps \( L_j^{(N)}, L_j : \text{TrP}_m^0 \rightarrow \text{TrP}_m^0 \) defined purely algebraically, such that \( (1/N)\Delta_j f = L_j^{(N)} f \) when \( f \) is viewed as a function on \( M_N(\mathbb{C})_{sa}^m \), \( L_j^{(N)} \) and \( L_j \) do not increase the degree of a trace polynomial, and \( \lim_{N \rightarrow \infty} L_j^{(N)} f = L_j f \) coefficient-wise.

A similar analysis holds for the Laplacian of \( f \in \text{TrP}_m^1 \) viewed as a function \( M_N(\mathbb{C})_{sa}^m \rightarrow M_N(\mathbb{C}) \). Here we follow the standard convention of using the same symbol \( \Delta \) for the Laplacians of vector-valued functions as for the Laplacians of scalar-valued functions; thus, the reader must be careful to distinguish scalar-valued and vector-valued functions based on context. We saw that there were linear transformations \( L_j^{(N)}, L_j : \text{TrP}_m^1 \rightarrow \text{TrP}_m^1 \) such that \( (1/N)\Delta_j f = L_j^{(N)} f \) as a function on matrices, \( L_j^{(N)} \) and \( L_j \) do not increase degree, and \( L_j^{(N)} f \rightarrow L_j f \) coefficient-wise.

We deduced as a consequence that \( e^{t(L_j^{(N)})/2} f = f \ast \sigma_{m,t}^{(N)} \) has a well-defined large \( N \) limit if \( f \) is a trace polynomial [17, Lemma 3.21], and that if \( \{u^{(N)}\} \) is asymptotically approximable by trace polynomials, then so is \( \{u^{(N)} \ast \sigma_{m,t}^{(N)}\} \) [17, Lemma 3.28].

In order to establish “conditional versions” of our earlier results, we must consider trace polynomials \( f(x_1, \ldots, x_m, y_1, \ldots, y_n) \) in \( m + n \) variables and take the Laplacian with respect to \( x = (x_1, \ldots, x_m) \) while treating \( y = (y_1, \ldots, y_n) \) as an auxiliary parameter. We denote by \( \Delta_x = \sum_{j=1}^m \Delta x_j, \Delta_y = \sum_{j=1}^n \Delta y_j \), and \( \Delta_x = \sum_{j=1}^m \Delta x_j, \Delta_y = \sum_{j=1}^n \Delta y_j \) the various Laplacian operators with respect to \( x \).

Because \( L_x^{(N)} \) and \( L_x \) map the finite-dimensional vector space trace polynomials of degree \( \leq d \) into itself, there are well-defined linear operators \( e^{tL_x^{(N)}/2} \) and \( e^{tL_x/2} \) on the space of trace polynomials in \( \text{TrP}_m^j \) of degree \( \leq d \) for each \( j = 0, 1 \) each \( d \in \mathbb{N} \), and each real \( t \geq 0 \). Since trace polynomials are the union of the subspaces of trace polynomials with degree \( \leq d \), there are linear operators \( e^{tL_x^{(N)}/2}, e^{tL_x/2} : \text{TrP}_m^j \rightarrow \text{TrP}_m^j \). Moreover, these operators form a semigroup.

**Lemma 4.8.** Let \( (X, Y) \) be a random variable in \( M_N(\mathbb{C})_{sa}^{m+n} \) with finite moments, and let \( S \sim \sigma_{m,t}^{(N)} \) be an independent GUE random variable in \( M_N(\mathbb{C})_{sa}^m \). Then we have

\[
E[f(X + S, Y)| (X, Y)] = [e^{tL_x^{(N)}/2} f](X, Y) \quad \text{for } f \in \text{TrP}_m^0 \text{ or } f \in \text{TrP}_m^1.
\]

Similarly, suppose that \( (X, Y) \) is a tuple of self-adjoint non-commutative random variables, and let \( S \) be a freely independent tuple with non-commutative law \( \sigma_{m,t} \). Then

\[
f(X + S, Y) = [e^{tL_x/2} f](X, Y) \quad \text{for } f \in \text{TrP}_m^0,
\]

and

\[
E_{W^*(X,Y)} [f(X + S, Y)] = [e^{tL_x/2} f](X, Y) \quad \text{for } f \in \text{TrP}_m^1,
\]

where \( E_{W^*(X,Y)} : W^*(X, Y, S) \rightarrow W^*(X, Y) \) is the unique trace-preserving conditional expectation.
Proof. Since $S$ is independent and distributed according to $\sigma_{m,t}^{(N)}$, we have
\[ E[f(X + S, Y)|(X, Y)] = \int f(X + z, Y) d\sigma_{m,t}^{(N)}(z). \]

On the other hand, for $(x, y) \in M_N(C)_{\mathbb{R}^+}^{m+n}$,
\[ \int f(x + z, y) d\sigma_{m,t}^{(N)}(z) = e^{tL_x/2}f(x, y), \]

because both sides are the solution to the heat equation on the space of coordinate-wise polynomials on $M_N(C)_{\mathbb{R}^+}^{m+n}$ of degree $\le d$. This shows (4.2).

To prove the free versions, we assume familiarity with the results of free probability (see e.g. [31, 23, 1, Chapter 5]). Suppose that $(X, Y)$ are non-commutative random variables and $S_t$ is a freely independent free semicircular $m$-tuple with law $\sigma_{m,t}^{(N)}$. We may assume that $(S_t)_{t \ge 0}$ is a free Brownian motion, so that $S_t - S_s \sim S_{t-s}$ for $0 \le s \le t$ and $S_t \sim t^{1/2}S_1$. Note that $e^{-tL_x/2}$ is a well-defined operator on trace polynomials. To prove (4.3), it suffices to show that $[e^{tL_x/2}f](X + S_t, Y) = f(X, Y)$ for $f \in \text{TrP}_m^1$. This will follow if we check that
\[ \frac{d}{dt} \left( e^{-tL_x/2}f(X + S_t, Y) \right) = 0. \]

From a free probabilistic computation sketched in [17, Lemma 3.23], we have
\[ \frac{d}{dt} f(X + S_t, Y) = \frac{1}{2} (L_x f)(X + S_t, Y), \]

and hence
\[ \frac{d}{dt} \left( e^{-tL_x/2}f(X + S_t, Y) \right) = \frac{d}{dt_1} \left[ e^{-t_1L_x/2} f \right](X + S_{t_2}, Y) \bigg|_{t_1 = t_2 = t} + \frac{d}{dt_2} \left[ e^{-t_1L_x/2} f \right](X + S_{t_2}, Y) \bigg|_{t_1 = t_2 = t} = 0. \]

Next, to prove (4.4), it suffices to show that for $g \in \text{TrP}_m^1$, we have
\[ \tau(f(X + S_t, Y)g(X, Y)) = \tau([e^{tL_x/2}f](X, Y)g(X, Y)), \]

since functions of the form $g(X, Y)$ for $g \in \text{TrP}_m^1$ are dense in $L^2(W^*(X, Y))$. Consider the function $F \in \text{TrF}_{m+n+m}$ given by $F(x, y, x') = \tau(f(x, y)g(x', y))$. Notice that
\[ L_x F(x, y, x') = \tau((L_x f)(x, y)g(x', y)) = \tau([L_x f](x, y)g(x', y)). \]

Here the first equality is checked directly from the definition of the Laplacian [17, see Def. 3.13 and 3.16, proof of Lemma 3.18]. The equality $L_x f(x, y)g(x', y) = L_x [f(x, y)g(x', y)]$ again is checked from the definition of the Laplacian; this equality is intuitive since $g(x', y)$ is independent of $x$. Since the same reasoning may be applied to compute the Laplacian $L_x$ of \( \tau([e^{tL_x/2}f](x, y)g(x, y)) \), we have
\[ e^{tL_x/2}F(x, y, x') = \tau([e^{tL_x/2}f](x, y)g(x', y)). \]

We can view $F(x, y, x')$ as a function of the $m$-tuple $x$ and the $(n + m)$-tuple $(y, x')$, that is, an element of $\text{TrP}_m^1 \times \text{TrP}_{n+m}^1$. We apply (4.3) to $f$ and the pair $(X, (Y, X))$ and obtain
\[ F(X + S_t, Y, X) = e^{tL_x/2}F(X, Y, X) \]

which means precisely that
\[ \tau(f(X + S_t, Y)g(X, Y)) = \tau([e^{tL_x/2}f](X, Y)g(X, Y)), \]

which completes the proof of (4.4). \( \square \)
Lemma 4.12. Moreover, we denote by \( P \) polynomials and let \((X,Y)\) be (deterministic) tuples of matrices with non-commutative laws converging to the law of \((X,Y)\) and let \(S(N) \sim \sigma_m(N)\). Then to prove (4.3), for instance, we could use the fact that \(E[f(X^N) + S(N), Y^N) = [e^{\frac{1}{2} f}(X^N)]Y^N)\) and take the limit as \(N \to \infty\) using Voiculescu’s theorem on asymptotic freeness [31, Thm. 2.2]. A similar proof could be done for (4.4).

Remark 4.13. The free conditional expectation formulas (4.3) and (4.4) could also be proved using random matrices provided that \(W^\omega(X,Y)\) is \(\mathcal{R}^\omega\)-embeddable. Indeed, let \((X^N), Y^N)\) be (deterministic) tuples of matrices with non-commutative laws converging to the law of \((X,Y)\) and set \(S(N) \sim \sigma_m(N)\). Then to prove (4.3), for instance, we could use the fact that \(E[f(X^N) + S(N), Y^N) = [e^{\frac{1}{2} f}(X^N)]Y^N)\) and take the limit as \(N \to \infty\) using Voiculescu’s theorem on asymptotic freeness [31, Thm. 2.2]. A similar proof could be done for (4.4).

Lemma 4.10. If \(f \in \text{TrP}^j_{m+n}\) for \(j = 0, 1\), then we have \(\|e^{tL_{L^2}} f\|_{u,R} \leq \|f\|_{u,R+2t^{1/2}}\) for \(t \geq 0\). In particular, \(f \mapsto e^{tL_{L^2}} f\) extends to a unique continuous linear operator \(\text{TrP}^j_{m+n} \to \text{TrP}^j_{m+n}\).

Proof. Let \((X,Y) \in (\mathcal{R}^\omega)_{sa}^{m+n}\) with \(\|(X,Y)\|_\infty \leq R\). Let \(S \sim \sigma_{m,t}\) be a freely independent semicircular tuple. If \(f \in \text{TrP}^0_{m+n}\), then
\[e^{tL_{L^2}} f(X,Y) = f(X + S, Y)\]
Since \(\|S\|_\infty = 2t^{1/2}\), we have \(\|(X + S, Y)\|_\infty \leq R + 2t^{1/2}\). Therefore, \(\|f\|_{u,R} \leq \|f\|_{u,R+2t^{1/2}}\) as desired. Similarly, if \(f \in \text{TrP}^1_{m+n}\), then we check \(\|f\|_{u,R} \leq \|f\|_{u,R+2t^{1/2}}\) using the conditional expectation formula (4.3). Now the continuous extension of \(e^{tL_{L^2}}\) to \(\text{TrP}^j_{m+n}\) is immediate. \(\Box\)

The semigroup \(e^{tL_{L^2}}\) acting on \(\text{TrP}^j_{m+n}\) describes the large \(N\) limit of the Gaussian convolution semigroup on \(M_N(\mathbb{C})_{sa}\), defined as follows.

Definition 4.11. For \(f : M_N(\mathbb{C})_{sa}^{m+n} \to \mathbb{C}\) or \(M_N(\mathbb{C})\), we denote
\[P_t^{(N)} f(x,y) = \int f(x + z, y) \, d\sigma_m^{(N)}(z)\]
Moreover, we denote by \(P_t^{\text{TrP}} : \text{TrP}^j_{m} \to \text{TrP}^j_{m}\) the continuous extension of \(e^{tL_{L^2}}\).

Lemma 4.12. Suppose that \(f^{(N)} : M_N(\mathbb{C})_{sa}^{m+n} \to \mathbb{C}\) is asymptotically approximable by trace polynomials and \(f^{(N)} \to f \in \text{TrP}^j_{m}\). Furthermore, assume that for some \(A, B > 0\) and \(k \in \mathbb{N}\), we have
\[\|f\|_{u,R} \leq A + BR^k \sup_{x \in M_N(\mathbb{C})_{sa}^{m+n}} \|f^{(N)}(x)\| \leq A + BR^k\]
Then \(P_t^{(N)} f^{(N)} \to P_t^{\text{TrP}} f\). The same holds for \(f^{(N)} : M_N(\mathbb{C})_{sa}^{m+n} \to M_N(\mathbb{C})\) and \(f \in \text{TrP}^1_{m}\) with \(|f^{(N)}(x)|\) replaced by \(\|f^{(N)}(x)\|_2\).

The proof of this lemma is the same as in [17, Lemma 3.28].

Remark 4.13. In both the scalar-valued and matrix-valued cases, the assumption (4.5) holds automatically with \(k = 1\) provided that \(f^{(N)}\) and \(f\) are \(\|\cdot\|_2\)-uniformly continuous (with modulus of continuity independent of \(N\)). Let us focus on the matrix-valued case of \(\text{TrP}^1_{m}\), there exists \(\delta > 0\) such that
\[\|x - y\|_2 < \delta \implies \|f(x) - f(y)\|_2 \leq 1\]
In particular, given \(x \in (\mathcal{R}^\omega)_{sa}^m\), we can choose an integer \(j\) such that \(j \delta < \|x\|_2 \leq 2j\delta\). Then we have
\[\|f(x) - f(0)\|_2 \leq \sum_{i=1}^{2j} \|f(ix/2k) - f((i - 1)x/2k)\|_2 \leq 2j \leq 2\|x\|_2/\delta\]
Thus,
\[\|f(x)\|_2 \leq \|f(0)\|_2 + \frac{2}{\delta} \|x\|_2 \leq \|f(0)\|_2 + \frac{2m^{1/2}}{\delta} \|x\|_\infty\]
which implies the first estimate of (4.3). The case for $f^{(N)}$ is handled similarly, and we note that $\|f^{(N)}(0)\|_2$ is bounded as $N \to \infty$ because of our assumption that $f^{(N)} \sim f$. The same argument works in the case of scalar-valued functions and $f \in \mathbb{TP}_m$.

5. Conditional Expectation for Free Gibbs States

5.1. Free Gibbs States from Convex Potentials. In [17] and in the present work, we focus on the following situation:

**Assumption 5.1.** We are given $0 < c \leq C$ and $V^{(N)} : M_N(\mathbb{C})^m_\text{sa} \to \mathbb{R}$ such that

1. $HV^{(N)} \geq c$, that is, $V^{(N)}(x) - \frac{1}{2}c\|x\|_2^2$ is convex.
2. $HV^{(N)} \leq C$, that is, $V^{(N)}(x) - \frac{1}{2}C\|x\|_2^2$ is concave.
3. $\{DV^{(N)}\}_{N \in \mathbb{N}}$ is asymptotically approximable by trace polynomials.

We denote by $\mu^{(N)}$ the probability measure on $M_N(\mathbb{C})^m_\text{sa}$ given by

$$d\mu^{(N)}(x) = e^{-N^2V^{(N)}(x)}dx.$$

Furthermore, we assume that the mean $\int x_j d\mu^{(N)}(x)$ is a scalar multiple of the identity matrix.

The following was proved in [17, Theorem 4.1].

**Theorem 5.2.** Let $V^{(N)}$ and $\mu^{(N)}$ be as in Assumption 5.1. Then there exists a non-commutative law $\lambda$ such that for every non-commutative polynomial $p$, we have

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

Moreover, we have for every $R > 0$ and $\epsilon > 0$ that

$$\lim_{N \to \infty} \frac{1}{N^2} \log \mu_N(\{x \in M_N(\mathbb{C})^m_\text{sa} : \|x\|_\infty \leq R, |\tau(p(x)) - \lambda(p)| > \epsilon\}) < 0.$$

**Corollary 5.3.** Let $\mu^{(N)}$ and $\lambda$ be as in Theorem 5.2. Let $X^{(N)}$ be a random $m$-tuple of matrices distributed according to $\mu^{(N)}$ and let $X$ be a non-commutative random $m$-tuple distributed according to $\lambda$. Let $f^{(N)}, g^{(N)} : M_N(\mathbb{C})^m_\text{sa} \to M_N(\mathbb{C})$. Suppose there are constants $A$ and $B > 0$ and $k \in \mathbb{N}$ such that

$$\max(\|f^{(N)}(x)\|_2, \|g^{(N)}(x)\|_2) \leq A + B\|x\|^k$$

Suppose that $f^{(N)} \sim f$ and $g^{(N)} \sim g$ where $f, g \in \mathbb{TP}_m$. Then

$$\lim_{N \to \infty} E[\tau_N(f^{(N)}(X^{(N)})g^{(N)}(X^{(N)}))] = \tau(f(X)g(X)).$$

**Proof.** Let $a_j^{(N)} = E[X_j^{(N)}]$ which we assumed to be a scalar multiple of the identity, and which we know has a limit as $N \to \infty$. By Lemma 2.2, we have

$$P(\|X_j^{(N)} - a_j^{(N)}\|_\infty \geq c^{-1/2}(\Theta + \delta) \leq e^{-cN\delta^2/2}. $$

In particular, letting $R > \sup_{N,j} |a_j^{(N)}| + c^{-1/2}\Theta$, we have

$$P(\|X^{(N)}\|_\infty \geq R) \to 0$$

and

$$E[1_{\|X^{(N)}\|_\infty \geq R}\tau_N(f^{(N)}(X^{(N)})g^{(N)}(X^{(N)}))] \to 0.$$

Therefore, in order to prove convergence of the expectation, it suffices to check that $\tau_N(f^{(N)}(X^{(N)})g^{(N)}(X^{(N)}))$ converges in probability to $\tau(f(X)g(X))$. 


We already know that $\tau_N(p(X^{(N)}))$ converges to $\tau(p(X))$ in probability for every non-commutative polynomial $p$. It follows that if $u$ is a scalar-valued trace polynomial, then $u(X^{(N)}) \to u(X)$ in probability. This also holds for $u \in \text{TrP}_m^0$; indeed, we know that $\|X^{(N)}\|_\infty \leq R$ with probability tending to 1 and $\|X\|_\infty \leq R$, whereas $u$ can be approximated in $\|\cdot\|_{u,R}$ by trace polynomials. Finally, if $u^{(N)}$ is a sequence of scalar-valued function such that $u^{(N)} \sim u \in \text{TrP}_m^0$, then $u^{(N)}(X^{(N)}) - u(X^{(N)})$ converges to 0 in probability, and hence $u^{(N)}(X^{(N)})$ converges in probability to $u(X)$. By Lemma 3.7 we can apply this statement to $u^{(N)} = \tau_N(f^{(N)}g^{(N)})$ and $u = \tau(fg)$, which completes the argument.

**Definition 5.4.** Let $V \in \text{TrP}_m^0$ and suppose $V$ extends to a function $L^2(\mathcal{R}^\omega)^m_\omega \to \mathbb{R}$ such that $V(x) - (c/2)||x||_2^2$ is convex and $V(x) - (C/2)||x||_2^2$ is concave. In this case, $V$ is differentiable as a function on the real Hilbert space $L^2(\mathcal{R}^\omega)^m_\omega$, as a consequence of the existence of supporting hyperplanes for convex functions on a Hilbert space. If we assume also that $DV \in \text{TrP}_m$, then we say that $V \in \mathcal{E}_m^\text{TrP}(c,C)$.

**Remark 5.5.** We did not prove or assume that the trace polynomials which approximate $DV$ are the gradients of the same trace polynomials that approximate $V$. Thus, this definition is technically different from that of [17, §8.2].

**Definition 5.6.** If $V \in \mathcal{E}_m^\text{TrP}(c,C)$, then we may define $V^{(N)} = V|_{M_N(\mathbb{C})^m_\omega}$, and in this case $DV^{(N)} = DV|_{M_N(\mathbb{C})^m_\omega}$. Clearly, $DV^{(N)}$ is asymptotically approximable by trace polynomials, and so by Theorem 6.2 there exists a non-commutative law $\lambda_V$ that arises as the large $N$ limit of the associated random matrix models. Furthermore, the limiting free Gibbs law $\lambda_V$ only depends on $V$, that is, every approximating sequence of functions $V^{(N)} \in \mathcal{E}_m^{(N)}(c,C)$ will produce the same free Gibbs law (see [17, §8.2]). We call $\lambda_V$ the free Gibbs state given by potential $V$.

**Remark 5.7.** One can check that if $V^{(N)}$ is as in Assumption 5.1 then there exists a $V \in \mathcal{E}_m^\text{TrP}(c,C)$ such that $V^{(N)} \sim V$ and $DV^{(N)} \Rightarrow DV$. Thus, the non-commutative laws that arise from these random matrix models are precisely $\lambda_V$ for $V \in \mathcal{E}_m^\text{TrP}(c,C)$.

**Remark 5.8.** Since $\lambda_V$ is independent of the choice of approximating sequence $V^{(N)}$, we can in particular take $V^{(N)} = V|_{M_N(\mathbb{C})^m_\omega}$, which produces a canonical unitarily invariant sequence of random matrices models.

### 5.2 Main Result on Conditional Expectation.

Our main result in this section is in some sense a generalization of [17, Theorem 4.1], which deals with conditional expectations rather than expectations. The proof of the earlier theorem was reduced to the following statement: Suppose $V^{(N)}$ satisfies Assumption 5.1 and that $u^{(N)} : M_N(\mathbb{C})^m_\omega \to \mathbb{C}$ is $\|\cdot\|_2$-Lipschitz (uniformly in $N$) and asymptotically approximable by trace polynomials. Then

$$\lim_{N \to \infty} \int u^{(N)} \, d\mu^{(N)}$$

exists.

Now, our goal is to prove the following.

**Theorem 5.9.** Consider functions $V^{(N)} : M_N(\mathbb{C})^m_\omega \to \mathbb{R}$, denoted as $V^{(N)}(x,y)$, which satisfy Assumption 5.1 as functions of $(x,y)$. Let $\mu^{(N)}$ be the associated probability measure on $M_N(\mathbb{C})^m_\omega$. Let $(X^{(N)}, Y^{(N)})$ be an $(m + n)$-tuple of random matrices distributed according to $\mu^{(N)}$, and let $(X,Y)$ be a $(m + n)$-tuple of non-commutative random variables distributed according to the limiting free Gibbs law $\lambda$ given by Theorem 5.2.
Let \( f^{(N)} : M_N(\mathbb{C})^{m+n} \rightarrow M_N(\mathbb{C}) \) be \( \| \cdot \|_2 \)-Lipschitz (uniformly in \( N \)) and suppose \( f^{(N)} \rightharpoonup f \in \mathbb{TP}^m_{m+n} \). Let \( g^{(N)} \) be the function given by
\[
g^{(N)}(Y^{(N)}) = E[f^{(N)}(X^{(N)}, Y^{(N)})|Y^{(N)}],
\]
which is well-defined function \( M_N(\mathbb{C})^{m+n} \) because \( \mu^{(N)} \) has positive density everywhere. Then \( g^{(N)} \) is Lipschitz with
\[
\|g^{(N)}\|_{\text{Lip}} \leq (1 + C/c) \|f^{(N)}\|_{\text{Lip}}.
\]
Moreover, there exists \( g \in \mathbb{TP}^m \) such that \( g^{(N)} \rightharpoonup g \) and hence
\[
g(Y) = E_{W^*(Y)}[f(X,Y)].
\]

The gist of the theorem is that the conditional expectation \( E[\cdot|Y^{(N)}] \) behaves in the large \( N \) limit like the \( W^* \)-algebraic expectation \( W^*(X,Y) \rightarrow W^*(Y) \). For instance, if \( f \in \mathbb{TP}^m_{m+n} \) is globally Lipschitz in \( \| \cdot \|_2 \), then the \( W^* \)-algebraic conditional expectation of \( f(X,Y) \) can be approximated by the classical conditional expectation \( E[f(X^{(N)}, Y^{(N)})|Y^{(N)}] \).

In fact, we can approximate \( E_{W^*}(Z) \) for every \( Z \in L^2(W^*(X,Y)) \) using classical conditional expectations in the same sense. Indeed, we showed in Proposition 3.12 that every \( Z \) can be expressed as \( f(X,Y) \) where \( f \in \mathbb{TP}^m \) is \( \| \cdot \|_2 \)-uniformly continuous, and there exist \( \| \cdot \|_2 \)-Lipschitz functions \( f_k \in \mathbb{TP}^m \) such that \( f_k \rightarrow f \) with respect to the uniform norm \( \| \cdot \|_u \). Let \( g_k^{(N)} \) and \( g^{(N)} \) be given by
\[
g_k^{(N)}(Y^{(N)}) = E[f_k(X^{(N)}, Y^{(N)})|Y^{(N)}],
\]
and the analogous relation for \( g^{(N)} \) and \( f \). Because conditional expectation is a contraction in \( L^\infty(\mu^{(N)}) \) (for functions taking values in \( M_N(\mathbb{C}) \) with \( \| \cdot \|_2 \)), we have
\[
\|g_k^{(N)} - g^{(N)}\|_{\text{Lip}} \leq \|f_k - f\|_u.
\]

By the theorem, there exists \( g_k \in \mathbb{TP}^m \) such that \( g_k^{(N)} \rightharpoonup g_k \). Given that \( \|g_k^{(N)} - g^{(N)}\|_{\text{Lip}} \leq \|f_k - f\|_u \rightarrow 0 \), a routine argument ("exchange of limits and uniform limits") shows that there exists \( g \in \mathbb{TP}^m \) such that \( g^{(N)} \rightharpoonup g \). In other words, the conclusion of Theorem 5.9 holds also for \( f \) and thus \( E_{W^*}(Y)|Z = E_{W^*}(Y)[f(X,Y)] \) can be viewed as the large \( N \) limit of \( E[f(X^{(N)}, Y^{(N)})|Y^{(N)}] \).

5.3. Strategy. Our proof will follow the same strategy as the special case in [17, §4]. In that paper, we showed that if \( V^{(N)} \) and \( \mu^{(N)} \) on \( M_N(\mathbb{C})^{m+n} \) are as in Assumption 5.1 and if \( u^{(N)} : M_N(\mathbb{C})^{m+n} \rightarrow \mathbb{C} \) is uniformly Lipschitz and asymptotically approximateable by trace polynomials, then \( \lim_{N \rightarrow \infty} \int u^{(N)} \, d\mu^{(N)} \) exists.

We considered the diffusion semigroup \( T^V_t \) that solves the equation
\[
\partial_t(T^V_t u^{(N)}) = \frac{1}{2N} \Delta(T^V_t u^{(N)}) - \frac{1}{2}(DV^{(N)}, D(T^V_t u^{(N)}))_2.
\]
As mentioned in [17, §4], this diffusion semigroup has an equivalent SDE formulation, and is a standard tool in proving the log-Sobolev inequality and concentration estimates (see for instance, [18, §4.4.2], [11]).

Now \( \int T^V_t u^{(N)} \, d\mu^{(N)} = \int u^{(N)} \, d\mu^{(N)} \) and \( \| T^V_t u^{(N)} \|_{\text{Lip}} \leq e^{-ct/2} \| u^{(N)} \|_{\text{Lip}} \). As \( t \rightarrow \infty \), the function \( T^V_t u^{(N)} \) converges to the constant function \( \int u^{(N)} \, d\mu^{(N)} \) at a rate independent of \( N \). On the other hand, we showed in [17, Lemma 4.10] that if \( \{u^{(N)}\}_{N \in \mathbb{N}} \) and \( \{DV^{(N)}\}_{N \in \mathbb{N}} \) are
asymptotically approximable by trace polynomials, then so is \( \{ T^{(N)}_t u^{(N)} \} \) for all \( t \geq 0 \). Hence, we concluded that the sequence of constant functions \( \{ f u^{(N)} d\mu^{(N)} \} \) is asymptotically approximable by trace polynomials, which means that the limit as \( N \to \infty \) exists.

Now we apply the same method in the conditional setting to prove Theorem 5.9. Let \( V^{(N)}(x,y) \) be a function satisfying Assumption 5.1. If we fix \( y \), then \( V^{(N)}(\cdot,y) \) is uniformly convex and semi-concave function of \( x \), so it defines a log-concave probability measure on \( M_N(\mathbb{C})^{n \times n}_{s_a} \). This produces a well-behaved conditional distribution of \( X^{(N)} \) given \( Y^{(N)} \), where \( (X^{(N)}, Y^{(N)}) \sim \mu^{(N)} \). Explicitly, for \( f \in L^1(\mu^{(N)}, M_N(\mathbb{C})) \), we have

\[
E[f(X^{(N)}, Y^{(N)}) | Y^{(N)}] = \frac{\int f(x, Y^{(N)}) e^{-N^2V^{(N)}(x, Y^{(N)})} \, dx}{\int e^{-N^2V^{(N)}(x, Y^{(N)})} \, dx}.
\]

We will evaluate this conditional expectation as the limit as \( t \to \infty \) of \( T^{(N)}_t f \), where \( T^{(N)}_t = T^{(N)}_t \) is the semigroup, acting on Lipschitz functions of \( (x, y) \), that solves

\[
\partial_t (T^{(N)}_t f) = \frac{1}{2N} \Delta_x (T^{(N)}_t f) - \frac{1}{2} J_x (T^{(N)}_t f)^* D_x V^{(N)},
\]

where \( J_x (T^{(N)}_t f) \) denotes the differential (Jacobian) of \( T^{(N)}_t f \) as a function \( x \) from \( M_N(\mathbb{C})^{n \times n}_{s_a} \) to \( M_N(\mathbb{C}) \) and \( * \) denotes the adjoint. In [5.4], we will analyze how \( T^{(N)}_t f \) affects the Lipschitz norms with respect to \( x \) and \( y \) separately and hence show that the conditional expectation is given by a Lipschitz function of \( y \). In [5.5] we will show that \( T^{(N)}_t \) preserves asymptotic approximability by trace polynomials of \( (x, y) \) and conclude our argument. The new aspect compared to [17] is that the functions are matrix-valued and depend on an extra parameter \( y \).

5.4. Conditional Diffusion Semigroup. To simplify notation, let us fix \( N \) and fix \( V : M_N(\mathbb{C})^{n \times n}_{s_a} \times M_N(\mathbb{C})^{n \times n}_{s_a} \to \mathbb{R} \) for the remainder of [5.4]. We will denote

\[
d\mu(x,y) = \frac{1}{\int e^{-N^2V(x,y)} \, dx} e^{-N^2V(x,y)} \, dx,
\]

which is a measure on \( M_N(\mathbb{C})^{n \times n}_{s_a} \) depending on the parameter \( y \). The associated semigroup \( T_t \) will be approximated by alternating two other operators \( P_t \) and \( S_t \) on short time intervals. Let \( P_t \) denote the semigroup of convolution with Gaussian with respect to \( x \), that is,

\[
P_t f(x, y) = \int f(x + z, y) \, d\mu(z).
\]

The semigroup \( S_t \) is given by

\[
S_t f(x, y) = f(W_t(x, y), y),
\]

where \( W_t : M_N(\mathbb{C})^{n \times n}_{s_a} \times M_N(\mathbb{C})^{n \times n}_{s_a} \to M_N(\mathbb{C})^{n \times n}_{s_a} \) is the solution to the initial value problem

\[
W_0(x, y) = x
\]

\[
\partial_t W_t(x, y) = -\frac{1}{2} D_x V(W_t(x, y), y).
\]

This solution is defined for all \( t \geq 0 \) by the Picard-Lindelöf theorem because \( D_x V(x, y) \) is globally Lipschitz in \( x \) (compare [4.1]).

Proposition 5.10. There exists a semigroup \( T_t \) acting on Lipschitz functions \( M_N(\mathbb{C})^{n \times n}_{s_a} \times M_N(\mathbb{C})^{n \times n}_{s_a} \) such that the following hold:

1. If \( t = n/2^\ell \) is a dyadic rational, let \( T_{t, \ell} f = (P_{2^{-\ell}} S_{2^{-\ell}})^n f \). Then \( T_{t, \ell} f \to T_t f \) as \( \ell \to \infty \) and more precisely

\[
\| T_{t, \ell} f(\cdot, y) - T_t f(\cdot, y) \|_{L^\infty} \leq \frac{C m^{1/2}}{c(2 - 2^{1/2})^{\ell/2}} 2^{-\ell/2} \| f(\cdot, y) \|_{Lip}.
\]
Lemma 5.11. With the setup above, we have for Lipschitz functions
\[ \|T_t f(x, y) - T_s f(x, y)\|_2 \leq e^{-ct/2} \left( \frac{C}{c^2} (6 + 5\sqrt{2})(t-s)^{1/2} + \|D_x V(x, y)\|_2 \right) \|f(\cdot, y)\|_{\text{Lip}}. \]

Proof. These results follow by freezing the variable $y$ and applying the results from our previous paper, specifically,

1. see [17, Lemma 4.5],
2. see [17, Lemma 4.6],
3. see [17, Lemma 4.6],
4. see [17, Lemma 4.8],
5. see [17, Lemma 4.9].

The results of [17, §4] were stated only for scalar-valued functions. However, the arguments hold for functions from $M_N(\mathbb{C})_m \times M_N(\mathbb{C})_n$ to any finite-dimensional normed vector space. The result (4) that $T_t$ is expectation-preserving follows immediately by applying the scalar-valued result to each coordinate of the vector-valued function in some basis. To verify the estimates, one simply replaces the “$\cdot$” in the arguments by the appropriate norm, which in our case would be $\|\cdot\|_2$ on $M_N(\mathbb{C})$. \qed

We will next show that $W_t(x, y)$ and $T_t f(x, y)$ depend in a Lipschitz manner upon $y$. Let us denote
\[
\|f\|_{\text{Lip}(dx)} = \sup_y \|f(\cdot, y)\|_{\text{Lip}}
\]
\[
\|f\|_{\text{Lip}(dy)} = \sup_x \|f(x, \cdot)\|_{\text{Lip}}.
\]

Lemma 5.11. With the setup above, we have for Lipschitz $f : M_N(\mathbb{C})_m \times M_N(\mathbb{C})_n \rightarrow M_N(\mathbb{C})$

1. $\|W_t\|_{\text{Lip},dx} \leq e^{-ct/2}$ and $\|W_t\|_{\text{Lip},dy} \leq (C/c)(1 - e^{-ct/2}).$
2. $\|S_t f\|_{\text{Lip},dx} \leq e^{-ct/2} \|f\|_{\text{Lip},dx}.$
3. $\|S_t f\|_{\text{Lip},dy} \leq \|f\|_{\text{Lip},dy} + (C/c)(1 - e^{-ct/2}) \|f\|_{\text{Lip},dx}.$
4. $\|P_t f\|_{\text{Lip},dy} \leq \|f\|_{\text{Lip},dy}$ and $\|P_t f\|_{\text{Lip},dx} \leq \|f\|_{\text{Lip},dx}.$
5. $\|T_t f\|_{\text{Lip},dx} \leq e^{-ct/2} \|f\|_{\text{Lip},dx}.$
6. $\|T_t f\|_{\text{Lip},dy} \leq \|f\|_{\text{Lip},dy} + (C/c)(1 - e^{-ct/2}) \|f\|_{\text{Lip},dx}.$

Proof. (1) Fix $x, x' \in M_N(\mathbb{C})_m$ and $y, y' \in M_N(\mathbb{C})_n$. Define
\[ \phi(t) = \|W_t(x, y) - W_t(x', y')\|_2. \]
Note that $\phi$ is locally Lipschitz in $t$ and hence absolutely continuous. Moreover, $\phi(t)^2$ is $C^1$ with

$$\partial_t[\phi(t)^2] = 2\langle \partial_t W_t(x, y) - \partial_t W_t(x', y'), W_t(x, y) - W_t(x', y') \rangle_2$$

$$= -\langle D_x V(W_t(x, y), y) - D_x V(W_t(x', y'), y), W_t(x, y) - W_t(x', y') \rangle_2$$

$$= -\langle D_x V(W_t(x, y), y) - D_x V(W_t(x', y'), y), W_t(x, y) - W_t(x', y') \rangle_2$$

$$- \langle D_x V(W_t(x', y'), y) - D_x V(W_t(x', y'), y), W_t(x, y) - W_t(x', y') \rangle_2$$

$$\leq -c\|W_t(x, y) - W_t(x', y')\|^2_2 + \|D_x V(W_t(x', y'), y) - D_x V(W_t(x', y'), y)\|_2\|W_t(x, y) - W_t(x', y')\|_2$$

Here we have employed the inequality $\langle D_x V(z, w) - D_x V(z', w), z - z' \rangle_2 \geq c\|z - z'\|^2_2$ coming from the uniform convexity of $V$ as well as the Cauchy-Schwarz inequality. This implies that

$$2\phi'(t)\phi(t) = \partial_t[\phi(t)^2] \leq -c\phi(t)^2 + C\|y - y'\|\phi(t).$$

Thus, $\phi'(t) \leq -(c/2)\phi(t) + (C/2)\|y - y'\|$, so that $\partial_t[e^{ct/2}\phi(t)] \leq (C/2)e^{ct/2}\|y - y'\|_2$. This implies that

$$e^{ct/2}\phi(t) - \phi(0) \leq \frac{C}{c}(e^{ct/2} - 1)\|y - y'\|_2.$$

But $\phi(t) = \|W_t(x, y) - W_t(x', y')\|_2$ and $\phi(0) = \|x - x'\|_2$. Hence,

$$\|W_t(x, y) - W_t(x', y')\|_2 \leq e^{-ct/2}\|x - x'\|_2 + \frac{C}{c}(1 - e^{-ct/2})\|y - y'\|_2.$$

This proves both estimates of (1).

(2) This is immediate since $S_t f(x, y) = f(W_t(x, y), y)$, as in [17, Lemma 4.4 (5)].

(3) Note that

$$\|S_t f(x, y) - S_t f(x', y')\|_2 = \|f(W_t(x, y), y) - f(W_t(x', y'), y')\|_2$$

$$\leq \|f(W_t(x, y), y) - f(W_t(x, y), y')\|_2 + \|f(W_t(x, y), y') - f(W_t(x', y'), y')\|_2$$

$$\leq \|f\|_{Lip, dy}\|y - y'\|_2 + \|f\|_{Lip, dx}\|W_t(x, y) - W_t(x', y')\|_2$$

$$\leq \|f\|_{Lip, dy}\|y - y'\|_2 + \frac{C}{c}(1 - e^{-ct/2})\|f\|_{Lip, dx}\|y - y'\|_2.$$

(4) This follows from basic properties of convolution of a function with a probability measure.

(5) By iterating the estimates (2) and (4), we obtain $\|T_{t, t} f\|_{Lip, dx} \leq e^{-ct/2}\|f\|_{Lip, dx}$. Then by Proposition \[5.10\] (2) and (3) we may take $\ell \to \infty$ and then extend to all real values of $t \geq 0$.

(6) First, consider $T_{t, \ell}$ for a dyadic rational $t = n/2^\ell$. Denote $\delta = 2^{-\ell}$. For $j = 0, \ldots, n - 1$, we have

$$\|T_{(j+1)\delta, t} f\|_{Lip, dy} = \|P_{\delta} S_{\delta} T_{j \delta, t} f\|_{Lip, dy}$$

$$\leq \|S_{\delta} T_{j \delta, t} f\|_{Lip, dy}$$

$$\leq \|T_{j \delta, t} f\|_{Lip, dy} + \frac{C}{c}(1 - e^{-c\delta/2})\|T_{j \delta, t} f\|_{Lip, dx}.$$
where the last inequality follows from (3). Therefore, by induction
\[
\|T_{t,\ell} f\|_{\text{Lip},dy} = \|T_{n,\ell} f\|_{\text{Lip},dy} \\
\leq \|f\|_{\text{Lip},dy} + \frac{C}{c} (1 - e^{-ct/2}) \sum_{j=0}^{n-1} \|T_{j,\ell} f\|_{\text{Lip},dx} \\
\leq \|f\|_{\text{Lip},dy} + \frac{C}{c} (1 - e^{-ct/2}) \sum_{j=0}^{n-1} e^{-cj/2} \|f\|_{\text{Lip},dx} \\
= \|f\|_{\text{Lip},dy} + \frac{C}{c} (1 - e^{-ct/2}) \|f\|_{\text{Lip},dx}.
\]
In light of Proposition 5.11 (2), we can take \( \ell \to \infty \) and conclude that \( \|T_{t} f\|_{\text{Lip},dy} \leq \|f\|_{\text{Lip},dy} + (C/c)(1 - e^{-ct/2}) \|f\|_{\text{Lip},dx} \) for dyadic rational \( t \). This inequality can then be extended to all real \( t \geq 0 \) by Proposition 5.11 (3).

**Corollary 5.12.** Let \( f : M_{N}(C)^{m} \times M_{N}(C)^{n} \to M_{N}(C) \) be Lipschitz with respect to \( \|\cdot\|_{2} \). Let \( g(y) = \int f(x, y) d\mu(x|y) \). Then \( g \) is Lipschitz with
\[
\|g\|_{\text{Lip}} \leq (1 + C/c) \|f\|_{\text{Lip}}.
\]

**Proof.** By the previous lemma,
\[
\|T_{t} f\|_{\text{Lip},dy} \leq \|f\|_{\text{Lip},dy} + \frac{C}{c} (1 - e^{-ct/2}) \|f\|_{\text{Lip},dx} \\
\leq \|f\|_{\text{Lip}} + \frac{C}{c} \|f\|_{\text{Lip}}.
\]
As \( t \to \infty \), we have \( T_{t} f(x, y) \to g(y) \) by Proposition 5.11 (5). Hence, \( \|g\|_{\text{Lip}} \leq (1 + C/c) \|f\|_{\text{Lip}}. \)

**5.5. Asymptotic Approximation and Convergence.** Let \( V^{(N)} \) and \( \mu^{(N)} \) be as in Theorem 5.9 and let \( (X^{(N)}, Y^{(N)}) \) be a random variable with distribution \( \mu^{(N)} \). Let \( \mu^{(N)}(x|y) \) denote the conditional distribution of \( X^{(N)} \) given \( Y^{(N)} \).

Let \( P_{t}^{(N)}, S_{t}^{(N)}, \) and \( T_{t}^{(N)} \) be the semigroups acting on Lipschitz functions defined as in §5.4 with respect to the potential \( V^{(N)} \).

**Lemma 5.13.** With the notation above, suppose that \( f^{(N)} : M_{N}(C)^{m} \to M_{N}(C) \), that \( f^{(N)} \) is \( K \)-Lipschitz for every \( N \), and that \( f^{(N)} \) is asymptotically approximable by trace polynomials. Then

1. \( \{P_{t}^{(N)} f^{(N)}\} \) is asymptotically approximable by trace polynomials.
2. \( \{S_{t}^{(N)} f^{(N)}\} \) is asymptotically approximable by trace polynomials.
3. \( \{T_{t}^{(N)} f^{(N)}\} \) is asymptotically approximable by trace polynomials.

**Proof.** (1) We proved in Lemma 4.12 that \( P_{t}^{(N)} \) preserves asymptotic approximability by trace polynomials.

(2) Recall that \( S_{t}^{(N)} f^{(N)}(x, y) = f^{(N)}(W_{t}^{(N)}(x, y), y) \), where
\[
W_{0}^{(N)}(x, y) = x \]
\[
\partial_{t} W_{t}^{(N)}(x, y) = -\frac{1}{2} D_{x} V^{(N)}(W_{t}(x, y), y).
\]
Now \( D_{x} V^{(N)}(x, y) \) is \( C \)-Lipschitz in \((x, y)\), asymptotically approximable by trace polynomials, and independent of \( t \), and thus it satisfies Assumption 4.16 so by Proposition 4.17, \( W_{t}^{(N)}(x, y) \)
is asymptotically approximable by trace polynomials (here we rely on Lemma 3.4 that asymptotic approximability is equivalent to being asymptotic to some element of $\text{TrP}_{m+n}$). Then because $f^{(N)}$ is $K$-Lipschitz in $(x, y)$, Lemma 3.10 implies asymptotic approximability of $f^{(N)}(W^{(N)}_{t}(x, y))$.

(3) Let $T_{t}^{(N)} = P_{2^{-t}}S_{2^{-t}}^{(N)}$ whenever $t = n2^{-\ell}$. From (1) and (2), it follows that $T_{t}^{(N)}f^{(N)}$ is asymptotically approximable by trace polynomials. Now for each dyadic $t$, Proposition 5.10(1) shows that $T_{t}^{(N)}f^{(N)} \to T_{t}^{(N)}f^{(N)}$ uniformly on $\| \cdot \|_{2}$-balls (and hence on $\| \cdot \|_{\infty}$). Therefore, by Lemma 3.11, $T_{t}^{(N)}f^{(N)}$ is asymptotically approximable by trace polynomials. Then we extend this property from dyadic $t$ to all real $t$ using Proposition 5.10(2) and Lemma 3.11.

**Proof of Theorem 5.9**. Let $f^{(N)} : M_{N}(\mathbb{C})^{m+n} \to M_{N}(\mathbb{C})$ be $K$-Lipschitz and asymptotically approximable by trace polynomials. Let

$$g^{(N)}(y) = \int f^{(N)}(x, y) \, d\mu^{(N)}(x|y)$$

We showed in Corollary 5.12 that $g^{(N)}$ is Lipschitz with $\|g^{(N)}\|_{\text{Lip}} \leq (1 + C/c)\|f^{(N)}\|_{\text{Lip}}$.

We know that $T_{t}^{(N)}f^{(N)}$ is asymptotically approximable by trace polynomials in $(x, y)$. By Proposition 5.10(5), we have $T_{t}^{(N)}f^{(N)}(x, y) \to g^{(N)}(x, y)$ as $t \to \infty$, with the error bounded by

$$e^{-ct/2} \left( 4C_{2}^{-1}(6 + 5\sqrt{2})t^{-1/2} + \frac{2}{c}\|D_{x}V(x, y)\|_{2} \right) .$$

Given that $\{DV^{(N)}\}$ is asymptotically approximable by trace polynomials, $\|D_{x}V^{(N)}\|_{u,R}$ is bounded as $N \to \infty$. This implies that the rate of convergence of $T_{t}^{(N)}f^{(N)}(x, y) \to g^{(N)}(x, y)$ as $t \to \infty$ is uniform on $\| (x, y) \|_{\infty} \leq R$ and independent of $N$. So by Lemma 3.11, $g^{(N)}$ is asymptotically approximable by trace polynomials of $(x, y)$. Yet $g^{(N)}$ is independent of $x$, and so we may approximate $g^{(N)}(y)$ by evaluating these trace polynomials at $(0, y)$, which reduces them to trace polynomials of $y$.

Since $g^{(N)}$ is asymptotically approximable by trace polynomials, let $g \in \text{TrP}_{m}$ such that $g^{(N)} \to g$. Then it remains to show that $g(Y) = E_{W_{\lambda}(Y)}[f(X, Y)]$, where $(X, Y)$ are non-commutative random variables for the free Gibbs law $\lambda$ as in the theorem statement. It suffices to check that

$$\tau(\phi(Y)g(Y)) = \tau(\phi(Y)f(X, Y))$$

whenever $\phi$ is a non-commutative polynomial. But using Corollary 5.3

$$\tau(\phi(Y)g(Y)) = \lim_{N \to \infty} E[\tau_{N}(\phi(Y^{(N)})g^{(N)}(Y))]$$

$$= \lim_{N \to \infty} E[\tau_{N}(\phi(Y^{(N)})f^{(N)}(X^{(N)}Y^{(N)}))]$$

$$= \tau(\phi(Y)f(X, Y)).$$

\(\square\)

**Remark 5.14**. We showed in [5.2] that $F_{t}^{(N)}$ has a large $N$ limit $F_{t}^{\text{TrP}}$ acting on $\text{TrP}_{m+n}$. Similarly, the results of [3.1] imply that $S_{t}^{(N)}$ has a large $N$ limit $S_{t}^{\text{TrP}}$ acting on $\text{TrP}_{m+n}$. This implies that the semigroup $T_{t}^{(N)}$ also has a large $N$ limit $T_{t}^{\text{TrP}}$ in light of Proposition 5.10(1) and (2) and Lemma 3.11. Future research should investigate in what sense $F(x, t) = T_{t}^{\text{TrP}}f(x)$ would solve the differential equation

$$\partial_{t}F = \frac{1}{2}L_{x}F - \frac{1}{2}(J_{x}F)^{\ast}(D_{x}V),$$

where $V$ is the large $N$ limit of $\{V^{(N)}\}$.
6. Conditional Entropy and Fisher’s Information

In this section, we show that for random matrix models satisfying Assumption 5.1, the conditional (classical) entropy $h(X^{(N)}|Y^{(N)})$ converges to the conditional non-microstates free entropy $\chi^*(X|Y)$ (also known as $\chi^*(X : W^*(Y))$).

6.1. Conditional Entropy and Fisher’s Information in the Classical Setting. We refer to [33, §3] and [17, §5] for background on classical entropy and Fisher’s information and motivation for the free case. The conditional setting is more technical, and we will state several standard results without proof, since the proofs in the non-conditional case were repeated in some detail in [17].

Recall that the classical entropy of a random variable $X$ in $\mathbb{R}^m$ with probability density $\rho$ is $h(X) = -\int \rho \log \rho$. Similarly, if $(X,Y)$ is a random variable in $\mathbb{R}^m \times \mathbb{R}^n$ with density $\rho_{X,Y}(x,y)$, then the conditional entropy $h(X|Y)$ is defined by

$$h(X|Y) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \rho_{X|Y}(x|y) \log \rho_{X|Y}(x|y) \, dx \rho_Y(y) \, dy = \int_{\mathbb{R}^m \times \mathbb{R}^n} (\log \rho_{X|Y}(x|y)) \rho(x,y) \, dx \, dy,$$

where $\rho_Y$ is the marginal density

$$\rho_Y(y) = \int_{\mathbb{R}^m} \rho_{X,Y}(x,y) \, dx$$

and $\rho_{X|Y}$ is the conditional density

$$\rho_{X|Y}(x|y) = \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)} \text{ defined when } \rho_Y(y) > 0.$$

It is a standard fact that if $X$ has finite variance, then $h(X|Y)$ is well-defined. The proof for the non-conditional entropy $h(X)$ was reviewed in [17, Lemma 5.1], and the conditional case can be handled similarly.

The conditional Fisher information given by

$$\mathcal{I}(X|Y) = \int_{\mathbb{R}^m \times \mathbb{R}^n} \left| \frac{\nabla_x \rho_{X|Y}(x|y)}{\rho_{X|Y}(x|y)} \right|^2 \rho_{X,Y}(x,y) \, dx \, dy,$$

whenever the right hand side makes sense and $\infty$ otherwise. It describes the rate of change of $h(X + t^{1/2}S|Y)$, where $S$ is a Gaussian random variable in $\mathbb{R}^m$ with covariance matrix $I$ independent of $(X,Y)$. Knowing that the density $\rho_{X + t^{1/2}S|Y}$ satisfies the heat equation

$$\partial_t \rho_{X + t^{1/2}S|Y} = \frac{1}{2} \Delta \rho_{X + t^{1/2}S|Y},$$

one can show that $\mathcal{I}(X + t^{1/2}S|Y)$ is well-defined and finite for $t > 0$ and that

$$\frac{d}{dt} h(X + t^{1/2}S|Y) = \frac{1}{2} \mathcal{I}(X + t^{1/2}S|Y).$$

The Fisher information is the $L^2$ norm of the $(\mathbb{R}^m$-valued) random variable $\Xi$ given by evaluating the score function $-\nabla_x \rho_{X|Y}/\rho_{X|Y}$ on the random variable $(X,Y)$, provided that this random variable is in $L^2$. In this case, the random variable $\Xi$ is known as the conjugate variable for $X$ given $Y$, and it is the unique element of $L^2$ satisfying the integration-by-parts relation

$$E[\Xi f(X,Y)] = E[\nabla_x f(X,Y)]$$

for all $f \in C_0^\infty(\mathbb{R}^m \times \mathbb{R}^n)$.

More generally, if there exists a random variable $\Xi$ in $L^2$ satisfying this integration-by-parts formula, then we define the conditional Fisher information to be $\mathcal{I}(X|Y) = E[|\Xi|^2]$ (and this extends our previous definition of $\mathcal{I}(X|Y)$). Otherwise, $\mathcal{I}(X|Y)$ is defined to be $\infty$. 
In light of the integration-by-parts characterization, conjugate variables behave well under conditionally independent sums. The following lemma is proved in the same way as the non-conditional case (see [17, Lemma 5.6]) and the free case (see [30, Proposition 3.7]).

**Lemma 6.1.** Let $Y$ be a random variable in $\mathbb{R}^n$ and let $X_1$ and $X_2$ be random variables in $\mathbb{R}^m$ that are conditionally independent given $Y$. Suppose that $\Xi$ is a conjugate variable for $X_i$ given $Y$. Then $E[\Xi|X_1 + X_2, Y]$ is a conjugate variable for $X_1 + X_2$ given $Y$. Hence,

$$\mathcal{I}(X_1 + X_2|Y) \leq \mathcal{I}(X_1|Y).$$

In particular, this holds if $X_2$ is independent from $(X_1,Y)$ or $X_1$ is independent of $(X_2,Y)$.

Conjugate variables also scale in the following way. The proof is straightforward from the integration-by-parts relation.

**Lemma 6.2.** If $\Xi$ is a conjugate variable for $X$ given $Y$ and $t > 0$, then $(1/t)\Xi$ is a conjugate variable for $tX$ given $Y$, and hence $\mathcal{I}(tX|Y) = t^{-2}\mathcal{I}(X|Y)$.

### 6.2. Random Matrix Renormalization

Suppose that $(X,Y)$ is a random variable in $M_N(\mathbb{C})^{m\times n} \times M_N(\mathbb{C})^{n\times m}$ with density $\rho_{X,Y}$. The trace on $M_N(\mathbb{C})^{m\times n}$ produces a real inner product. But to study the large $N$ limit, we use the normalized trace $\tau_N = (1/N)\text{Tr}$. The corresponding normalized Gaussian is the GUE ensemble $S = (S_1, \ldots, S_m)$ where $S_j$ has variance 1 with respect to $\tau_N$. We use the following renormalized entropy, which is motivated by computation of the Gaussian case and by (6.5) below,

$$h^{(N)}(X|Y) = \frac{1}{N^2}h(X|Y) + \frac{m}{2}\log N.$$

Due to the normalization of Gaussian, the evolution of the density for $(X + t^{1/2}S, Y)$ is given by the renormalized heat equation

$$\partial_t \rho_{X+t^{1/2}S,Y} = \frac{1}{2N^3}\Delta \rho_{X+t^{1/2}S,Y}.$$

This results in

$$\partial_t h^{(N)}(X + t^{1/2}S|Y) = \frac{1}{2N^3} \mathcal{I}(X + t^{1/2}S|Y) = \frac{1}{2} \mathcal{I}^{(N)}(X + t^{1/2}S|Y),$$

where $\mathcal{I}^{(N)}(X|Y) := N^{-3} \mathcal{I}(X|Y)$, assuming that $X$ has finite variance and $t > 0$.

Another heuristic for the normalization $\mathcal{I}^{(N)} = N^{-3} \mathcal{I}$ comes from analyzing the case where $(X,Y)$ have density $(1/Z)e^{-N^2V(x,y)}dxdy$ where $V$ is uniformly convex and semi-concave. Indeed, in this case, the classical conjugate variable for $X$ given $Y$ is $-N^2 \nabla_x V(X,Y)$. Recall that $D_x V = N \nabla_x V$ is the gradient of $V$ with respect to the normalized inner product $\langle \cdot, \cdot \rangle_2$. Thus,

$$\frac{1}{N^3} \mathcal{I}(X|Y) = \frac{1}{N} E\|N \nabla_x V(X,Y)\|_2^2 = E\|D_x V(X,Y)\|_2^2$$

is dimension-independent normalization. Furthermore, the normalized conjugate variable $\xi = (1/N)\Xi$ (which would be $D_x V(X,Y)$ in the case where the law is given by a potential $V$) satisfies the integration-by-parts relation

$$E(\xi_j, f(X,Y))_2 = \frac{1}{N^2} \text{Div}_x, f(X,Y),$$

where $\xi = (\xi_1, \ldots, \xi_m)$ and where Div is the divergence with respect to the classical coordinates (not normalized). But if $f$ is a non-commutative polynomial, then

$$\frac{1}{N^2} \text{Div}_x, f(x,y) = \frac{1}{N^2} \text{Tr} \otimes \text{Tr}(D_x f(x,y)) = \tau_N \otimes \tau_N(D_x f(x,y)).$$
where $\mathcal{D}_{x_j}$ denotes the non-commutative derivative or free difference quotient with respect to $x_j$. Thus, applying the integration-by-parts relation to non-commutative polynomials results in the dimension-independent relation

$$E(\xi_j, f(X,Y))_2 = \sum_{j=1}^m E[\tau_N \otimes \tau_N(\mathcal{D}_{x_j}, f(x,y))]$$

that characterizes the normalized conjugate variable.

As a consequence of (6.5), $h^{(N)}(X|Y)$ can be recovered by integrating $\mathcal{I}^{(N)}(X + t^{1/2}S|Y)$ and modifying the integral to converge at $\infty$. This results in

$$h^{(N)}(X|Y) = \frac{1}{2} \int_0^\infty \left( \frac{m}{1+t} - \mathcal{I}^{(N)}(X + t^{1/2}S|Y) \right) dt + \frac{m}{2} \log 2\pi e$$

provided that $(X, Y)$ has a density $\rho_{X,Y}$ and that $X$ has finite variance. The proof is similar to [14, Lemma 5.7]. Convergence of the integral at $\infty$ can be deduced from the following estimate, and it also shows convergence of the integral at 0 if $\mathcal{I}(X|Y)$ is finite. Compare [30, Corollary 6.14 and Remark 6.15] and [14, Lemma 5.7].

**Lemma 6.3.** Let $(X, Y)$ be a random variable in $\mathcal{M}_N(\mathbb{C})_{sa} \times \mathcal{M}_N(\mathbb{C})_{sa}$ such that $a = (1/m) \sum_{j=1}^m E[\tau_N(\xi_j^2)] < \infty$, and let $S$ be an independent GUE $m$-tuple. Then

$$\frac{m}{a+t} \leq \mathcal{I}^{(N)}(X + t^{1/2}S|Y) \leq \min\left(\frac{m}{t}, \mathcal{I}^{(N)}(X|Y)\right).$$

**Proof.** We observe that $\xi_i = E[t^{-1/2}S|X + t^{1/2}S,Y]$ is a normalized conjugate variable for $X + t^{1/2}S$ given $Y$ by Lemma 6.1. This yields $\mathcal{I}^{(N)}(X + t^{1/2}S|Y) \leq m/t$. On the other hand, if $\xi$ is a normalized conjugate variable for $X$ given $Y$, we also have $\xi_i = E[|\xi|X + t^{1/2}S,Y]$, which yields the upper bound $\mathcal{I}^{(N)}(X|Y)$. The lower bound follows from observing $(E[|\xi|^2]^{1/2}(E[|X + t^{1/2}S|^2]^{1/2} \geq E(\xi_i, X + t^{1/2}S)_2$ and evaluating the right hand side using integration by parts. \hfill \Box

### 6.3. Convergence to Conditional Free Entropy

Motivated by the normalized entropy and Fisher’s information in the previous section, Voiculescu defined the free versions as follows. Let $(X, Y)$ be an $(m+n)$-tuple of self-adjoint non-commutative random variables in a tracial $W^*$-algebra $(\mathcal{M}, \tau)$. We say that $\xi = (\xi_1, \ldots, \xi_m) \in L^2(\mathcal{M}, \tau)_{sa}$ is a **free conjugate variable for $X$ given $Y$** if for every non-commutative polynomial $f(X,Y)$, we have

$$\tau(\xi_j f(X,Y)) = \tau(\mathcal{D}_{x_j} f(X,Y)).$$

The **free Fisher information** $\Phi^*(X|Y)$ is defined to be $\|\xi\|^2_2$ if such a $\xi$ exists, and $\infty$ otherwise. The **non-microstates free entropy** $\chi^*(X|Y)$ is defined to be

$$\chi^*(X|Y) = \frac{1}{2} \int_0^\infty \left( \frac{m}{1+t} - \Phi^*(X + t^{1/2}S|Y) \right) dt + \frac{m}{2} \log 2\pi e.$$

Convergence of the integral at $\infty$ follows from the free analogue of Lemma 6.3 so that $\chi^*(X|Y)$ is well-defined in $[-\infty, \infty)$ whenever $X$ has finite variance.

**Remark 6.4.** Voiculescu’s original notation in [30, §7] was $\chi^*(X : W^*(Y))$ rather than $\chi^*(X|Y)$, since the definition of the conjugate variable can be rephrased so as to depend only on $W^*(Y)$ rather than $Y$. Indeed, one can show that the conjugate variables $\xi$ for $X$ given $Y$ as defined above would satisfy

$$\tau(\xi_j p(X,Z)) = \tau(\mathcal{D}_{x_j} p(x,z)),$$

whenever $p$ is non-commutative polynomial in $m+k$ variables $Z = (Z_1, \ldots, Z_k)$ is a self-adjoint tuple in $W^*(Y)$ for some $k \in \mathbb{N}$. This is clearly true if the $Z_j$’s polynomials of $Y$, and the case of general $Z_j \in W^*(Y)_{sa}$ follows by a routine density argument.
Theorem 6.6. Let $\mathcal{X}$ be distributed according to $\mu$. We want to show that the law of $(X, Y)$ for microstates entropy of $X$ in the presence of $Y$ used in [25, §1].

The following lemma gives sufficient conditions for classical Fisher information for random matrix models to converge to free Fisher information. The main hypotheses are that the non-commutative laws converge, the score functions $D_x V(N)$ for the $N \times N$ matrix models are asymptotically approximable by trace polynomials, and some mild growth conditions on score functions and probability measures as $\| (x, y) \| \to \infty$. We omit the proof since it is a direct adaptation of the proof of [17, Proposition 5.10].

Lemma 6.5. Let $V(N) : M_N(\mathbb{C})_{sa}^{m+n} \to \mathbb{R}$ be a potential with $\int_{M_N(\mathbb{C})_{sa}^{m+n}} e^{-N^2 V_N(x, y)} \, dx \, dy < +\infty$, let $\mu(N)$ be the associated probability density, and let $(X(N), Y(N))$ be a random variable distributed according to $\mu(N)$. Let $(X, Y)$ be an $(m + n)$-tuple of self-adjoint non-commutative random variables in the tracial $W^*$-algebra $(\mathcal{M}, \tau)$. Assume that:

(A) The non-commutative law of $(X(N), Y(N))$ with respect to $\tau_N$ converges in probability to the non-commutative law of $(X, Y)$.

(B) $D_x V(N)$ is defined and continuous, and the sequence $\{D_x V(N)\}$ is asymptotically approximable by trace polynomials, and hence $D_x V(N) \to g \in (\text{TrP}_{m+n})_{sa}$.

(C) For some $k \geq 0$ and $a, b > 0$, we have

$$\|D_x V(N)(x, y)\|^2_2 \leq a + b \| (x, y) \|^k$$

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

$$\lim_{N \to \infty} E \left[ 1_{\|(X(N), Y(N))\| \geq R_0} \left( 1 + \|(X(N), Y(N))\|^k \right) \right] = 0.$$  

Then $\mathcal{I}(X|Y)$ is finite. Moreover, $g(X, Y)$ is in $L^2(\mathcal{M}, \tau)$ and it is the free conjugate variable for $X$ given $Y$, and we have

$$\mathcal{I}(X|Y) = E[\|D_x V(N)(X(N), Y(N))\|^2_2] \to \|g(X, Y)\|^2_2 = \Phi^*(X|Y).$$

Theorem 6.6. Let $V(N) : M_N(\mathbb{C})_{sa}^m \times M_N(\mathbb{C})_{sa}^n \to \mathbb{R}$ satisfy Assumption 5.1 for some $0 < c \leq C$. Let $\mu(N)$ be the corresponding measure, let $X(N), Y(N)$ be random variables chosen according to $\mu(N)$, and let $S(N)$ be an independent $m$-tuple of GUE matrices.

Let $X = (X_1, \ldots, X_m)$ and $Y = (Y_1, \ldots, Y_n)$ be non-commutative random variables with non-commutative law $\mu = \mu_V$, and let $S$ be a freely independent free semicircular $m$-tuple. Then for every $t \geq 0$, we have

$$\Phi^*(X + t^{1/2} S|Y) = \lim_{N \to \infty} \mathcal{I}(X(N) + t^{1/2} S(N)|Y(N))$$

and

$$\chi^*(X + t^{1/2} S|Y) = \lim_{N \to \infty} h(N)(X(N) + t^{1/2} S(N)|Y(N)).$$

Proof. We want to show that the law of $(X(N) + t^{1/2} S(N), Y(N))$ satisfies the assumptions of Lemma 6.5 for each $t \geq 0$. The joint law of $(X(N), Y(N), S(N))$ is given by the convex potential $U(N)(x, y, s) = V(x, y) + (1/2)\|s\|^2_2$. Now $U(N)$ satisfies $\min(c, 1) \leq \text{HU}(N) \leq \max(C, 1)$ and $D U(N)$ is asymptotically approximable by trace polynomials. Thus, the law of $(X(N), Y(N), S(N))$ has a large $N$ limit given by Theorem 5.2. In fact, the large $N$ limit must be non-commutative law of $(X, Y, S)$ because of Voiculescu’s asymptotic freeness theorem 31 and because the non-commutative law of $S(N)$ converges to the non-commutative law of $S$. (Alternatively, this could be proved the same way as [17, Lemma 7.4].)
Since the non-commutative law of \((X^{(N)}, Y^{(N)}, S^{(N)})\) converges in probability to that of \((X, Y, S)\), the non-commutative law of \((X^{(N)} + t^{1/2}S^{(N)}, Y^{(N)})\) converges in probability to that of \((X + t^{1/2}S, Y)\), and thus (A) of Lemma 6.5 holds. Moreover, Lemma 2.12 shows that

\[ P(||X^{(N)} - E(X^{(N)}), Y^{(N)} - E(Y^{(N)}), S^{(N)})||_\infty \geq \min(1, c)^{-1/2}(\Theta + \delta)) \leq e^{-N\delta^2/2}. \]

From this it is not hard to show that \((X^{(N)} + t^{1/2}S^{(N)}, Y^{(N)})\) satisfies (D).

It remains check (B) and (C). The potential for \((X^{(N)} + t^{1/2}S^{(N)}, Y^{(N)})\) is given by

\[ W_t^{(N)}(\tilde{x}, y, s) = U^{(N)}(\tilde{x} - t^{1/2}s, y, s), \]

which follows by applying the change of variables formula for the density. Here we write \(\tilde{x}\) to emphasize that this variable corresponds to \(X^{(N)} + t^{1/2}S^{(N)}\) rather than \(X^{(N)}\). Note that \(W_t^{(N)}\) is uniformly convex and semiconcave since it is the composition of \(U^{(N)}\) with an invertible linear transformation. Also,

\[ DW_t^{(N)}(\tilde{x}, y, s) = (DV_t^{(N)}(\tilde{x} - t^{1/2}s, y), s - t^{1/2}DV_t^{(N)}(\tilde{x} - t^{1/2}s, y)) \]

is asymptotically approximable by trace polynomials. The potential corresponding to \((X^{(N)} + t^{1/2}S^{(N)}, Y^{(N)})\) is

\[ V_t^{(N)}(\tilde{x}, y) = -\frac{1}{N^2} \log \int e^{-N^2W_t^{(N)}(\tilde{x}, y, s)} ds. \]

Since \(W_t^{(N)}\) is uniformly convex, the integrand vanishes rapidly at \(t\), and thus it is straightforward to differentiate under the integral by dominated convergence, and deduce that \(V_t^{(N)}\) is continuously differentiable. Furthermore,

\[ DV_t^{(N)}(\tilde{x}, y)e^{-N^2V_t^{(N)}(\tilde{x}, y)} = \int DW_t^{(N)}(\tilde{x}, y, s)e^{-N^2W_t^{(N)}(\tilde{x}, y, s)} ds, \]

so that

\[ DV_t^{(N)}(\tilde{x}, y) = \frac{\int DW_t^{(N)}(\tilde{x}, y, s)e^{-N^2W_t^{(N)}(\tilde{x}, y, s)} ds}{\int e^{-N^2W_t^{(N)}(\tilde{x}, y, s)} ds}, \]

or in other words \(DV_t^{(N)}\) is given by the conditional expectation

\[ DV_t^{(N)}(X^{(N)} + t^{1/2}S^{(N)}, Y^{(N)}) \]

(6.10)

\[ = E \left[ D_{(x,y)W_t^{(N)}(X^{(N)} + t^{1/2}S^{(N)}, Y^{(N)}, S^{(N)})}|X^{(N)} + t^{1/2}S^{(N)}, Y^{(N)} \right] \]

\[ = E \left[ DV_t^{(N)}(X^{(N)}, Y^{(N)})|X^{(N)} + t^{1/2}S^{(N)}, Y^{(N)} \right] \]

Now we apply Theorem 5.9 using the potential \(W_t^{(N)}\) and conditioning on \((X^{(N)} + t^{1/2}S^{(N)}, Y^{(N)})\) to conclude that \(DV_t^{(N)}(\tilde{x}, y)\) is asymptotically approximable by trace polynomials, which establishes (B).

Furthermore, Theorem 5.9 implies that

\[ \|DV_t^{(N)}\|_{\text{Lip}} \leq (1+C/c)\|D_{(x,y)W_t^{(N)}(X^{(N)} + t^{1/2}S^{(N)}, Y^{(N)}, S^{(N)})}|X^{(N)} + t^{1/2}S^{(N)}, Y^{(N)} \|_{\text{Lip}} \leq (1+C/c)(1+t^{1/2})C. \]

This implies that (C) of Lemma 6.5 holds with \(k = 1\), using Remark 4.13.

Therefore, we may apply Lemma 6.5 to \((X^{(N)} + t^{1/2}S^{(N)}, Y^{(N)})\) to obtain that (6.8) holds for every \(t \geq 0\), that is,

\[ \lim_{N \to \infty} \mathcal{T}^{(N)}(X^{(N)} + t^{1/2}S^{(N)}|Y^{(N)}) \to \Phi^*(X + t^{1/2}S|Y), \]

For the second claim (6.9) regarding \(h^{(N)}\) and \(\chi^*\), it remains to show that

\[ \lim_{N \to \infty} \frac{1}{2} \int_0^\infty \left( \frac{m}{1 + t} - \mathcal{T}^{(N)}(X + t^{1/2}S|Y) \right) dt = \frac{1}{2} \int_0^\infty \left( \frac{m}{1 + t} - \Phi^*(X + t^{1/2}S|Y) \right) dt. \]
We just showed the integrand converges pointwise. But we can take the limit inside the integral by the dominated convergence theorem, because by Lemma 6.3 we have
\[
\frac{m}{a + t} \leq \mathcal{I}^{(N)}(X^{(N)} + t^{1/2}S^{(N)})Y^{(N)} \leq \min\left(\frac{m}{t}, \mathcal{I}^{(N)}(X^{(N)}|Y^{(N)})\right),
\]
and we also know that \(\mathcal{I}^{(N)}(X^{(N)}|Y^{(N)})\) is bounded as \(N \to \infty\) because it converges to \(\Phi(\cdot|\cdot)\).

\[\square\]

Remark 6.7. Of course, (6.10) leads to the same conclusion as Lemma 6.1. Indeed, \(\xi_t = D_{x}V_t^{(N)}(X^{(N)} + t^{1/2}S^{(N)}, Y^{(N)})\) is the conjugate variable for \(X^{(N)} + t^{1/2}S^{(N)}\), and Lemma 6.1 says that \(\xi_t\) is the conditional expectation of \(\xi_0 = D_{x}V^{(N)}(X^{(N)}, Y^{(N)})\) given \(X^{(N)} + t^{1/2}S^{(N)}\) and \(Y^{(N)}\).

Remark 6.8. In [17, §7], we did not use the conditional expectation method to prove \(DV_t^{(N)}\) is asymptotically approximable by trace polynomials, but rather we analyzed the evolution of \(DV_t^{(N)}\) directly using PDE semigroups. The proof given here for convergence of entropy is thus considerably shorter. However, our results on the evolution of \(DV_t^{(N)}\) will come in handy for our construction of transport in the next section.

7. Conditional Transport to Gaussian

In this section, we prove our main results about transport (Theorems 7.10 and 7.13). Suppose that \(V^{(N)}(x, y)\) is a potential as in Assumption 5.1, \(\mu^{(N)}\) is the corresponding probability distribution and that \((X^{(N)}, Y^{(N)})\) is a random variable with this law. Let \(Z^{(N)}\) be an independent \(m\)-tuple of GUE matrices. Let \(\mu_t^{(N)}\) be the law of \((X^{(N)} + t^{1/2}Z^{(N)}, Y^{(N)})\).

The evolution of the potential \(V_t^{(N)}\) corresponding to \(\mu_t^{(N)}\) was studied in [17], and in particular, we established a dimension-independent way to obtain \(DV_t^{(N)}\) from \(DV^{(N)}\) using operations that preserve asymptotic approximability by trace polynomials. By solving an ODE in terms of \(DV_t^{(N)}\), we will obtain transport maps \(F_{s,t}^{(N)} : M_N(C)^{m+n} \to M_N(C)^{m+n}\) such that
\[
(F_{s,t}^{(N)})(X^{(N)} + t^{1/2}Z^{(N)}, Y^{(N)}), Y^{(N)}) \sim (X^{(N)} + s^{1/2}Z^{(N)}, Y^{(N)}).
\]
Upon renormalizing and taking the limit as \(s\) or \(t\) goes to \(\infty\), we obtain transport to the law of \((Z^{(N)}, Y^{(N)})\).

To make each part of the proof more computationally tractable, we proceed in stages. Up until §7.3, we fix \(N\) (and thus suppress it in the notation). First, in §7.4 we describe the basic construction of transport for functions of \(x\) alone (imagining that we have frozen the variable \(y\)). In §7.5 we describe the properties of \(V_t^{(N)}(x, y)\). Next, §7.6 proves Lipschitz estimates for the transport maps \(F_{s,t}^{(N)}(x, y)\).

In §7.7 we introduce renormalized transport maps \(\tilde{F}_{s,t}^{(N)}\) that transport \(\tilde{\mu}_t\) to \(\tilde{\mu}_s\), where \(\tilde{\mu}_t\) is the law of \((e^{-t/2}X^{(N)} + e^{-t/2}(e^t - 1)^{1/2}Z^{(N)}, Y^{(N)})\). The renormalized transport map \(\tilde{F}_{s,t}\) is the same one used by Otto and Villani in their proof of the Talagrand transportation-entropy inequality [24, §4, proof of Lemma 2], in the special case where the target measure is Gaussian (and generalized to the conditional setting). We will explain this inequality further in §8.3.

The new element in our paper is the analysis of the large \(t\) and large \(N\) limits of the transport maps. In §7.4 we show that the limit as \(s\) or \(t\) tends to \(\infty\) exists. Then in §7.5 we use the machinery of asymptotic approximation by trace polynomials to study the large \(N\) limit of \(\tilde{F}_{s,t}^{(N)}\). In order to get dimension-independent estimates for convergence as \(s\) or \(t\) tends to \(\infty\), we conduct a finer analysis of convexity properties of \(V_t\) and Lipschitz properties of \(\tilde{F}_{s,t}\). It is
convenient to carry out the earlier stages of this analysis in \((7.2)\) and \((7.3)\) for \(F_{s,t}\) rather than \(\rho_{s,t}\).

7.1. Basic Construction of Transport. In this section, we will fix \(N\) and fix a function \(V : M_N(\mathbb{C})^m_{sa} \to \mathbb{R}\) in \(\mathcal{E}_{\varphi,C}\) for some \(0 < c < C\). Later, we will allow \(V\) to depend on \(N\) and to depend on another self-adjoint tuple \(y\), but we prefer to simplify notation for the sake of carrying out the basic computation.

Let \(\mu\) be the probability measure with density \((1/Z)e^{-N^2V}\) where \(Z = \int_{M_N(\mathbb{C})^m_{sa}} e^{-N^2V}.\) We showed in \((7.1)\) that the density of \(\mu_t := \mu * \sigma_t, N\) is \((1/Z)e^{-N^2V_t}\), where \(V_t\) solves the equation

\[
\frac{\partial}{\partial t} V_t = \frac{1}{2N} \Delta V_t - \|DV_t\|_2^2.
\]

Because \((1/Z)e^{-N^2V_t}\) solves the heat equation, we know that \(V_t\) is a smooth function of \((x, t)\) for \(t > 0\) and a continuous function of \((x, t)\) for \(t > 0\). Moreover, \(V_t \in \mathcal{E}(c(1 + ct)^{-1}, C(1 + Ct)^{-1})\) for each \(t\) as proved in Theorem 6.1 (1) of \([7.2]\).

Now we can describe explicit transport functions \(F_{s,t}\) such that \((F_{s,t})_* \mu_s = \mu_t\) for all \(s, t \in [0, +\infty)\).

**Proposition 7.1.** Let \(V, \mu, V_t, \) and \(\mu_t\) be as above.

1. There exists a unique family of functions \(F_{s,t} : M_N(\mathbb{C})^m_{sa} \to M_N(\mathbb{C})^m_{sa}\) for \(0 \leq s \leq t < +\infty\) such that

\[
F_{t,t}(x) = x \quad \text{for} \quad t \in [0, +\infty)
\]

\[
\frac{\partial_s}{\partial s} F_{s,t}(x) = \frac{1}{2} \text{DV}_s(F_{s,t}(x)) \quad \text{for} \quad s, t \in [0, +\infty).
\]

2. \(F_{t, t_2} \circ F_{t_2, t_3} = F_{t, t_3}\) and in particular \(F_{t, s} = F_{s, t}^{-1}\).

3. \((F_{s,t})_* \mu_s = \mu_t\).

**Proof.** (1) Because \(V_s \in \mathcal{E}(c(1 + cs)^{-1}, C(t + Cs)^{-1})\), we know that \(DV_s(x)\) is \(C\)-Lipschitz with respect to \(\|\|_2\). Hence, given \(t \in [0, +\infty)\), by the Picard-Lindelöf theorem, the initial value problem \((7.2)\) has a solution defined for all \(s \in [0, +\infty)\).

(2) Fix \(t_1, t_2, t_3\) and \(x \in M_N(\mathbb{C})^m_{sa}\). Let \(G(t)\) be the function defined by \(G(t_3) = x\) and \(\partial_t G(t) = DV_t(G(t))\). By definition of the functions \(F_{s,t}\), we have \(G(t_1) = F_{t_1, t_3}(x)\) and \(G(t_2) = F_{t_2, t_3}(x)\). So \(G\) also satisfies the initial value problem \(\partial_t G(t) = DV_t(G(t))\) and \(G(t_2) = F_{t_2, t_3}(x)\). Therefore, \(G(t_3) = F_{t_1, t_2}(F_{t_2, t_3}(x))\), so that \(F_{t_1, t_2}(F_{t_2, t_3}(x)) = F_{t_1, t_3}(x)\).

(3) We first prove the claim for \(s, t > 0\). Because \(V_s\) is smooth, it follows that \(F_{s,t}\) is smooth for \(s, t > 0\) by standard theory of smooth dependence for ODE. Let \(JF_{s,t}\) denote the Jacobian linear transformation (differential) of \(F_{s,t}\). Let \(\rho_t = (1/Z)e^{-V_t}\) be the density of \(\mu_t\). As a consequence of the change-of-variables formula for multivariable integrals, we see that the density of \((F_{s,t})_* \mu_s = (F_{s,t})_* \mu_t\) is

\[
(\rho_t \circ F_{s,t})|\text{det } JF_{s,t}| = \frac{1}{Z} \exp \left[ -N^2 \left( V_t \circ F_{s,t} - \frac{1}{N^2} \log |\text{det } JF_{s,t}| \right) \right].
\]

Fix \(s\). If \(t = s\), then clearly this reduces to \(\rho_s\). Therefore, it suffices to show that \((\rho_t \circ F_{s,t})|\text{det } JF_{t,s}|\) is a constant function of \(t\), or equivalently that

\[
\frac{\partial_t}{\partial t} \left[ V_t \circ F_{s,t} - \frac{1}{N^2} \log |\text{det } JF_{t,s}| \right] = 0.
\]
Recalling smoothness \( V_t \) and \( F_{t,s} \) for \( s, t > 0 \) and using the differential equations (7.1) for \( V_t \) and (7.2) for \( F_{t,s} \), we obtain

\[
\partial_t[V_t \circ F_{t,s}] = (\partial_t V_t) \circ F_{t,s} + (DV_t(F_{t,s}), \partial_s F_{t,s})_2
\]

\[
= \frac{1}{2N} \Delta V_t \circ F_{t,s} - \frac{1}{2} \|DV_t(F_{t,s})\|^2 + \frac{1}{2} \langle DV_t(F_{t,s}), DV_t(F_{t,s}) \rangle_2
\]

\[
= \frac{1}{2N} \Delta V_t.
\]

Meanwhile, to compute \( \partial_t \log |\det F_{t,s}| \), note that for small \( \epsilon \in \mathbb{R} \),

\[
JF_{t+\epsilon,s} = J(F_{t+\epsilon,t} \circ F_{t,s}) = JF_{t+\epsilon,t}(F_{t,s})JF_{t,s},
\]

so that

\[
\log |\det JF_{t+\epsilon,s}| = \log |\det JF_{t,s}| + \log |\det JF_{t+\epsilon,t}(F_{t,s})|.
\]

Using smoothness,

\[
\frac{d}{d\epsilon}|_{\epsilon=0} JF_{t+\epsilon,t} = J \left( \frac{d}{d\epsilon}|_{\epsilon=0} F_{t+\epsilon,t} \right)
\]

\[
= \frac{1}{2} J(DV_t) = \frac{1}{2} NJ(\nabla V_t),
\]

Since \( JF_{t+\epsilon,t} \) becomes the identity when \( \epsilon = 0 \), we know that for small enough \( \epsilon \), the linear transformation \( JF_{t+\epsilon,t} \) has positive determinant and \( \log JF_{t+\epsilon,t} \) is well-defined by power series, so that

\[
\frac{d}{d\epsilon}|_{\epsilon=0} \log |\det JF_{t+\epsilon,t}| = \frac{d}{d\epsilon}|_{\epsilon=0} \log |\det JF_{t+\epsilon,t}|
\]

\[
= \text{Tr} \left( \frac{d}{d\epsilon}|_{\epsilon=0} JF_{t+\epsilon,t} \right)
\]

\[
= \text{Tr}(N(J\nabla V_t)) = \frac{1}{2} N \Delta V_t.
\]

Hence, \( \partial_t \log |\det JF_{t,s}| = \frac{N}{2} \Delta V_t \circ F_{t,s} \). This implies that

\[
\partial_t \left[ V_t \circ F_{t,s} - \frac{1}{N^2} \log |\det JF_{t,s}| \right] = \frac{1}{2N} \Delta V_t - \frac{1}{N^2} \cdot \frac{N}{2} \Delta V_t \circ F_{t,s} = 0,
\]

completing the proof of the claim for \( s, t > 0 \). The equality \((F_{s,t})_\ast \mu_t = \mu_s\) extends to the case where \( s \) or \( t \) is zero because both sides depend continuously on \( s \) and \( t \) with respect to the weak topology on measures. \( \square \)

In particular, the map \( F_{0,t} \) transports \( \mu_t = \mu \ast \sigma_{1,N} \) to the original law \( \mu \). In other words, if \( X \sim \mu \) and \( Z \sim \sigma_{1,N} \), then \( F_{0,t}(X + t^{1/2}Z) \sim X \) and \( F_{t,0}(X) \sim X + t^{1/2}Z \). This implies that \((1 + t)^{-1/2} F_{t,0}(X) \sim (1 + t)^{-1/2}(X + t^{1/2}Z)\). This suggests that we can find a transport map from the law of \( X \) to the law of \( Z \) as the large t limit of \((1 + t)^{-1/2} F_{t,0}\). In the interest of efficiency, we postpone the details of this argument until after we introduce the dependence on the other set of parameters \( y \).

### 7.2. Conditional Hamilton-Jacobi-Bellman Equation and Semigroups.

Let us now fix \( N \) and fix a potential \( V : M_N(\mathbb{C})^{sa}_+ \times M_N(\mathbb{C})^{sa}_+ \rightarrow \mathbb{R} \) in \( \mathcal{E}_{c_1,c_2}^N(\mu, \nu) \) for some \( 0 \leq c \leq C \). Let \( \mu \) be the corresponding law and let \((X, Y)\) be a random variable in \( M_N(\mathbb{C})^{sa}_+ \times M_N(\mathbb{C})^{sa}_+ \) distributed according to \( \mu \). Let \( \mu_t \) be the law of \((X + t^{1/2}Z, Y)\), where \( Z \) is an independent tuple of independent GUE.

Our goal is to transport the law \( \mu_s \) to the law \( \mu_t \). Upon freezing the variable \( y \), the methods of the previous section will produce a transport map \( F_{s,t}(x, y) \) such that \( F_{s,t}(\cdot, y) \) pushes forward
the conditional distribution of $X + s^{1/2}Z$ given $Y$ to the conditional distribution of $X + t^{1/2}Z$

given $Y$. Specifically, $F_{s,t} : M_N(C)^m \times M_N(C)^n \to M_N(C)^{m+n}$ is the solution to the initial value problem

$$F_{t,t}(x, y) = x,$$
$$\partial_x F_{s,t}(x, y) = D_x V_s(F_{s,t}(x, y), y).$$

Then $(F_{s,t}(X + t^{1/2}Z, Y), Y) \sim (X + s^{1/2}Z, Y)$.

We seek to understand the large $t$ and large $N$ behavior of $F_{s,t}(x, y)$ as a function of $(x, y)$ rather than simply as a function of $x$ for a fixed $y$. To achieve this, we must understand the behavior of $V(x, y)$ and $D_x V(x, y)$ as functions of $(x, y)$. We will first import the results of [17, §6] regarding $D_x V(x, y)$ as a function of $x$, then we will extend them to handle the dependence on $y$.

The potential $V_t$ satisfies

$$(7.3) \quad \partial_t V_t = \frac{1}{2N} \Delta_x V_t - \frac{1}{2} \|D_x V_t\|^2.$$  

We express $V_t = R_t V$, where $R_t$ is a semigroup acting on convex and semiconcave functions defined as follows. Let

$$P_t u(x, y) = \int_{M_N(C)^m} u(x + z, y) d\sigma_{m,t}^{(N)}(z)$$
$$Q_t u(x, y) = \inf_z \left[ u(z, y) + \frac{1}{2t} \|z - x\|^2 \right].$$

Then as suggested by Trotter’s formula, we want to express $R_t u = \lim_{n \to \infty}(P_{t/n} Q_{t/n})^n u$, but for technical convenience we only apply this dyadic rational $t$ and values of $n$ that are powers of 2. The following is a direct application of [17, Theorems 6.1 and 6.17] to $V(\cdot, y)$.

**Theorem 7.2.** There exists a semigroup of nonlinear operators $R_t : \bigcup_{C > 0} \mathcal{E}_m^{(N)}(0, C) \to \bigcup_{C > 0} \mathcal{E}_m^{(N)}(0, C)$ with the following properties:

1. **Change in Convexity:** If $u(\cdot, y) \in \mathcal{E}_m^{(N)}(c, C)$, then $R_t u(\cdot, y) \in \mathcal{E}_m^{(N)}(c(1 + Ct)^{-1}, C(1 + Ct)^{-1})$.

2. **Approximation by Iteration:** For $\ell \in \mathbb{Z}$ and $t \in 2^{-\ell} \mathbb{N}_0$, denote $R_{t, \ell} u = (P_{2^{-\ell}} Q_{2^{-\ell}})^{2^\ell} u$.

Suppose $t \in \mathbb{Q}^+_{2}$ and $u \in \mathcal{E}_m^{(N)}(0, C)$.

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

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

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

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

(a) $R_{t, u} \leq R_t u + \frac{3}{2} \log(1 + Ct) - \log(1 + C s)$.

(b) $R_{t, u} \geq R_t u - \frac{3}{2} \left( \log(1 + Ct) + \|D_x u\|^2_2 \right)$.

(c) If $C(t - s) \leq 1$, then $\|D_x(R_{t, u}) - D_x(R_t u)\|_2 \leq 5 C m_{1/2}^1(2(t - s)^{1/2} + C(t - s)) \|D_x u\|_2$.

4. **Differential Equation:** $R_t u(\cdot, t)$ is smooth as a function of $(x, t)$, it satisfies (7.3), and we have $P_t[\exp(-N^2 u)] = \exp(-N^2 R_t u)$.

Result (1) regarding convexity and semiconcavity only applies to $R_t u$ as a function of $x$ for a fixed $y$. We now extend this result to control the dependence on $y$, using the same techniques as in [17] Lemma 6.6. As remarked in that paper, this type of analysis of $Q_t$ is standard in the PDE literature.
Lemma 7.3. Suppose that \( u : M_N(\mathbb{C})_{sa}^n \times M_N(\mathbb{C})_{sa}^n \to \mathbb{R} \) and that
\[
cI_m + c'I_n \leq Hu \leq CI_m + C'I_n
\]
in the sense of Definition 2.1. Then
1. \( cI_m + c'I_n \leq H(P_t u) \leq CI_m + C'I_n \).
2. \( D(Q_t u)(x, y) = Du(x - tD_x(Q_t u)(x, y), y) \).
3. \( c(1 + ct)^{-1}I_m + c'I_n \leq H(Q_t u) \leq C(1 + Ct)^{-1}I_m + C'I_n \).
4. \( c(1 + ct)^{-1}I_m + c'I_n \leq H(R_t u) \leq C(1 + Ct)^{-1}I_m + C'I_n \).

Proof. (1) This is left as an exercise.

(2) The proof is a modification of that of [17, Lemma 6.6], which proves an analogous result in the simpler case of functions of \( x \) without the extra variable \( y \). Fix \( x_0 \) and \( y_0 \). Because the function \( u(z, y_0) + \frac{1}{2}\|z - x_0\|_2^2 \) is uniformly convex with respect to \( z \), it has a unique minimizer \( z_0 \). This minimizer must be a critical point with respect to the first variable, and hence
\[
D_x u(z_0, y_0) + \frac{1}{t}(z_0 - x_0),
\]
that is,
\[
z_0 = x_0 - tD_x u(z_0, y_0).
\]
Let \( p = D_x u(z_0, y_0) \) and \( q = D_y u(z_0, y_0) \), so that \( Du(z_0, y_0) = (p, q) \). Our assumption \( cI_m \oplus c'I_n \leq Hu \leq CI_m \oplus C'I_n \) implies that
\[
u(x, y) \leq u(x, y) \leq \nu(x, y),
\]
where
\[
\nu(x, y) = u(z_0, y_0) + \langle p, x - z_0 \rangle + \langle q, y - y_0 \rangle + \frac{c}{2}\|x - z_0\|_2^2 + \frac{c'}{2}\|y - y_0\|_2^2
\]
\[
\nu(x, y) = u(z_0, y_0) + \langle p, x - z_0 \rangle + \langle q, y - y_0 \rangle + \frac{C}{2}\|x - z_0\|_2^2 + \frac{C'}{2}\|y - y_0\|_2^2.
\]
Note that \( \nu \leq u \leq \nu \) implies \( Q_{t\nu} \leq Q_t u \leq Q_{t\nu} \) since monotonicity of \( Q_t \) is immediate from the definition. One can compute \( Q_{t\nu} \) and \( Q_{t\nu} \) directly as in Lemma 6.4 (2) and the proof of Lemma 6.6 in [17] and obtain
\[
Q_{t\nu}(x, y) = u(z_0, y_0) - \frac{t}{2}\|p\|_2^2 + \langle p, x - z_0 \rangle + \langle q, y - y_0 \rangle + \frac{c}{2(1 + ct)}\|x - tp - z_0\|_2^2 + \frac{c'}{2}\|y - y_0\|_2^2
\]
\[
= u(z_0, y_0) + \frac{1}{2t}\|z_0 - x_0\|_2^2 + \langle p, x - x_0 \rangle + \langle q, y - y_0 \rangle + \frac{c}{2(1 + ct)}\|x - tp - z_0\|_2^2 + \frac{c'}{2}\|y - y_0\|_2^2
\]
\[
= Q_t u(x_0, y_0) + \langle p, x - x_0 \rangle + \langle q, y - y_0 \rangle + \frac{c}{2(1 + ct)}\|x - x_0\|_2^2 + \frac{c'}{2}\|y - y_0\|_2^2,
\]
where the last two lines following from substituting \( z_0 = x_0 - tp \) and that the infimum defining \( Q_t u \) is achieved at \( z_0 \). The analogous formula for \( Q_{t\nu}(x, t) \) holds as well. The functions \( Q_{t\nu} \) and \( Q_{t\nu} \) thus provide second-order Taylor expansions from above and below for the function \( Q_t u \) with respect to \( (x, y) \) at the point \( (x_0, y_0) \). Looking at the first-order terms in the expansions shows that \( Q_t u \) is differentiable at \( (x_0, y_0) \) with
\[
D(Q_t u)(x_0, y_0) = (p, q) = Du(z_0, x_0) = Du(x_0 - tp, y_0) = Du(x_0 - tD_x(Q_t u)(x_0, y_0), y_0),
\]
which proves (2).

(3) We examine the second-order terms of upper and lower Taylor expansions \( Q_{t\nu} \) and \( Q_{t\nu} \) and apply the claim \( (2) \implies (1) \) from Lemma 2.2. This is the same argument as in the proof of [17, Proposition 2.13 (2)].
(4) Recall that if \( A \leq Hu \leq B \), then \( A \leq H(P_tu) \leq B \). Using this fact together with (3) iteratively, we see that if \( t \) is a dyadic rational and \( t = 2^{-\ell}k \), then 
\[
\begin{align*}
c(1 + ct)^{-1}I_m \oplus c'I_n \leq H(R_{t,x}u) \leq C(1 + Ct)^{-1}I_m \oplus C'I_n.
\end{align*}
\]
In light of Theorem 7.2 (2), this will also hold in the limit as \( \ell \to \infty \), since for any two self-adjoint matrices \( A \) and \( B \), the family of functions with \( A \leq Hu \leq B \) is closed under pointwise limits. Similarly, using Theorem 7.2 (3), we extend this to all real \( t \geq 0 \). \( \square \)

Remark 7.4. The convexity conditions of Lemma 7.3 (4) can alternatively be deduced from [7, Theorem 4.3]. However, it is convenient for us to use Theorem 7.2 here because we want the dimension-independent time-continuity estimates Theorem 7.2 (3) in the proof of Theorem 7.11 below.

7.3. Lipschitz Estimates for Conditional Transport. This subsection proves the technical estimate Lemma 7.5 on the Lipschitz seminorm of \( F_{x,t} \). This depends crucially on the convexity properties of \( V_{t}(x, y) \).

Lemma 7.5.
\[
\begin{align*}
\langle D_xV_{t}(x, y) - D_xV_{t}(x', y'), x - x' \rangle_2 &\leq \frac{C}{1 + Ct} \|x - x'\|_2^2 + \frac{C - c}{(1 + Ct)^{1/2}(1 + ct)^{1/2}} \|x - x'\|_2 \|y - y'\|_2 \\
\langle D_xV_{t}(x, y) - D_xV_{t}(x', y'), x - x' \rangle_2 &\geq \frac{c}{1 + ct} \|x - x'\|_2^2 - \frac{C - c}{(1 + Ct)^{1/2}(1 + ct)^{1/2}} \|x - x'\|_2 \|y - y'\|_2.
\end{align*}
\]

Proof. First, note that
\[
\begin{align*}
\langle D_xV_{t}(x, y) - D_xV_{t}(x', y'), x - x' \rangle_2 &= \langle D_xV_{t}(x, y) - D_xV_{t}(x', y), x - x' \rangle_2 + \langle D_xV_{t}(x', y') - D_xV_{t}(x', y), x - x' \rangle_2.
\end{align*}
\]
By Lemma 2.2, the first term on the right hand side of (7.4) can be estimated by
\[
\frac{c}{1 + ct} \|x - x'\|_2^2 \leq \langle D_xV_{t}(x, y) - D_xV_{t}(x', y), x - x' \rangle_2 \leq \frac{C}{1 + Ct} \|x - x'\|_2^2.
\]
To handle the second term on the right hand side of (7.4), define
\[
\begin{align*}
\nabla_i(x, y) &= V_{t}(x, y) - \frac{c}{2(1 + ct)} \|x\|_2^2 - \frac{c}{2} \|y\|_2^2 \\
\nabla_j(x, y) &= V_{t}(x, y) - \frac{C}{2(1 + Ct)} \|x\|_2^2 - \frac{C}{2} \|y\|_2^2
\end{align*}
\]
and recall that \( \nabla_i \) is convex and \( \nabla_j \) is concave and in particular
\[
0 \leq H\nabla_i \leq \left( \frac{C}{1 + Ct} - \frac{c}{1 + ct} \right) I_m \oplus (C - c)I_n = \frac{C - c}{(1 + Ct)(1 + ct)} I_m \oplus (C - c)I_n.
\]
Note that
\[
\begin{align*}
D_xV_{t}(x', y) - D_xV_{t}(x', y') &= \left( D_xV_{t}(x', y) - \frac{c}{1 + ct} x' \right) - \left( D_xV_{t}(x', y) - \frac{c}{1 + ct} x' \right) \\
&= D_x\nabla_i(x', y) - D_x\nabla_i(x', y').
\end{align*}
\]
Therefore,
\[
\begin{align*}
\langle D_xV_{t}(x', y) - D_xV_{t}(x', y'), x - x' \rangle_2 &= \langle D_x\nabla_i(x', y) - D_x\nabla_i(x', y'), x - x' \rangle_2 \\
&= \langle D\nabla_i(x', y) - D\nabla_i(x', y'), (x - x', 0) \rangle_2.
\end{align*}
\]
Now we apply Lemma 7.3 to $\nabla_t$ with the matrix $A = \frac{C-c}{(1+Ct)(1+ct)}I_n \oplus (C-c)I_n$ and conclude that

$$
||D\nabla_t(x', y) - D\nabla_t(x', y'), (x - x', 0)||_2 \leq \left( (C-c)\|y - y'\|_2^2 \right)^{1/2} \left( \frac{C-c}{(1+Ct)(1+ct)} \|x - x'\|_2^2 \right)^{1/2}
$$

$$
\leq \frac{C-c}{(1+Ct)^{1/2}(1+ct)^{1/2}} \|x - x'\| \|y - y'\|.
$$

Combining this estimate for the second term of (7.4) with our earlier estimate for the first term completes the proof. \hfill \square

**Lemma 7.6.** We have

$$
\|F_{s,t}\|_{Lip, dx} \leq \begin{cases} 
\frac{(1+Cs)^{1/2}}{(1+Ct)^{1/2}} \frac{1}{(1+cs)^{1/2}} & s \geq t \\
\frac{(1+Cs)^{1/2}}{(1+Ct)^{1/2}} & s < t.
\end{cases}
$$

and

$$
\|F_{s,t}\|_{Lip, dy} \leq \begin{cases} 
\frac{(C/c - 1)(1+Cs)^{1/2}}{(1+Ct)^{1/2}} \left( \frac{1}{(1+ct)^{1/2}} - \frac{1}{(1+cs)^{1/2}} \right), & s \geq t, \\
\frac{(C/c - 1)(1+cs)^{1/2}}{(1+Ct)^{1/2}} \left( \frac{1}{(1+cs)^{1/2}} - \frac{1}{(1+Ct)^{1/2}} \right), & s < t.
\end{cases}
$$

**Proof.** Fix $t \geq 0$ and $(x, y)$ and $(x', y')$ in $M_N(C)^m \times M_N(C)^n$ and define

$$
\phi(s) = \|F_{s,t}(x, y) - F_{s,t}(x', y')\|_2.
$$

Note that $\phi$ is locally Lipschitz, hence absolutely continuous. Also,

$$
2\phi(s)\phi'(s) = \partial_s[\phi(s)^2] = 2(\partial_s F_{s,t}(x, y) - \partial_s F_{s,t}(x', y'), F_{s,t}(x, y) - F_{s,t}(x', y'))_2
$$

$$
= (D\phi_{s,t}(x, y), y - F_{s,t}(x', y'), F_{s,t}(x, y) - F_{s,t}(x', y'))_2
$$

$$
\leq \frac{C}{1+Ct} \|F_{s,t}(x, y) - F_{s,t}(x', y')\|_2^2
$$

$$
+ \frac{C-c}{(1+Ct)^{1/2}(1+ct)^{1/2}} \|F_{s,t}(x, y) - F_{s,t}(x', y')\|_2 \|y - y'\|_2
$$

$$
= \frac{C}{1+Cs} \phi(s)^2 + \frac{C-c}{(1+Cs)^{1/2}(1+cs)^{1/2}} \phi(s) \|y - y'\|_2^2,
$$

where we have applied Lemma 7.3. It follows that whenever $\phi(s) > 0$,

$$
\phi'(s) \leq \frac{C}{2(1+Cs)} \phi(s) + \frac{C-c}{2(1+Cs)^{1/2}(1+cs)^{1/2}} \|y - y'\|_2.
$$

On the other hand, since $\phi(s) \geq 0$, any point where $\phi$ is zero and $\phi$ is differentiable must be a critical point, so when $\phi(s) = 0$ the estimate is vacuously true. This inequality implies

$$
\frac{d}{ds} \left[ \frac{1}{(1+Cs)^{1/2}} \phi(s) \right] \leq \frac{C-c}{2(1+Cs)(1+cs)^{1/2}} \|y - y'\|_2
$$

$$
\leq \frac{C(C-c)}{2c(1+Cs)^{3/2}} \|y - y'\|_2,
$$

where in the last line we have observed that $(1+cs)^{1/2} \geq (c/C)^{1/2}(1+Cs)^{1/2} \geq (c/C)(1+Cs)^{1/2}$. Hence for $s \geq t$

$$
\frac{1}{(1+Cs)^{1/2}} \phi(s) - \frac{1}{(1+Ct)^{1/2}} \phi(t) \leq \frac{C-c}{c} \left( \frac{1}{(1+Ct)^{1/2}} - \frac{1}{(1+Cs)^{1/2}} \right) \|y - y'\|_2
$$
A brief computation shows that the corresponding potential is (here the potential is only well-defined up to an additive constant because the probability measure \( \mu \))

\[
\phi(t) = \| x - x' \|_2 \quad \text{and} \quad \phi(t) = \| x - x' \|_2.
\]

This is proves the asserted estimates in the case where \( s \geq t \). The argument for the case \( s \leq t \) is similar. Here we use the lower bound rather than the upper bound in Lemma 7.5 and get

\[
\phi'(s) \geq \frac{C - \rho}{2(1 + cs)} \phi(s) - \rho \frac{C - \rho}{2(1 + cs)^{1/2}} \| y - y' \|_2
\]

so that

\[
\frac{d}{ds} \left[ \frac{1}{(1 + cs)^{1/2}} \phi(s) \right] \geq -\frac{C - \rho}{2c(1 + cs)^{3/2}} \| y - y' \|_2.
\]

Now we take \( s \leq t \) and obtain

\[
\frac{1}{(1 + ct)^{1/2}} \| x - x' \|_2 - \frac{1}{(1 + cs)^{1/2}} \| F_{s,t}(x, y) - F_{s,t}(x', y') \|_2 \geq -\frac{C - \rho}{c} \left( \frac{1}{(1 + cs)^{1/2}} - \frac{1}{(1 + ct)^{1/2}} \right) \| y - y' \|_2,
\]

which yields the desired estimates. \( \square \)

7.4. Transport in the Large \( t \) Limit. We remind the reader here that we are still working in the finite-dimensional setting for a fixed values of \( N \) which is suppressed in the notation. To understand the large \( t \) limit of our transport maps, consider the renormalized law

\[
\tilde{\mu}_t := \text{law of } (\tilde{X}_t, Y) := \left( e^{-t/2} X + (1 - e^{-t})^{1/2} Z, Y \right).
\]

A brief computation shows that the corresponding potential is

\[
\tilde{V}_t(x, y) := V_{t-1}(e^{t/2} x, y),
\]

(here the potential is only well-defined up to an additive constant because the probability measure \( \tilde{\mu}_t \) includes a normalizing constant \( 1/\tilde{Z}_t \) anyway, so we made a convenient choice of the additive constant). This potential satisfies the equation

\[
\partial_t \tilde{V}_t = \frac{1}{2N} \Delta_x \tilde{V}_t - \frac{1}{2} \left\| D_x \tilde{V}_t \right\|_2^2 + \frac{1}{2} \left\langle D_x \tilde{V}_t, x \right\rangle.
\]

We remark that if \( \tilde{\rho}_t = (1/\tilde{Z}_t) e^{-N^2 \tilde{V}_t} \) is the density at time \( t \) and \( r(x, y) = \text{const} e^{-\| (x, y) \|_2^2} \) is the Gaussian density, then

\[
\partial_t \tilde{\rho}_t = \frac{1}{2N} \Delta_x \tilde{\rho}_t + \frac{1}{2} \left\langle \nabla_x \tilde{\rho}_t, x \right\rangle_{T_N} + \frac{N m}{2} \tilde{\rho}_t \frac{\tilde{\rho}_t}{r}
\]

In other words, \( \tilde{\rho}_t \) evolves according to the diffusion semigroup with respect to Gaussian measure (compare equation (33) of [24]), while the heat equation represents diffusion with respect to Lebesgue measure.

The transport functions are renormalized as follows. Because \( (F_{s,t}(x, y), y) \) pushes forward \( \mu_t \) to \( \mu_s \), we may compute that \( \tilde{F}_{s,t}(x, y) \) pushes forward \( \tilde{\mu}_t = \mu_s \), where

\[
\tilde{F}_{s,t}(x, y) := e^{-s/2} F_{e^{-s} t, e^{-s} t} (e^{s/2} x, y).
\]
Moreover, from the differential equation (7.2), we deduce that
\[
\partial_s \tilde{F}_{s,t}(x, y) = \frac{1}{2} \left( D_x \tilde{V}_{s,t}(x, y) - \tilde{F}_{s,t}(x, y) \right).
\]

As \( t \to \infty \), the law \( \tilde{\mu}_t \) converges to the law of \((Z, Y)\), which we denote \( \tilde{\mu}_\infty \). Thus, if we show that \( \tilde{F}_{s,t} \) has a limit as \( s \to +\infty \) or \( t \to +\infty \), we will be able to transport our given law \( \mu = \tilde{\mu}_0 \) of \((X, Y)\) to the law of \((Z, Y)\). As the first step, we deduce from Lemma 7.6 the following Lipschitz estimates on \( \tilde{F}_{s,t} \) which are uniform in \( s \) and \( t \). Note also that the coefficient of \( \| y - y' \|_2 \) goes to zero as \( s, t \to \infty \).

**Lemma 7.7.** We have
\[
\left\| \tilde{F}_{s,t} \right\|_{\text{Lip}, dx} \leq \max(C, 1/c)^{1/2}
\]
and
\[
\left\| \tilde{F}_{s,t} \right\|_{\text{Lip}, dy} \leq (C/c - 1) \max(C, 1/C)^{3/2} |e^{-s/2} - e^{-t/2}|.
\]
In particular,
\[
\left\| \tilde{F}_{s,t} \right\|_{\text{Lip}} \leq \max(C, 1/c)^{7/2}.
\]

**Proof.** For the first estimate, for the case where \( s \geq t \), direct substitution of (7.9) into (7.5) of Lemma 7.6 shows that
\[
\left\| \tilde{F}_{s,t} \right\|_{\text{Lip}, dx} \leq e^{-s/2} \frac{(1 + C(e^s - 1))^{1/2}}{(1 + C(e^t - 1))^{1/2}} e^{t/2} = \frac{(C + (1 - C)e^{-s})^{1/2}}{(C + (1 - C)e^{-t})^{1/2}}.
\]
The function \( C + (1 - C)e^{-s} \) is clearly monotone on \([0, +\infty)\) and achieves the values 1 and \( C \) at 0 and \( +\infty \) respectively, and hence is between \( \min(1, C) \) and \( \max(1, C) \). Hence,
\[
\left\| \tilde{F}_{s,t} \right\|_{\text{Lip}, dx} \leq \frac{\max(C, 1/C)^{1/2}}{\min(1, C)^{1/2}} = \max(C, 1/C)^{1/2} \leq \max(C, 1/c)^{1/2}.
\]
The case where \( s \leq t \) follows by the same argument, where the bound this time is \( \max(c, 1/c)^{1/2} \leq \max(C, 1/c)^{1/2} \).

For the second estimate, we apply (7.8). Note in (7.4), in the case \( s \leq t \), we may use \( (1 + cs)^{1/2} \leq (1 + Cs)^{1/2} \) and thus in both cases \( s \geq t \) or \( s \leq t \),
\[
\left\| F_{s,t} \right\|_{\text{Lip}, dy} \leq (C/c - 1)(1 + Cs)^{1/2} \left\| \int_s^t \frac{C}{2(1 + Cu)^{3/2}} \, du \right\| - \frac{1}{(1 + Ct)^{1/2}}
\]
\[= (C/c - 1)(1 + Cs)^{1/2} \left| \int_s^t \frac{C}{2(1 + Cu)^{3/2}} \, du \right|.
\]
This implies that
\[
\left\| \tilde{F}_{s,t} \right\|_{\text{Lip}, dy} \leq e^{-s/2} (C/c - 1)(1 + C(e^s - 1))^{1/2} \left| \int_s^t \frac{C}{2(1 + Cu)^{3/2}} \, du \right|
\]
\[= (C/c - 1)e^{-s/2} (1 + C(e^s - 1))^{1/2} \left| \int_s^t \frac{Ce^w}{2(1 + C(e^w - 1))^{3/2}} \, dw \right|
\]
\[\leq (C/c - 1) \max(1, C)^{1/2} \left| \int_s^t \frac{Ce^w}{2 \min(1, C)^{3/2} e^{3w/2}} \, dw \right|
\]
\[\leq (C/c - 1) \max(1, C)^{1/2} C \min(1, C)^{3/2} |e^{-t/2} - e^{-s/2}|
\]
\[\leq (C/c - 1) \max(C, 1/C)^{3/2} |e^{-t/2} - e^{-s/2}|.
\]
where we have again applied $\min(1, C)e^s \leq 1 + C(e^s - 1) \leq \max(1, C)e^s$.

For the last estimate (7.13), observe that
\[
\left\| \tilde{F}_{s,t} \right\|_{\text{Lip}} \leq \left\| \tilde{F}_{s,t} \right\|_{\text{Lip,} dx} + \left\| \tilde{F}_{s,t} \right\|_{\text{Lip,} dy}
\leq \max(C, 1/c)^{1/2} + (C/c - 1) \max(C, 1/C)^{3/2} |e^{-s/2} - e^{-t/2}|
\leq \max(C, 1/c)^{3/2} + (\max(C, 1/c)^2 - 1) \max(C, 1/c)^{3/2}
= \max(C, 1/c)^{7/2}.
\]

**Lemma 7.8.** Let $\pi_1$ denote the function $\pi_1(x, y) = x$. Then
\[
\left\| \tilde{F}_{s,t} - \pi_1 \right\|_{\text{Lip,} dx} \leq \frac{1}{2} \max(C, 1/c) - 1 \max(C, 1/C)^{1/2} |e^{-s} - e^{-t}|
\]
and
\[
\left\| \tilde{F}_{s,t} - \pi_1 \right\|_{\text{Lip}} \leq \max(C, 1/c)^3 - 1 \max(C, 1/c)^{1/2} |e^{-s/2} - e^{-t/2}|.
\]

**Proof.** Let $U_s(x, y) = \tilde{V}_s(x, y) - (1/2)\|x\|_2^2$. Then (7.10) says that
\[
\partial_s \tilde{F}_{s,t}(x, y) = \frac{1}{2} D_x U_s(\tilde{F}_{s,t}(x, y), y).
\]

Moreover, we have
\[
\frac{ce^s}{1 + c(e^s - 1)} \leq H_x \tilde{V}_s \leq \frac{C e^s}{1 + C(e^s - 1)}.
\]

We can bound $H_x U_s$ above and below by subtracting 1 from both sides, which after some computation reduces to
\[
\frac{c - 1}{1 + c(e^s - 1)} \leq H_x U_s \leq \frac{C - 1}{1 + C(e^s - 1)}.
\]

Therefore, we have $-L \leq H_x U_s \leq L$ where
\[
L = \max \left( \frac{1 - c}{1 + c(e^s - 1)}, \frac{C - 1}{1 + C(e^s - 1)} \right)
\]

We claim that $L \leq (\max(C, 1/c) - 1)e^{-s}$. If the first term $(1 - c)/(1 + c(e^s - 1))$ is negative, then it is $\leq L$ automatically, but if it is positive, then $c \leq 1$ and hence
\[
\frac{1 - c}{1 + c(e^s - 1)} \leq \frac{1 - c}{c + c(e^s - 1)} = (1/c - 1)e^{-s} \leq (\max(C, 1/c) - 1)e^{-s}.
\]

Similarly, if $(C - 1)/(1 + C(e^s - 1))$ is negative, there is nothing to prove, but otherwise $C \geq 1$, and hence
\[
\frac{C - 1}{1 + C(e^s - 1)} \leq \frac{C - 1}{1 + (e^s - 1)} = (C - 1)e^{-s} \leq (\max(C, 1/c) - 1)e^{-s}.
\]

But $-L \leq H_x U_s \leq L$ implies that $D_x U_s$ is $L$-Lipschitz in $x$. Therefore,
\[
\left\| \partial_x \tilde{F}_{s,t}(x, y) - \partial_x \tilde{F}_{s,t}(x', y) \right\|_2 \leq \frac{1}{2} \left\| D_x U_s(\tilde{F}_{s,t}(x, y), y) - D_x U_s(\tilde{F}_{s,t}(x', y), y) \right\|_2
\leq \frac{1}{2} (\max(C, 1/c) - 1)e^{-s} \left\| \tilde{F}_{s,t}(x, y) - \tilde{F}_{s,t}(x', y) \right\|_2.
\]

Applying (7.11) in the case where $s \geq t$, we get
\[
\left\| \partial_x \tilde{F}_{s,t}(x, y) - \partial_x \tilde{F}_{s,t}(x', y) \right\|_2 \leq \frac{1}{2} (\max(C, 1/c) - 1) \max(C, 1/c)^{1/2} e^{-s} \|x - x'\|_2.
\]
Hence,
\[
\left\| \tilde{F}_{s,t}(x,y) - \tilde{F}_{s,t}(x',y) - (x-x') \right\|_2 \leq \left\| \int_t^s \left[ \partial_u \tilde{F}_{u,t}(x,y) - \partial_u \tilde{F}_{u,t}(x',y) \right] du \right\|_2 
\leq \frac{1}{2} (\max(C,1/c) - 1) \max(C,1/c)^{1/2} |e^{-s} - e^{-t}| \|x-x'\|_2.
\]
which proves the desired estimate (7.14).

To check the second estimate (7.15), first observe
\[
\frac{1}{2} |e^{-s} - e^{-t}| = \int_{\min(s,t)}^{\max(s,t)} \frac{1}{2} e^{-u} du \leq \int_{\min(s,t)}^{\max(s,t)} \frac{1}{2} e^{-u/2} du = |e^{-s/2} - e^{-t/2}|,
\]
Moreover, \( \| \tilde{F}_{s,t} - \pi_1 \|_{\text{Lip},dy} = \| \tilde{F}_{s,t} \|_{\text{Lip},dy} \). Therefore, using (7.12) and (7.14),
\[
\| \tilde{F}_{s,t} - \pi_1 \|_{\text{Lip}} \leq \| \tilde{F}_{s,t} \|_{\text{Lip},dx} + \| \tilde{F}_{s,t} - \pi_1 \|_{\text{Lip},dy} 
\leq (\max(C,1/c) - 1) \max(C,1/c)^{1/2} \frac{1}{2} |e^{-s} - e^{-t}| + (C/c - 1) \max(C,1/c)^{3/2} |e^{-s/2} - e^{-t/2}| 
\leq (\max(C,1/c) - 1) \max(C,1/c)^{1/2} + (\max(C,1/c) - 1) \max(C,1/c)^{3/2} |e^{-s/2} - e^{-t/2}| 
= (\max(C,1/c)^3 - 1) \max(C,1/c)^{1/2} |e^{-s/2} - e^{-t/2}|.
\]

**Proposition 7.9.** The limits \( \tilde{F}_{s,\infty} := \lim_{t \to \infty} \tilde{F}_{s,t} \) and \( \tilde{F}_{\infty,t} = \lim_{s \to \infty} \tilde{F}_{s,t} \) exist for \( s,t \geq 0 \).

More precisely, let \( (X,Y) \) and \( (\tilde{X}_t, Y) \) be a pair of random variables with the laws \( \mu \) and \( \tilde{\mu}_t \) as above. Then
\[
\| \tilde{F}_{s,\infty}(x,y) - \tilde{F}_{s,t}(x,y) \|_2 \leq \max(C,1/c)^{1/2} e^{-t/2} \| E(X) \|_2 + e^{-t/2} \max(C,1/c)^3 - 1 \max(C,1/c) \left( \| (x, y - E(Y)) \|_2 + (m + \text{Var}(Y))^{1/2} \right)
\]
and
\[
\| \tilde{F}_{s,t}(x,y) - \tilde{F}_{\infty,t}(x,y) \|_2 \leq \frac{1}{2} (\max(C,1/c) - 1) \max(C,1/c)^{1/2} \left( e^{-s/2} \| E(X) \|_2 + \text{Var}(X) + (1 - e^{-t}) \| \text{Var}(Y) \|^{1/2} \right)
\]
The estimates of Lemmas 7.7 and 7.8 extend to the cases where \( s \) or \( t \) is infinite, where we define \( \tilde{F}_{\infty}s(x,y) = x \). Moreover, if \( (\tilde{X}_t, Y) \sim \tilde{\mu}_t \), then we have the relation \( \tilde{F}_{s,t}(\tilde{X}_t, Y) \sim (\tilde{X}_s, Y) \) when \( s,t \in [0,\infty) \).

**Remark 7.10.** We have written the explicit form of the estimates here in order to emphasize that the bounds are dimension-independent; they only depend on the parameters \( m, c, C, \| E(X) \|_2, \| E(Y) \|_2, \text{Var}(X), \) and \( \text{Var}(Y) \). The estimates also become sharper when \( c \) and \( C \) are close to 1, which would include the situation where \( V(x, y) \) is a perturbation of the quadratic potential \( (1/2)[\| x \|_2^2 + \| y \|_2^2] \). This perturbative setting was studied first in the literature, for instance in [12] and [14]; see [17], §8.3 for further discussion.

**Proof.** We first consider the case where \( s \) is fixed and \( t \to +\infty \). Note that by (7.11),
\[
\| \tilde{F}_{s,t}(x,y) - \tilde{F}_{s,t}(x',y) \|_2 = \| \tilde{F}_{s,t}(\tilde{F}_{t,t}(x,y), y) - \tilde{F}_{s,t}(x,y) \|_2 
\leq \| \tilde{F}_{s,t} \|_{\text{Lip},dx} \| \tilde{F}_{t,t}(x,y) - x \|_2 
\leq \max(C,1/c)^{1/2} \| \tilde{F}_{t,t}(x,y) - x \|_2.
\]
Lemma 2.5. By Lemma 7.8
\[ \| \tilde{F}_{t,v} - \pi_t \|_{\text{Lip}} \leq L |e^{-s/2} - e^{t/2}|, \]
where \( \max(C,1/c)^3 - 1 \) \( \max(C,1/c)^{1/2} \). Then we apply Lemma 2.5 to \( G(x,y) = \tilde{F}_{t,v}(x,y) - x \) with the random variable \((\tilde{X}_{t,v},Y)\). Note that \((\tilde{X}_{t,v},Y)\) has mean \( (e^{-t/2} E(X), E(Y)) \) and variance \( e^{-t'} \text{Var}(X) + (1 - e^{-t'})m + \text{Var}(Y) \). Moreover,
\[ E[G(\tilde{X}_{t,v},Y)] = E[\tilde{X}_t] - E[\tilde{X}_{t,v}] = (e^{-t/2} - e^{-t'/2})E(X). \]
Thus, by Lemma 2.5
\[ (7.19) \| \tilde{F}_{t,v}(x,y) - x \|_2 \leq |e^{-t/2} - e^{-t'/2}| \| E(X) \|_2 \]
\[ + L|e^{-t/2} - e^{-t'/2}| \left( \| (x - e^{-t/2} E(X), y - E(Y)) \|_2 + (e^{-t'} \text{Var}(X) + (1 - e^{-t'})m + \text{Var}(Y))^{1/2} \right). \]
Plugging this into (7.18), we see that \( \tilde{F}_{s,t} \) is Cauchy in \( t \) as \( t \to +\infty \). Moreover, we obtain the estimate (7.10) by taking \( t' \to \infty \) in (7.19) and multiplying by \( \| \tilde{F}_{s,t} \|_{\text{Lip},dx} \leq \max(c,1/c)^{1/2} \).

Now let us fix \( t \) and consider when \( s' \) and \( s \) approach \( \infty \). The argument for this case is similar but antisymmetrical. We estimate
\[ \| \tilde{F}_{s',t}(x,y) - \tilde{F}_{s,t}(x,y) \|_2 \leq \| \tilde{F}_{s',s}(\tilde{F}_{s,t}(x,y),y) - \tilde{F}_{s,t}(x,y) \|_2 \]
\[ \leq \| \tilde{F}_{s',s} - \pi_t \|_{\text{Lip,}d} \| \tilde{F}_{s,t}(x,y) \|_2 \]
\[ \leq \frac{1}{2} \max(C,1/c - 1) \max(C,1/c)^{1/2} |e^{-s} - e^{-s'}| \| \tilde{F}_{s,t}(x,y) \|_2, \]
where the last line follows from (7.14). Then by applying Lemma 2.5 to the function \( \tilde{F}_{s,t}(x,y) \) and the random variable \((\tilde{X}_t,Y)\), together with (7.13), we obtain
\[ \| \tilde{F}_{s,t}(x,y) \|_2 \leq e^{-s/2} \| E(X) \|_2 \]
\[ + \max(C,1/c)^{1/2} \left( \| (x - e^{-t/2} E(X), y - E(Y)) \|_2 + (e^{-t} \text{Var}(X) + (1 - e^{-t})m + \text{Var}(Y))^{1/2} \right) \]
This produces an estimate on \( \| \tilde{F}_{s',t} - \tilde{F}_{s,t} \|_2 \) which shows that \( \tilde{F}_{s,t} \) is Cauchy as \( s \to \infty \), so that \( \tilde{F}_{\infty,t} \) is well-defined. The explicit bound on the rate of convergence follows fixing \( s \) and \( t \), combining the above estimates, and taking \( s' \to \infty \).

Finally, since we have established convergence of \( \tilde{F}_{s,t} \) as \( s \) or \( t \) approaches \( \infty \), a routine argument with limits will extend the estimates of Lemmas 7.7 and 7.8 and the transport relations, to the cases where \( s \) or \( t \) is \( +\infty \). \( \square \)

7.5. Transport in the Large \( N \) Limit. If \( V^{(N)} \in \mathcal{C}_{m+n}^{(N)}(c,C) \) and \( \{DV^{(N)}\} \) is asymptotically approximable by trace polynomials, then we must show that the associated sequence of transport maps is asymptotically approximable by trace polynomials, and hence conclude that they define transport for the non-commutative random variables in the large \( N \) limit.

**Theorem 7.11.** For \( N \in \mathbb{N} \), let \( V^{(N)}(x,y) \) be a potential on \( M_N(\mathbb{C})^{m+n} \) satisfying Assumption 5.7 for some \( 0 < c \leq C \), and let \( \mu^{(N)} \) be the corresponding probability measures on \( M_N(\mathbb{C})^{m+n} \). Let \((X^{(N)},Y^{(N)})\) be a random variable given by \( \mu^{(N)} \) and let \( Z^{(N)} \) be an independent GUE \( m \)-tuple. Let \( \bar{\mu}^{(N)}_t \) be the law of \((\bar{X}_t,Y) = (e^{-t/2}X^{(N)} + (1 - e^{-t})^{1/2}Z^{(N)},Y^{(N)})\),
and let \( \hat{V}_t^{(N)}(x,y) = \hat{R}_t^{(N)} V^{(N)}(e^{t/2}x,y) \) be the corresponding potential. Similarly, let \( \mu_\infty^{(N)} \) be the law of \( (Z^{(N)}, Y^{(N)}) \). For \( s,t \in [0, \infty) \), let \( \tilde{F}_{s,t}^{(N)} : M_N(\mathbb{C})^{n+n} \to M_N(\mathbb{C})^{n} \) be the solution of the initial value problem

\[
\begin{align*}
\tilde{F}_{s,t}^{(N)}(x,y) &= x \\
\partial_s \tilde{F}_{s,t}^{(N)}(x,y) &= \frac{1}{2} \left( D_x \hat{V}_s(\tilde{F}_{s,t}(x,y), y) - \tilde{F}_{s,t}(x,y) \right).
\end{align*}
\]

Then

(1) The family \( \tilde{F}_{s,t}^{(N)} \) extends continuously to \( (s,t) \in [0, \infty]^2 \).

(2) \( \tilde{F}_{s,t}^{(N)} \circ \tilde{F}_{t,u}^{(N)} = \tilde{F}_{s,u}^{(N)} \).

(3) \( \tilde{F}_{s,t}^{(N)}(\tilde{X}_t^{(N)}, Y^{(N)}), Y^{(N)}) \sim (\tilde{X}_s^{(N)}, Y^{(N)}) \).

(4) For \( s,t \in [0, \infty) \), the sequence \( \{\tilde{F}_{s,t}^{(N)}\}_{N \in \mathbb{N}} \) is \( (C/c) \max(1, 1/c)^{1/2} \)-Lipschitz for all \( s, t, \) and \( N \), and it is asymptotically approximable by trace polynomials as \( N \to \infty \).

Proof. Recall in \( \Box \) we defined \( \tilde{F}_{s,t}^{(N)} \) by renormalizing \( F_{s,t}^{(N)} \). However, that definition is equivalent to the definition of \( \tilde{F}_{s,t} \) given in this theorem because both definitions produce a solution to the ODE \( \Box \). Of course, global uniqueness of the solution holds because the vector field \( D_x \hat{V}_t^{(N)}(x,y) - x \) is uniformly Lipschitz in \((x,y)\) on any compact time interval (as we discuss in more detail below).

So claims (1), (2), and (3) follow immediately from Proposition \( \Box \). The estimate for the Lipschitz norm of \( \tilde{F}_{s,t}^{(N)} \) was shown in \( \Box \).

We finish by showing asymptotic approximability using the results of \( \Box \) Let \( V_t^{(N)} = R_t^{(N)} V^{(N)} \). By Theorem \( \Box \) (3c), \( D_x V_t^{(N)} \) is uniformly continuous in \( t \) on \([0, \infty)\). Since \( D_x \hat{V}_t^{(N)}(x,y) = e^{t/2} D_x V_{e^{t/2}-1}^{(N)}(e^{t/2}x,y) \), it follows that \( D_x \hat{V}_t^{(N)} \) is uniformly continuous in \( t \) on \([0, T] \) for every \( T > 0 \), with modulus of continuity independent of \( N \), and recall it is also uniformly Lipschitz in \((x,y)\), since \( 0 \leq H V_t^{(N)} \leq \max(C, Ce^t/(1 + C(e^t - t))) \).

Consequently, \( (1/2)(D_x \hat{V}_t^{(N)}(x,y) - x) \) is uniformly continuous in \( t \) on \([0, T] \) and uniformly Lipschitz in \((x,y)\). Also, we showed that \( D_x V_t^{(N)} \) is asymptotically approximable by trace polynomials as \( N \to \infty \) and hence so is \( D_x \hat{V}_t^{(N)} \). Thus, \( (1/2)(D_x \hat{V}_t^{(N)}(x,y) - x) \) satisfies Assumption \( \Box \) so we may apply Proposition \( \Box \) to deduce that \( \tilde{F}_{s,t}^{(N)} \) is asymptotically approximable by trace polynomials for \( s, t \in [0, \infty) \). This property extends to the case where \( s \) or \( t \) is infinite using Lemma \( \Box \) and Proposition \( \Box \). \( \square \)

Remark 7.12. Rather than citing the proof of Theorem \( \Box \) one could also argue that \( D_x V_t^{(N)} \) is asymptotically approximable directly from the construction of the semigroup \( R_t^{(N)} \) using the same reasoning as \( \Box \) Proposition \( \Box \). Moreover, this method would also show that \( D(R_t^{(N)} V^{(N)}) \) is asymptotically approximable by trace polynomials provided we can prove analogues of Theorem \( \Box \) (2) and (3) for \( D(R_t^{(N)} V^{(N)}) \) rather than only \( D_x (R_t^{(N)} V^{(N)}) \). However, all this is unnecessary work for our present purpose.

Theorem 7.13. With all the notation of the previous theorem, let \((X,Y)\) be a non-commutative random variable distributed according to the limiting free Gibbs law \( \lambda \), let \( S \) be a freely independent free semicircular m-tuple, and let \( \tilde{X}_t = e^{-t/2} X + (1 - e^{-t})^{1/2} S \). Define \( \tilde{F}_{s,t} \) by \( \tilde{F}_{s,t}^{(N)} \to \tilde{F}_{s,t} \). For \( s, t, u \in [0, +\infty] \), we have

(1) \( \tilde{F}_{s,t} \) is \( (C/c) \max(C, 1/c)^{1/2} \)-Lipschitz with respect to \( \|\cdot\|_2 \).

(2) \( \tilde{F}_{s,t} \circ \tilde{F}_{t,u} = \tilde{F}_{s,u} \).
(3) \( \tilde{F}_{s,t}(\tilde{X}_t, Y) \sim (\tilde{X}_s, Y) \) in non-commutative law.

(4) We have
\[
\left\| \tilde{F}_{s,t}(\tilde{X}_t, Y) - \tilde{F}_{s,t}'(\tilde{X}_t, Y) - (e^{-s/2} - e^{-s'/2})\tau(X) \right\|_{\infty} \\
\leq (\max(C, 1/c)^3 - 1) \max(C, 1/c)|e^{-s/2} - e^{-s'/2}||\Theta|.
\]

where \( \Theta \) is the universal constant from Proposition [3,4].

In particular, \( W^*(X, Y) \) is isomorphic to \( W^*(S, Y) \), which is the free product \( W^*(S) * W^*(Y) \).

Proof. We know that there exists \( \tilde{F}_{s,t} \) such that \( \tilde{F}_{s,t} \sim \tilde{F}_{s,t} \) because of Lemma [3,4]. Then (1) and (2) follow from the corresponding properties of \( \tilde{F}_{s,t} \) by straightforward limit arguments.

As remarked in the last proof \( D\tilde{V}_t^{(N)} \) is asymptotically approximable by trace polynomials. We also know \( D\tilde{V}_t^{(N)} \) is uniformly convex and semi-concave, and thus by Theorem [5,2] the non-commutative law of \( (\tilde{X}_t^{(N)}, Y^{(N)}) \) converges in probability to some non-commutative law. Of course, the limiting non-commutative law must be the non-commutative law of \( (\tilde{X}_t, Y) \) because the joint non-commutative law of \( (X^{(N)}, Y^{(N)}, S^{(N)}) \) converges in probability to that \( (X, Y, S) \) (as in the proof of Theorem [6,9]).

With this relation between the laws of \( (\tilde{X}_t^{(N)}, Y^{(N)}) \) and \( (\tilde{X}_t, Y) \) in hand, we can prove (3) by taking the large \( N \) limit using Corollary [6,3]. Indeed, if \( f \in \text{TrP}_{m+n}^1 \) is \( \| \cdot \|_2 \)-uniformly continuous, then \( f(\tilde{F}_{s,t}^{(N)}(x, y)) \) is also \( \| \cdot \|_2 \)-uniformly continuous and asymptotically approximable by trace polynomials by Lemma [3,10]. Thus, applying Corollary [6,3] to this function and the function 1, we get
\[
\tau(f(\tilde{F}_{s,t}(\tilde{X}_t, Y), Y)) = \lim_{N \to \infty} E[\tau_N(f(\tilde{F}_{s,t}(\tilde{X}_t^{(N)}, Y^{(N)}), Y^{(N)}))] \\
= \lim_{N \to \infty} E[\tau_N(f(\tilde{X}_s, Y))] \\
= \tau(f(\tilde{X}_s, Y)).
\]

Hence, \( \tau(f(\tilde{F}_{s,t}(\tilde{X}_t, Y), Y)) = \tau(f(\tilde{X}_s, Y)) \) for all \( f \in \text{TrP}_{m+n}^1 \) that are uniformly continuous in \( \| \cdot \|_2 \). But by Proposition [3,12] such functions \( f \) can realize every element in the \( W^* \)-algebra generated by \( (\tilde{X}_s, Y) \), and in particular all the non-commutative polynomials in \( (\tilde{X}_s, Y) \). Hence, \( (\tilde{F}_{s,t}(\tilde{X}_t, Y), Y) \sim (\tilde{X}_s, Y) \) in non-commutative law as desired.

(4) Note that
\[
\tilde{F}_{s,t}(\tilde{X}_t, Y) - \tilde{F}_{s,t}'(\tilde{X}_t, Y) = (\pi_1 - \tilde{F}_{s,t}') \circ (\tilde{F}_{s,t}'(\tilde{X}_t, Y), Y),
\]
but \( (\tilde{F}_{s,t}'(\tilde{X}_t, Y), Y) \sim (\tilde{X}_s', Y) \) in non-commutative law. Hence, it suffices to prove the desired estimate for \( \tilde{F}_{s,t}(\tilde{X}_s', Y) - \tilde{X}_s' \) rather than \( \tilde{F}_{s,t}(\tilde{X}_t, Y) - \tilde{F}_{s,t}'(\tilde{X}_t, Y) \). Now \( (\tilde{X}_s', Y) \) arises as the large \( N \) limit of the matrix models given by potential \( \tilde{V}_s^{(N)} \). By Lemma [7,0] (4), we have
\[
H\tilde{V}_s^{(N)} \geq c(1 + ct)^{-1} I_m \oplus cI_n, \text{ so that }
\]
\[
H\tilde{V}_s^{(N)} \geq \frac{ce^s}{1 + c(e^s - 1)} I_m \oplus cI_n \geq \min(1, c) I_{m+n} \geq \frac{1}{\max(C, 1/c)} I_{m+n}.
\]

By Remark [5,8] there exists a sequence of random matrix models for \( (\tilde{X}_s', Y) \) given by uniformly convex potentials which are also unitarily invariant (even if this is not true of our original model), with the same lower bound \( 1/\max(C, 1/c) \) for the Hessian of the potential. Therefore, by Proposition [3,15]
\[
\left\| \tilde{F}_{s,t'}(\tilde{X}_s', Y) - \tilde{X}_s' - \tau(\tilde{F}_{s,t'}(\tilde{X}_s', Y) - \tilde{X}_s') \right\|_{\infty} \leq \max(C, 1/c)^{1/2} \Theta \left\| \tilde{F}_{s,t'} - \pi_1 \right\|_{\text{Lip}}.
\]
We finish by substituting the estimate
\[ \|\hat{F}_{s,s'} - \pi_1\|_{\text{Lip}} \leq (\max(C, 1/c)^3 - 1) \max(C, 1/c)^{1/2} |e^{-s/2} - e^{-s'/2}| \]
which follows from (7.15) and Lemma 3.9 (the latter lemma is needed since the original statement of (7.15) is for the finite-dimensional setting for a fixed \( N \)).

The last claim regarding \( W^* \)-algebras follows from (3) by examining the case with \( s = 0 \) and \( t = \infty \) or vice versa. \( \square \)

8. Applications

We show that Assumption 5.1 is preserved under independent joins, marginals, convolution, and linear changes of variables. We conclude that for the convex free Gibbs laws considered here, \( \chi^* \) satisfies additivity under conditioning. Moreover, by iterating our conditional transport results, we obtain “upper triangular” transport maps from a convex free Gibbs law to the law of a free semicircular family, which also satisfy the entropy-cost inequality relative to the semicircular law.

8.1. Operations on Convex Gibbs Laws. Recall that Assumption 5.1 for a sequence \( \{V(N)\} \) of potentials \( M_N(\mathbb{C})^{m_n} \to \mathbb{R} \) states that \( c \leq HV^{(N)} \leq C \) for some constants \( c \) and \( C \), the sequence \( \{DV^{(N)}\} \) is asymptotically approximable by trace polynomials, and \( \int x_j d\mu^{(N)}(x) \) is a scalar matrix for each \( j \), where \( \mu^{(N)} \) is the measure associated to \( DV^{(N)} \).

**Proposition 8.1.** Suppose that \( V_1^{(N)} : M_N(\mathbb{C})^{m_n} \to \mathbb{R} \) and \( V_2^{(N)} : M_N(\mathbb{C})^{n_n} \to \mathbb{R} \) satisfy Assumption 5.1 for some \( 0 < c \leq C \). Then \( V^{(N)}(x,y) := V_1^{(N)}(x) + V_2^{(N)}(y) \) also satisfies Assumption 5.1 for the same \( c \) and \( C \).

Moreover, let \( \mu_1^{(N)} \), \( \mu_2^{(N)} \), and \( \mu^{(N)} \) be the measures associated to \( V_1^{(N)} \), \( V_2^{(N)} \), and \( V^{(N)} \) respectively, and let \( \lambda_1 \), \( \lambda_2 \), and \( \lambda \) be the respective limiting free Gibbs laws given by Theorem 5.2. Then \( \mu^{(N)} \) is the independent join of \( \mu_1^{(N)} \) and \( \mu_2^{(N)} \) and \( \lambda \) is the freely independent join of \( \lambda_1 \) and \( \lambda_2 \).

**Proof.** The claim \( c \leq HV^{(N)} \leq C \) follows because \( HV^{(N)}(x,y) = HV_1^{(N)}(x) \oplus HV_2^{(N)}(y) \). The claim about asymptotic approximation by trace polynomials follows because \( DV^{(N)}(x,y) = \langle DV_1^{(N)}(x), DV_2^{(N)}(y) \rangle \) and each component is asymptotically approximable by trace polynomials.

It is clear that the probability density for \( \mu^{(N)} \) is the tensor product of the probability densities for \( \mu_1^{(N)} \) and \( \mu_2^{(N)} \) and hence \( \mu^{(N)} \) is the independent join of these two marginal laws. It follows that \( \int x_j d\mu^{(N)}(x) \) and \( \int y_j d\mu^{(N)}(y) \) are scalar matrices, hence Assumption 5.1 holds for \( V^{(N)} \).

Let \( (X^{(N)}, Y^{(N)}) \sim \mu^{(N)} \) be random variables and let \( (X,Y) \sim \lambda \) be non-commutative random variables. Then by Theorem 6.6
\[ \Phi^*(X,Y) = \lim_{N \to \infty} E \left[ \|DV^{(N)}(X^{(N)}, Y^{(N)})\|_2^2 \right] \]
\[ = \lim_{N \to \infty} E \left[ \|DV_1^{(N)}(X^{(N)})\|_2^2 \right] + \lim_{N \to \infty} E \left[ \|DV_2^{(N)}(Y^{(N)})\|_2^2 \right] \]
\[ = \Phi^*(X) + \Phi^*(Y). \]

It was shown in Proposition 5.18(c) that \( \Phi^*(X,Y) = \Phi^*(X) + \Phi^*(Y) \) implies that \( X \) and \( Y \) are freely independent. \( \square \)

**Proposition 8.2.** Suppose that \( V^{(N)} : M_N(\mathbb{C})^{m_n+n} \to \mathbb{R} \) satisfies Assumption 5.1. Let \( \mu^{(N)} \) be the corresponding law, let \( (X^{(N)}, Y^{(N)}) \sim \mu^{(N)} \) and let \( \mu_1^{(N)} \) and \( \mu_2^{(N)} \) be the laws of \( X^{(N)} \) and
Then \( \mu_1^{(N)} \) and \( \mu_2^{(N)} \) are given by a potential \( W^{(N)} \) that also satisfies Assumption B.1 for the same values of \( c \) and \( C \).

**Proof.** By symmetry, it suffices to prove the claims for \( \mu_2^{(N)} \). First, it is immediate that the mean of \( y_j \) under \( \mu_2^{(N)} \) is a scalar, since it is \( E[Y_j^{(N)}] \). Moreover, if we define

\[
V_2^{(N)}(x) = -\frac{1}{N^2} \int e^{-N^2 V^{(N)}(x,y)} \, dx,
\]

then (as in the proof of Theorem 4.1) we may compute \( DV_1^{(N)} \) by differentiating under the integral and obtain

\[
DV_2^{(N)}(Y^{(N)}) = E[D_y V(X^{(N)}, Y^{(N)})] Y^{(N)}.
\]

It follows by Theorem 5.1 that \( \{DV_2^{(N)}\} \) is asymptotically approximable by trace polynomials.

Finally, the fact that \( c \leq HV_2^{(N)} \leq C \) follows from [7, Theorem 4.3], or alternatively by the following reasoning. Let \( \mu_1^{(N)} \) be the law of \( (e^{-t/2}X^{(N)} + (1-e^{-t})/2S^{(N)}, Y^{(N)}) \), where \( S^{(N)} \) is an independent GUE tuple. The corresponding potential \( \hat{V}_t^{(N)} \) is given by (7.8) and it satisfies

\[
\frac{ce^t}{1 + c(e^t - 1)} I_m + c I_n \leq H\hat{V}_t^{(N)} \leq \frac{Ce^t}{1 + C(e^t - 1)} I_m + CI_n
\]

by direct substitution of (7.8) into Lemma 7.3 (4) and hence

\[
\min(c,1) I_m + c I_n \leq H\hat{V}_t^{(N)} \leq \max(C,1) I_m + CI_n
\]

Now as \( t \to \infty \), the law \( \hat{\mu}_t \) converges to the law \( \hat{\mu}_\infty \) of \( (S^{(N)}, Y^{(N)}) \). By applying Lemma 2.7, \( \hat{\mu}_\infty \) is given by some potential \( W^{(N)}(x, y) \) satisfying

\[
\min(c,1) I_m + c I_n \leq HW^{(N)} \leq \max(C,1) I_m + CI_n.
\]

However, we know that \( W^{(N)}(x, y) = (1/2)\|x\|_2^2 + V_2^{(N)}(y) \) because the potential corresponding to a law is unique. This implies that \( c \leq HV_2^{(N)} \leq C \) as desired. \( \square \)

**Proposition 8.3.** Let \( V^{(N)} : M_N(\mathbb{C})_{sa}^m \to \mathbb{R} \) satisfy Assumption B.1, for some \( 0 < c \leq C \), and let \( X^{(N)} \) be the corresponding random variable. Let \( A \) be an invertible \( m \times m \) matrix with real entries and let \( A^{(N)} \) denote the linear transformation \( M_N(\mathbb{C})_{sa}^m \to M_N(\mathbb{C})_{sa}^m \) given by

\[
(A^{(N)}x)_i = \sum_{j=1}^m A_{i,j} x.
\]

Then \( \hat{V}^{(N)} = V^{(N)}((A^{-1})^{(N)}) \) is the potential corresponding to \( A^{(N)}X^{(N)} \), and \( \hat{V}^{(N)} \) satisfies Assumption 5.1 with constants \( c/\|A\| \) and \( C\|A^{-1}\| \).

**Proof.** The fact that \( \hat{V}^{(N)} \) is the potential corresponding to \( A^{(N)}X^{(N)} \) follows from change of variables. Now it is immediate that the expectation of \( (A^{(N)}X^{(N)})_i \) is a scalar multiple of identity for each \( i \). Next, by the chain rule

\[
DV^{(N)}(x) = ((A^{-1})^{(N)})^T DV^{(N)}((A^{-1})^{(N)}x) = ((A^{-1})^T)^{(N)} DV^{(N)}((A^{-1})^{(N)}x),
\]

and from this it follows that \( \{DV^{(N)}\} \) is asymptotically approximable by trace polynomials. Similarly, by the chain rule,

\[
HV^{(N)}(x) = ((A^{-1})^T)^{(N)} HV^{(N)}((A^{-1})^{(N)}x)(A^{-1})^{(N)}.
\]

The maximum and minimum singular values of \( (A^{-1})^{(N)} \) are the same as those of \( A^{-1} \), which are \( \|A^{-1}\| \) and \( 1/\|A\| \) respectively. By a basic linear algebra argument, it follows that \( c/\|A\| \leq H\hat{V}^{(N)} \leq C\|A^{-1}\| \). \( \square \)
Proof. Let \( \mu \) and \( \nu \) be two potentials \( M_N(C)_\mu \rightarrow \mathbb{R} \) satisfying Assumption 5.2 with constants \( c \) and \( C \). Let \( X(N) \) and \( Y(N) \) be the corresponding random tuples of matrices. Then the law of \( X(N) + Y(N) \) is given by another potential \( \tilde{V}(N) \) satisfying Assumption 5.1 with constants \( \sqrt{2c} \) and \( \sqrt{2C} \). Moreover, the free Gibbs state corresponding to \( \{ \tilde{V}(N) \} \) is the free convolution of those corresponding to \( \{ V_1(N) \} \) and \( \{ V_2(N) \} \).

Proof. Let \( V(N)(x, y) = V_1(N)(x) + V_2(N)(y) \), which satisfies Assumption 5.1 (with the same constants) by Proposition 5.1. Now let \( A \) be the \( 2m \times 2m \) matrix

\[
A = \begin{pmatrix} I & I \\ -I & I \end{pmatrix}.
\]

Since \( A/\sqrt{2} \) is an isometry, we have \( \|A\| = \sqrt{2} \) and \( \|A^{-1}\| = 1/\sqrt{2} \). Therefore, by Proposition 8.3 the law of \( (X(N) + Y(N), -X(N) + Y(N)) \) is given by a potential satisfying Assumption 5.1 with constants \( \sqrt{2c} \) and \( \sqrt{2C} \). Then by Proposition 8.2 the law of \( X(N) + Y(N) \) is given by such a potential with the same constants \( \sqrt{2c} \) and \( \sqrt{2C} \).

We showed in Proposition 8.1 that the large \( N \) limit of the law of \( (X(N), Y(N)) \) given a freely independent join of the corresponding marginals. Hence, the large \( N \) limit of the law of \( X(N) + Y(N) \) is given by the free convolution.

As a consequence, we have additivity of entropy under conditioning.

Corollary 8.5. Let \( V(N)(x, y) \) be a potential satisfying Assumption 5.1 as in the set of Theorem 6.6. Let \( X, Y \) be a tuple of non-commutative random variables distributed according to the limiting free Gibbs law associated to \( V(N) \). Then

\[
\chi^*(X, Y) = \chi^*(X|Y) + \chi^*(Y).
\]

Proof. From standard classical results, we have

\[
h(X(N), Y(N)) = h(X(N)|Y(N)) + h(Y(N)).
\]

Dividing by \( N^2 \) and adding \( \frac{1}{2} (m + n) \log N \) to both sides, we obtain the normalized version

\[
h(N)(X(N), Y(N)) = h(N)(X(N)|Y(N)) + h(N)(Y(N)).
\]

By the previous theorem, we obtain the desired relation for \( \chi^* \) in the limit as \( N \rightarrow \infty \). More precisely, we apply the theorem as stated to \( h(N)(X(N)|Y(N)) \). Meanwhile, for \( h(N)(X(N), Y(N)) \) and \( h(N)(Y(N)) \) we apply the special case of the theorem where we condition on \( 0 \) variables.

8.2. Entropy and Fisher Information Relative to Gaussian. As background for our discussion of the entropy-cost inequality in [8.3] we review the entropy of one probability measure relative to another. If \( \nu \) is a measure on \( \mathbb{R}^m \), then the entropy of \( \mu \) relative to \( \nu \) is

\[
h(\mu|\nu) := -\int \rho \log \rho \ d\mu, \text{ where } \rho = \frac{d\mu}{d\nu}
\]

whenever the integral is well-defined. The standard entropy \( h(\mu) = -\int \rho \log \rho \) corresponds to the choice of Lebesgue measure for \( \nu \).

Remark 8.6. The reader should be careful to distinguish between the relative entropy \( h(\mu|\nu) \) and the conditional entropy \( h(X|Y) \). The first changes the ambient measure while the second describes conditioning on \( Y \).

Remark 8.7. If \( \mu \) and \( \nu \) are both probability measures, then \( h(\mu|\nu) \leq 0 \). For this reason, many authors choose to change the sign. We will keep the sign convention given above to be consistent with our convention for \( h(\mu) \) relative to Lebesgue measure, but we will write absolute value signs around relative entropy when it is natural to use the positive version.
For probability measures on $M_N(\mathbb{C})_{sa}^m$, we may study entropy relative to the Gaussian measure $\sigma_{m,t}^{(N)}$ on $M_N(\mathbb{C})_{sa}^m$. A direct computation shows that if $X \sim \mu$ is a random variable in $M_N(\mathbb{C})_{sa}^m$, then we have

$$h(\mu|\sigma_{m,t}^{(N)}) = h(\mu) - \frac{N^2}{2} E\|X\|_2^2 + \frac{mN^2}{2} \log \frac{N}{2\pi}.$$  

We denote the normalized version by

$$h_g^{(N)}(\mu) := h_g^{(N)}(X) := \frac{1}{N^2} h(\mu|\sigma_{m,t}^{(N)}) = h^{(N)}(X) - \frac{1}{2} E\|X\|_2^2 - \frac{m}{2} \log 2\pi.$$  

Similarly, if $\mu$ is a measure on $M_N(\mathbb{C})_{sa}^{m+n}$ which absolutely continuous with respect to Lebesgue measure and $(X,Y)$ is the corresponding random variable, we define

$$h_g^{(N)}(X|Y) = h^{(N)}(X|Y) - \frac{1}{2} E\|X\|_2^2 - \frac{m}{2} \log 2\pi,$$

which is equivalent to

$$h_g^{(N)}(X|Y) = \int h_g^{(N)}(\mu_{X|Y=y}) d\mu_Y(y),$$

where $\mu_{X|Y=y}$ is the conditional distribution of $X$ given $Y = y$, and $\mu_Y$ is the marginal law of $Y$. Similarly, if $(X,Y)$ is an $(m+n)$-tuple of non-commutative random variables, we define the free entropy $\chi^*$ relative to Gaussian by

$$\chi^*_g(X|Y) = \chi^*(X|Y) - \frac{1}{2} E\|X\|_2^2 - \frac{m}{2} \log 2\pi.$$  

We define the normalized conditional Fisher information relative to Gaussian by

(8.1)  

$$\mathcal{T}^{(N)}_g(X|Y) = \mathcal{T}^{(N)}(X|Y) - 2m + E\|X\|_2^2.$$  

Note that if this Fisher information is finite and if $\xi$ is the normalized conjugate variable for $X$ given $Y$ as in [6.2] then

$$\mathcal{T}^{(N)}_g(X|Y) = E\|\xi - X\|_2^2$$

because

$$E\|\xi - X\|_2^2 = E\|\xi\|_2^2 - 2E(\xi, X)_2 + E\|X\|_2^2 = \mathcal{T}^{(N)}(X|Y) - 2m + E\|X\|_2^2,$$

where we have evaluated the middle term on the right hand side using integration by parts. Similarly, for an $(m+n)$-tuple $(X,Y)$ of non-commutative random variables, we define

(8.2)  

$$\Phi^*_g(X|Y) = \Phi^*(X|Y) - 2m + \|X\|_2^2 = \|\xi - X\|_2^2,$$

where the second equality holds provided that $\Phi^*$ is finite and $\xi$ is the free conjugate variable. We have the following version of (6.6) and Lemma 6.3 for entropy and Fisher information relative to Gaussian.

**Lemma 8.8.** Let $(X,Y)$ be a random variable in $M_N(\mathbb{C})_{sa}^{m+n}$ with a density and with finite variance and let $S$ be an independent GUE $m$-tuple. Then

$$|h_g^{(N)}(X|Y)| = \frac{1}{2} \int_{0}^{\infty} \mathcal{T}^{(N)}_g(e^{-t/2}X + (1 - e^{-t})^{1/2}S|Y) dt.$$  

Similarly, suppose that $(X,Y)$ is an $(m+n)$-tuple of non-commutative random variables and let $S$ be a freely independent free semicircular $m$-tuple. Then

$$|\chi^*_g(X|Y)| = \frac{1}{2} \int_{0}^{\infty} \Phi^*_g(e^{-t/2}X + (1 - e^{-t})^{1/2}S|Y) dt.$$
Proof. The first formula follows from [24, §4, Lemma 1] after renormalization. However, we will give an argument by a change of variables in (6.6) that will apply to both $h^g(N)$ and $\chi^*_g$. Note that by (6.6)

$$h^g(N)(X|Y) = h^g(N)(X|Y) - \frac{1}{2}E\|X\|_2^2 - \frac{m}{2} \log 2\pi$$

$$= \frac{1}{2} \int_0^\infty \left( \frac{m}{1+t} - T(N)(X + t^{1/2}S|Y) \right) dt + \frac{m}{2} - \frac{1}{2}E\|X\|_2^2,$$

and in particular, we know that the integral is well-defined in $[-\infty, +\infty)$. Now we do a change of variables in the integral $t = e^u - 1$, $dt = e^u \, du$ and obtain

$$\frac{1}{2} \int_0^\infty \left( \frac{m}{1+t} - T(N)(X + t^{1/2}S|Y) \right) dt = \frac{1}{2} \int_0^\infty \left( \frac{m}{e^u} - T(N)(X + (e^u - 1)^{1/2}S|Y) \right) e^u \, du$$

$$= \frac{1}{2} \int_0^\infty \left( m - T(N)(e^{-u/2}X + (1 - e^{-u})^{1/2}S|Y) \right) du,$$

where we have applied the scaling relation Lemma 6.2 for Fisher information. On the other hand,

$$\frac{m}{2} - \frac{1}{2}E\|X\|_2^2 = \frac{1}{2} \left( E\|S\|_2^2 - E\|X\|_2^2 \right)$$

$$= \frac{1}{2} \int_0^\infty e^{-u} \left( E\|S\|_2^2 - E\|X\|_2^2 \right) du$$

$$= \frac{1}{2} \int_0^\infty \left( m - E\|e^{-u/2}X + (1 - e^{-u})^{1/2}S\|_2^2 \right) du.$$

Therefore, altogether

$$h^g(N)(X|Y) = \frac{1}{2} \int_0^\infty \left( 2m - T(N)(e^{-u/2}X + (1 - e^{-u})^{1/2}S|Y) - E\|e^{-u/2}X + (1 - e^{-u})^{1/2}S\|_2^2 \right) du$$

$$= -\frac{1}{2} \int_0^\infty T(N)(e^{-u/2}X + (1 - e^{-u})^{1/2}S|Y) du,$$

which is the desired formula. The statement for $\chi^*$ can be proved by exactly the same computation, since the definition of $\chi^*$ in (6.7) is completely analogous to (6.6). \qed

Furthermore, the log-Sobolev inequality for the Gaussian measure has the following interpretation for entropy and Fisher’s information. This in fact generalizes to entropy and Fisher’s information relative to any measure $\nu$ satisfying LSI, see [24, Definition 1], but we only use the case where $\nu$ is Gaussian and $\mu$ is sufficiently regular.

**Lemma 8.9.** Let $X$ be a random variable in $\mathcal{M}_N(\mathbb{C})_{sa}$ that has a $C^1$ density with respect to Lebesgue measure. Then

$$|h^g(N)(X|Y)| \leq \frac{1}{2} T(N)(X|Y).$$

**Proof.** First, it suffices to check the non-conditional version $h^g(N)(X) \leq \frac{1}{2} T(N)(X)$. Indeed, in the conditional case, the left hand side is $\int h^g(N)(\mu_X|Y=y) \, d\mu_Y(y)$ and the right hand side is $\int T(N)(\mu_X|Y=y) \, d\mu_Y(y)$, and solving the non-conditional case would allow us to compare the integrands pointwise.

Now suppose that $X$ has density $\rho$ with respect to Lebesgue measure and let $\tilde{\rho}$ be the density with respect to Gaussian, so that

$$\rho(x) = \tilde{\rho}(x) \frac{1}{(2\pi/N)^{N^2/2}} e^{-N^2\|x\|_2^2} \, dx.$$
By Corollary 2.11, the measure $\sigma_{m,t}^{(N)}$ satisfies the normalized log-Sobolev inequality (2.4) with $c = 1$, so that
\[
\frac{1}{N^2} \int f^2 \log \frac{f^2}{\int f^2 d\sigma_{m,t}^{(N)}} \, d\sigma_{m,t}^{(N)} \leq \frac{2}{N^4} \int \|Df\|_2^2 d\sigma_{m,t}^{(N)}.
\]
Let $f = \tilde{\rho}^{1/2}$. Then $\int f^2 \, d\sigma_{m,t}^{(N)}$ reduces to 1, so the right hand side is $|h_2^{(N)}(X)|$. On the other hand, letting $V(x) = -(1/N^2) \log \rho$, we get
\[
f(x) = \tilde{\rho}(x)^{1/2} = (\text{constant})e^{-N^2V(x)/2-N^2\|x\|^2/4},
\]
and hence on the support of $f$, we have
\[
Df(x) = -\frac{N^2}{2} (DV(x) - x)\tilde{\rho}(x)^{1/2}.
\]
Thus,
\[
\frac{2}{N^2} \int \|Df\|_2^2 \, d\sigma_{m,t}^{(N)} = \frac{1}{2} \int \|DV(x) - x\|^2 \, d\mu(x) = \frac{1}{2} I_g^{(N)}(X).
\]
Hence, the log-Sobolev inequality implies the desired inequality. \hfill \square

### 8.3. Conditional Transport and the Entropy-Cost Inequality

Now we will show that the transport maps constructed in Theorem 7.3 satisfy the Talagrand entropy-cost inequality. It was shown in [24, Theorem 1] that if a measure $\nu$ satisfies the log-Sobolev inequality (2.1) with some constant $c$ (and some regularity conditions), then it satisfies the Talagrand inequality
\[
W_2(\mu, \nu)^2 \leq \frac{2h(\mu|\nu)}{\rho} \quad \text{for all } \mu,
\]
where $W_2$ is the $L^2$-Wasserstein distance, which is equivalent to the infimum of $\|X - Y\|_2$ over all coupled random variables $X$ and $Y$ with $X \sim \mu$ and $Y \sim \nu$.

Adapting Otto and Villani’s argument, we will show that the transport maps constructed in Theorem 7.3 witness the (conditional) entropy-cost inequality relative to the GUE law for the $N \times N$ matrix models and the corresponding free entropy-cost inequality for the non-commutative random variables. This is claim (5) below, while the other claims in Theorem 8.10 summarize the results of our earlier construction.

We remark that the free Talagrand inequality for self-adjoint tuples was studied in greater generality in [16] and [4, §3.3]. Although we restricted ourselves to the case where the target measure is Gaussian/semicircular, our goal in this paper was not merely to estimate the Wasserstein distance using some coupling, but rather to exhibit a coupling that arises from a transport map, and to show Lipschitzness of this transport map.

**Theorem 8.10.** As in Theorem 7.11, let $V^{(N)}(x,y)$ be a potential on $M_N(\mathbb{C})_{sa}^{m+n}$ satisfying Assumption 5.2 for some $0 < c \leq C$, and let $\mu^{(N)}$ and $(X^{(N)}, Y^{(N)})$ be the corresponding probability measures and random variables. Let $S^{(N)}$ be an independent GUE $m$-tuple. Let $(X, Y)$ be a tuple of non-commutative random variables given by the limiting free Gibbs law $\lambda$ and let $S$ be a freely independent free semicircular $m$-tuple. Let $\pi_1(x, y) = x$ and $\pi_2(x, y) = y$. Then there exist functions $F^{(N)}$, $G^{(N)} : M_N(\mathbb{C})_{sa}^{m+n} \to M_N(\mathbb{C})_sa$ and $F, G \in \mathbb{TIP}_m^{m+n}$ such that

(1) We have $(F^{(N)}(X^{(N)}, Y^{(N)}), Y^{(N)}) \sim (S^{(N)}, Y^{(N)})$ and $(G^{(N)}(S^{(N)}, Y^{(N)}), Y^{(N)}) \sim (X^{(N)}, Y^{(N)})$ in law, and $(F(X, Y), Y) \sim (S, Y)$ and $(G(S, Y), Y) \sim (X, Y)$ in non-commutative law.

(2) $(F^{(N)} \circ (G^{(N)} \circ \pi_2)) = (G^{(N)} \circ (F^{(N)} \circ \pi_2)) = \text{id}$ and the same holds for $F$ and $G$.

(3) $F^{(N)} \approx F$ and $G^{(N)} \approx G$. 

(4) We have \( \|F^N - \pi_1\|_{\text{Lip}} \) and \( \|G^N - \pi_1\|_{\text{Lip}} \leq (\max(C,1/c)^3 - 1) \max(C,1/c)^{1/2} \), and the same holds for \( F \) and \( G \).

(5) We have

\[
\|F^N(X^N), Y^N) - X^N\|_{L^2}^2 = \|G^N(S^N), Y^N) - S^N\|_{L^2}^2 \leq 2h^N_g(X^N|Y^N).
\]

and

\[
\|F(X, Y) - X\|_2^2 = \|G(S, Y) - S\|_2^2 \leq 2\chi_g^*(X|Y).
\]

Proof. Let \( \tilde{F}_{s,t}^N \) and \( \tilde{F}_{s,t} \) be as in Theorems 7.11 and 7.13. Then let

\[
F^N = \tilde{F}_{\infty,0}^N,
G^N = \tilde{F}_{0,\infty}^N,
F = \tilde{F}_{\infty,0},
G = \tilde{F}_{0,\infty}.
\]

The only property that was not shown in the earlier theorems is (5). First, note that as a consequence of (1),

\[
\|F^N(X^N), Y^N) - X^N\|_{L^2} = \|S^N - G^N(S^N), Y^N)\|_{L^2}.
\]

The rest of the proof of (5) proceeds as in [24, §4]. As in §7.3 let \( \tilde{V}_{s,t}^N \) denote the potential corresponding to \( (\tilde{X}_s^N, Y^N)) = (e^{-t/2}X^N + (1 - e^{-t})^{1/2}\tilde{S}(N), Y^N)) \) and recall that

\[
\partial_s \tilde{F}_{s,t}^N(x, y) = \frac{1}{2}(D_x \tilde{V}_{s,t}^N(\tilde{F}_{s,t}^N(x, y) - \tilde{F}_{s,t}^N(x, y)),
\]

and hence

\[
\left\| \tilde{F}_{\infty,0}^N(x, y) - x \right\|_2 \leq \frac{1}{2} \int_0^\infty \left\| D_x \tilde{V}_{s,t}^N(\tilde{F}_{s,t}^N(x, y) - \tilde{F}_{s,t}^N(x, y)) \right\|_2 ds.
\]

Then we apply Minkowski’s inequality with respect to integration \( d\mu^N(x, y) \) to obtain

\[
\left( \int \left\| \tilde{F}_{\infty,0}^N(x, y) - x \right\|_2^2 d\mu^N(x, y) \right)^{1/2} \leq \frac{1}{2} \int_0^\infty \left( \int \left\| D_x \tilde{V}_{s,t}^N(\tilde{F}_{s,t}^N(x, y) - \tilde{F}_{s,t}^N(x, y)) \right\|_2^2 d\mu^N(x, y) \right)^{1/2} ds
\]

which can be rewritten as

\[
\|F^N(X^N), Y^N) - X^N\|_{L^2} \leq \frac{1}{2} \int_0^\infty \left\| D_x \tilde{V}_{s,t}^N(\tilde{F}_{s,t}^N(X^N), Y^N)) - \tilde{F}_{s,t}^N(X^N), Y^N)) \right\|_{L^2} ds
= \frac{1}{2} \int_0^\infty \left\| D_x \tilde{V}_{s,t}^N(\tilde{X}_s^N, Y^N)) - \tilde{X}_s^N \right\|_{L^2} ds
= \frac{1}{2} \int_0^\infty \mathcal{T}_g^N(\tilde{X}_s^N|Y^N)^{1/2} ds,
\]

where we have applied the fact that \( (\tilde{F}_{s,0}^N(X^N), Y^N)) \sim (\tilde{X}_s^N, Y^N)) \). It follows from Lemma 8.8 and a change of variables that

\[
(8.3) \quad h^N_g(\tilde{X}_s^N|Y^N) = \int_t^\infty \mathcal{T}_g^N(\tilde{X}_s^N|Y^N) ds.
\]

It is easy to see that \( s \rightarrow \mathcal{T}_g^N(\tilde{X}_s^N|Y^N) \) is bounded on compact sets because of Lemma 6.3 and 5.1. Therefore, we have for almost every \( t, \)

\[
\frac{d}{dt} |h^N_g(\tilde{X}_t^N|Y^N))| = \frac{1}{2} \mathcal{T}_g^N(\tilde{X}_t^N|Y^N).\]
Hence, for almost every \( t \),
\[
\frac{d}{dt} |h^{(N)}_g(\tilde{X}^{(N)}_t|Y^{(N)})|^{1/2} \geq \frac{1}{4} \tau^{(N)}_g(\tilde{X}^{(N)}_t|Y^{(N)}) |h^{(N)}_g(\tilde{X}^{(N)}_t|Y^{(N)})|^{-1/2}
\]
\[
\geq \frac{1}{2\sqrt{2}} \tau^{(N)}_g(\tilde{X}^{(N)}_t|Y^{(N)})^{1/2},
\]
where the last line follows from Lemma 8.9. Therefore,
\[
\left\| F^{(N)}(X^{(N)}, Y^{(N)}) - X^{(N)} \right\|_{L^2} \leq \sqrt{\frac{1}{2}} \int_0^\infty \frac{d}{dt} |h^{(N)}_g(\tilde{X}^{(N)}_t|Y^{(N)})|^{1/2} dt
\]
\[
= \sqrt{\frac{1}{2}} |h^{(N)}_g(X^{(N)}|Y^{(N)})|,
\]
where we have employed the fact that \( \lim_{t \to \infty} |h^{(N)}_g(\tilde{X}^{(N)}_t|Y^{(N)})| = 0 \) by (8.3). This establishes the first claim of (5).

The second claim of (5) follows by taking the large \( N \) limit using Corollary 5.3 and Theorem 6.6. More precisely, for the left hand side, we take the limit using Corollary 5.3. Meanwhile, for the right hand side, note that \( h^{(N)}_g(X^{(N)}|Y^{(N)}) \to \chi^{(N)}_g(X|Y) \) because \( h^{(N)}_g(X^{(N)}|Y^{(N)}) \to \chi^*(X|Y) \) and \( E\|X^{(N)}\|_2^2 \to \|X\|_2^2 \) by Corollary 5.3.

\[ \square \]

8.4. Construction of Triangular Transport. Finally, by iterating Theorem 8.11, we obtain the following result concerning “upper-triangular transport.”

Theorem 8.11. Let \( V^{(N)} : M_N(\mathbb{C})^m \to \mathbb{R} \) be a potential satisfying Assumption 7.1. Let \( \mu^{(N)} \) and \( X^{(N)} \) be the corresponding law and random variable. Let \( \lambda \) be the limiting free Gibbs law, and let \( X \sim \mu \) be an \( m \)-tuple of non-commutative random variables. Let \( S^{(N)} \) be an independent GUE \( m \)-tuple and let \( S \) be a freely independent free semicircular family. Then there exist functions \( \Phi^{(N)}, \Psi^{(N)} : M_N(\mathbb{C})^m \to M_N(\mathbb{C})^m \) and \( \Phi, \Psi \in (\text{Tri}_m)^m \) such that

1. \( \Phi^{(N)}(X^{(N)}) \sim S^{(N)} \) and \( \Psi^{(N)}(S^{(N)}) \sim X^{(N)} \) in law, and similarly, \( \Phi(X) \sim S \) and \( \Psi(S) \sim X \) in non-commutative law.
2. \( \Phi^{(N)} \) and \( \Psi^{(N)} \) are inverse functions of each other, and the same holds for \( \Phi \) and \( \Psi \).
3. \( \Phi^{(N)} \to \Phi \) and \( \Psi^{(N)} \to \Psi \).
4. \( \Phi^{(N)} \) is upper triangular in the sense that
   \[
   \Phi^{(N)}(x_1, \ldots, x_m) = (\Phi^{(N)}_1(x_1, \ldots, x_m), \Phi^{(N)}_2(x_2, \ldots, x_m), \ldots, \Phi^{(N)}_m(x_m))
   \]
   and the same holds for \( \Psi^{(N)} \), \( \Phi \), and \( \Psi \). In particular, the isomorphism \( W^*(X) \to W^*(S) \) induced by \( \Phi \) maps \( W^*(X_k, \ldots, X_m) \) onto \( W^*(S_k, \ldots, S_m) \) for each \( k = 1, \ldots, m \).
5. We have \( \| \Phi^{(N)} - \text{id} \|_{\text{Lip}} \leq m^{1/2}(\max(C, 1/c)^3 - 1) \max(C, 1/c)^{1/2} \) and \( \| \Psi^{(N)} - \text{id} \|_{\text{Lip}} \) is bounded by some constant \( L(c, C, m) \) which goes to zero as \( c, C \to 1 \).
6. We have
   \[
   \left\| \Phi^{(N)}(X^{(N)}) - X^{(N)} \right\|_{L^2}^2 = \left\| \Psi^{(N)}(S^{(N)}) - S^{(N)} \right\|_{L^2}^2 \leq 2h^{(N)}_g(X^{(N)}).
   \]
   and
   \[
   \left\| \Phi(X) - X \right\|_2^2 = \left\| \Psi(S) - S \right\|_2^2 \leq 2\chi^{(N)}_g(X).
   \]
7. We have
   \[
   \| \Phi_j(X_j, \ldots, X_m) - (X_j - \tau(X_j)) \|_\infty = \|(\Psi_j(S_j, \ldots, S_m) - \tau(\Psi_j(S))) - S_j \|_\infty \leq (\max(C, 1/c)^3 - 1) \max(C, 1/c) \Theta,
   \]
   where \( \Theta \) is the universal constant from Proposition 3.12.
Proof. First, by Proposition 8.2, the marginal law of \((X_j^{(N)}, \ldots, X_m^{(N)})\) is given by a convex potential satisfying the same assumptions.

For each \(j\), we apply Theorem 8.10 with \(X_j^{(N)}\) as the first variable and \((X_j^{(N)}, \ldots, X_m^{(N)})\) as the second variable. We thus obtain maps \(\Phi_j^{(N)} : M_N(\mathbb{C})_{sa}^{m-j+1} \to M_N(\mathbb{C})_{sa}\) such that

\[
(\Phi_j^{(N)}(X_j^{(N)}, \ldots, X_m^{(N)}), X_j^{(N)}, \ldots, X_m^{(N)}) \sim (S_j^{(N)}, X_j^{(N)}, \ldots, X_m^{(N)}).
\]

Let

\[
\Phi^{(N)}(x_1, \ldots, x_m) = (\Phi_1^{(N)}(x_1, \ldots, x_m), \Phi_2^{(N)}(x_2, \ldots, x_m), \ldots, \Phi_m^{(N)}(x_m)).
\]

Let \(Y^{(N)} = \Phi^{(N)}(X^{(N)})\). Then we can check by induction on \(j\) that

\[
(Y_1^{(N)}, \ldots, Y_j^{(N)}, X_{j+1}^{(N)}, \ldots, X_m^{(N)}) \sim (S_1^{(N)}, \ldots, S_j^{(N)}, X_{j+1}^{(N)}, \ldots, X_m^{(N)}),
\]

where the last line follows because \((Y_1^{(N)}, X_{j+1}^{(N)}, \ldots, X_m^{(N)}) \sim (S_1^{(N)}, X_{j+1}^{(N)}, \ldots, X_m^{(N)})\) and because \(S_1^{(N)}, \ldots, S_m^{(N)}\) are independent of the other variables. By Theorem 8.10, \(\Phi_j^{(N)}\) is asymptotic to some \(\Phi_j \in \mathbb{T}M_{sa}^m\), and the objects \(\Phi, X, Y, S\) satisfy the analogous transport relations in the non-commutative setting. Now because each \(\Phi_j^{(N)} - \pi_x\) is \((\max(C, 1/c)^3 - 1)\) Lipschitz, we see that \(\Phi^{(N)} - \pi\) is \(m^{1/2}(\max(C, 1/c)^3 - 1)\) Lipschitz.

By Theorem 8.10 there is a map \(G_{j}^{(N)} : M_N(\mathbb{C})_{sa}^{m-j+1} \to M_N(\mathbb{C})_{sa}\) such that \((G_{j}^{(N)}(x_1, \ldots, x_m), x_{j+1}, \ldots, x_m)\) is the inverse of \((\Phi_j(x_1, \ldots, x_m), x_{j+1}, \ldots, x_m)\). Define \(\Psi_j^{(N)}\) by backwards induction by

\[
\Psi_j^{(N)}(x_j, \ldots, x_m) = G_j^{(N)}(x_j, \Psi_{j+1}(x_{j+1}, \ldots, x_m), \ldots, \Psi_m^{(N)}(x_m)).
\]

Then \(\Psi^{(N)} = (\Psi_1^{(N)}, \ldots, \Psi_m^{(N)})\) is the inverse of \(\Phi^{(N)}\). Since \(G_j^{(N)} - \pi\) is \((\max(C, 1/c)^3 - 1)\) Lipschitz, we can show by induction that \(\|\Psi_j^{(N)}\|_{\text{Lip}}\) is bounded by a constant depending only on \(c, C, m - j\), and which goes to zero as \(C \to 1\). Moreover, by Lemma 3.10, \(\Psi^{(N)}\) is asymptotic to some \(\Psi \in \mathbb{T}M_{sa}^m\).

This concludes the verification of (1) - (5). Now to prove (6), we apply Theorem 8.10 (5) and get

\[
\|\Phi^{(N)}(X^{(N)}) - X^{(N)}\|_L^2 \leq \sum_{j=1}^m \left(\|\Phi_j^{(N)}(X_j^{(N)}, \ldots, X_m^{(N)}) - X_j^{(N)}\|_L^2 \right)
\]

\[
= 2 \sum_{j=1}^m \left(2H_{j}^{(N)}(X_j^{(N)}, X_{j+1}^{(N)}, \ldots, X_m^{(N)}) \right)
\]

\[
= 2 \left(2H_{j}^{(N)}(X^{(N)}) - \frac{1}{2}E\|X^{(N)}\|_2^2 - \frac{m}{2}\log 2\pi \right)
\]

\[
= 2H_{j}^{(N)}(X^{(N)}),
\]
where we have applied the definition of $h_{\Phi}^{(N)}$ and the classical fact that $h^{(N)}$ is additive under conditioning. As before, because $\Phi^{(N)}(X^{(N)}) \sim S^{(N)}$, we see that $\|S^{(N)} - \Phi^{(N)}(S^{(N)})\|_{L^2} = \|\Phi^{(N)}(X^{(N)}) - X^{(N)}\|_{L^2}$. Finally, the second claim of (6) regarding the free case follows by taking the limit as $N \to \infty$.

Finally, to prove (7), recall that the map $\Phi_j$ is a special case of the map $\tilde{F}_{0,\infty}$ in Theorem 7.13. Thus, by applying Theorem 7.13 (4) in the case where $s = \infty$ and $s' = t = 0$, we obtain $\|\Phi_j(X_1, \ldots, X_m) - (X_j - \tau(X_j))\|_\infty \leq (\max(C, 1/c)^3 - 1) \max(C, 1/c) \Theta$. Moreover, the middle quantity in claim (7) equals the left hand side because $\Phi(X) \sim S$.

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. Anantharaman and S. Popa, *An Introduction to $II_1$ Factors*, 2017. preprint available at http://www.math.ucla.edu/~popa/Books/IIun-v10.pdf.

[3] P. Biane, M. Capitaine, and A. Guionnet, *Large deviation bounds for matrix Brownian motion*, Invent. Math., 152 (2003), pp. 433–459.

[4] S. G. Bobkov and M. Ledoux, *From Brunn-Minkowski to Braskamp-Lieb and to logarithmic Sobolev inequalities*, Geom. Funct. Anal., 10 (2000), pp. 1028–1052.

[5] R. Boutonnet and A. Carderi, *Maximal amenable von neumann subalgebras arising from maximal amenable subgroups*, Geometric and Functional Analysis, 25 (2015), pp. 1688–1705.

[6] R. Boutonnet and C. Houdayer, *Amenable absorption in amalgamated free product von neumann algebras*, Kyoto J. Math., 58 (2018), pp. 583–593.

[7] H. J. Brascamp and E. H. Lieb, *On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation*, Journal of Functional Analysis, 22 (1976), pp. 366–389.

[8] Y. Dabrowski, *A Laplace principle for Hermitian Brownian motion and free entropy I: the convex functional case*, arXiv e-prints, (2017).

[9] Y. Dabrowski, *A non-commutative path space approach to stationary free stochastic differential equations*, arXiv e-prints, (2010).

[10] Y. Dabrowski, A. Guionnet, and D. Shlyakhtenko, *Free transport for convex potentials*, arXiv e-prints, (2016).

[11] G. B. Folland, *Real Analysis: Modern Techniques and their Applications*, Pure and Applied Mathematics, John Wiley & Sons, New York, 2 ed., 1999.

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

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

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

[15] M. Hartglass and B. Nelson, *Free transport for interpolated free group factors*, Journal of Functional Analysis, 274 (2018), pp. 222 – 251.

[16] F. Hiai and Y. Ueda, *Free transportation cost inequalities for noncommutative multivariables*, Infinite Dimensional Analysis, Quantum Probability and Related Topics, 09 (2006), pp. 391–412.

[17] D. Jekel, *An elementary approach to free entropy theory for convex potentials*, arXiv e-prints, (2018).

[18] 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.
[19] E. S. Meckes, *The Random Matrix Theory of the Classical Compact Groups*, Cambridge Tracts in Mathematics, Cambridge University Press, 2019.

[20] E. S. Meckes and M. W. Meckes, *Spectral powers of random matrices*, Electron. Comm. Probab., 18 (2013).

[21] B. Nelson, *Free monotone transport without a trace*, Communications in Mathematical Physics, 334 (2015), pp. 1245–1298.

[22] B. Nelson, *Free transport for finite depth subfactor planar algebras*, Journal of Functional Analysis, 268 (2015), pp. 2586–2620.

[23] 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.

[24] F. Otto and C. Villani, *Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality*, Journal of Functional Analysis, 173 (2000), pp. 361–400.

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

[26] S. Popa, *Maximal injective subalgebras in factors associated with free groups*, Advances in Mathematics, 50 (1983), pp. 27–48.

[27] T. Tao, *An Introduction to Random Matrix Theory*, vol. 132 of Graduate Texts in Mathematics, American Mathematical Society, Providence, 2012.

[28] D. Voiculescu, *The analogues of entropy and of Fisher’s information measure in free probability, I*, Comm. Math. Phys., 155 (1993), pp. 71–92.

[29] D. Voiculescu, *The analogues of entropy and of Fisher’s information measure in free probability, III: Absence of Cartan subalgebras*, Geometric and Functional Analysis, 6 (1996), pp. 172–199.

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

[31] D. Voiculescu, *A strengthened asymptotic freeness result for random matrices with applications to free entropy*, International Mathematics Research Notices, 1998 (1998), pp. 41–63.

[32] D. Voiculescu, *The analogues of entropy and of fishers information measure in free probability theory, vi: Liberation and mutual free information*, Advances in Mathematics, 146 (1999), pp. 101–166.

[33] D. Voiculescu, *Free entropy*, Bulletin of the London Mathematical Society, 34 (2002), pp. 257–278.

[34] D. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables*, vol. 1 of CRM Monograph Series, American Mathematical Society, 1992.

Department of Mathematics, UCLA, Los Angeles, CA 90095

E-mail address: davidjekel@math.ucla.edu

URL: www.math.ucla.edu/~davidjekel/