Fractional Free Convolution Powers

DIMITRI SHLYAKHTENKO & TERENCE TAO

ABSTRACT. The extension $k \mapsto \mu^{\boxplus k}$ of the concept of a free convolution power to the case of non-integer $k \geq 1$ was introduced by Bercovici-Voiculescu and Nica-Speicher, and is related to the minor process in random matrix theory. In this paper, we give two proofs of the monotonicity of the free entropy and free Fisher information of the (normalized) free convolution power in this continuous setting, and also establish an intriguing variational description of this process.

1. INTRODUCTION

1.1. Integer-free convolution powers. In this paper we assume familiarity with noncommutative probability, particularly the concept of free independence (see, e.g., [37]).

In [31], Voiculescu introduced the notion of the *free convolution* $\mu \boxplus \nu$ of two compactly supported probability measures μ, ν on $\mathbb{R}$. There are multiple ways to define this operation. One is to define $\mu \boxplus \nu$ to be the law of $X + Y$, where X, Y are freely independent (real) noncommutative random variables with law μ, ν, respectively. Another is to define $\mu \boxplus \nu$ to be the asymptotic empirical spectral distribution of $A + B$ as $N \to \infty$, where A, B are classically independent bounded $N \times N$ random Hermitian matrices, each invariant under unitary conjugation, and whose empirical spectral distribution converges to μ, ν, respectively. A third way is to introduce the *Cauchy transform*[1] $G_\mu \colon \mathbb{C} \setminus \operatorname{supp}(\mu) \to \mathbb{C}$ of a compactly

[1]One can also write $G_\mu = -s_\mu$, where $s_\mu(z) := \int_{\mathbb{R}} d\mu(x)/(x - z)$ is the *Stieltjes transform* of μ; however, it will be slightly more convenient to work with the Cauchy transform instead of the Stieltjes transform to reduce the number of minus signs in our formulae.

Indiana University Mathematics Journal ©, Vol. 71, No. 6 (2022)

supported probability measure μ by the formula

$$G_\mu(z) := \int_{\mathbb{R}} \frac{\mathrm{d}\mu(x)}{z - x} \tag{1.1}$$

for $z \in \mathbb{C} \setminus \mathrm{supp}(\mu)$ (in particular, one has $G_\mu(z) = 1/z + O(1/|z|^2)$ as $|z| \to \infty$), and then define the *R-transform* $R_\mu(s)$ for su ciently small complex numbers s by requiring that

$$\frac{1}{G_\mu(z)} + R_\mu(G_\mu(z)) = z \tag{1.2}$$

for all su ciently large z. For su ciently small s one has the convergent Taylor expansion

$$R_\mu(s) = \sum_{n=0}^{\infty} \kappa_{n+1}(\mu) s^n,$$

where

$$\begin{aligned}
\kappa_1(\mu) &= \int_{\mathbb{R}} x \,\mathrm{d}\mu, \\
\kappa_2(\mu) &= \int_{\mathbb{R}} x^2 \,\mathrm{d}\mu - \Big(\int_{\mathbb{R}} x \,\mathrm{d}\mu\Big)^2, \\
\kappa_3(\mu) &= \int_{\mathbb{R}} x^3 \,\mathrm{d}\mu - 3\Big(\int_{\mathbb{R}} x \,\mathrm{d}\mu\Big)\Big(\int_{\mathbb{R}} x^2 \,\mathrm{d}\mu\Big)^2 + 2\Big(\int_{\mathbb{R}} x \,\mathrm{d}\mu\Big)^3 \\
&\cdots
\end{aligned}$$

are the *free cumulants* of μ.

Example 1.1. If μ_{sc} is the *semicircular distribution*

$$\mu_{\mathrm{sc}} := \frac{1}{2\pi}(4 - x^2)_+^{1/2} \,\mathrm{d}x,$$

then one easily verifies that $R_{\mu_{\mathrm{sc}}}(s) = s$, and so $\kappa_1(\mu_{\mathrm{sc}}) = 1$ and $\kappa_n(\mu_{\mathrm{sc}}) = 0$ for $n > 1$.

It is not di cult to see that a compactly supported probability measure μ is uniquely determined by its R-transform R_μ.

The free convolution $\mu \boxplus \nu$ is then the unique compactly supported measure for which $R_{\mu \boxplus \nu}(s) = R_\mu(s) + R_\nu(s)$ for all su ciently small s, or equivalently $\kappa_n(\mu \boxplus \nu) = \kappa_n(\mu) + \kappa_n(\nu)$ for all $n \geq 1$ (see, e.g., [37]); this is a commutative and associative operation on such measures. If k is a positive integer, one can then define $\mu^{\boxplus k} = \mu \boxplus \cdots \boxplus \mu$ to be the free convolution of k copies of μ, and one clearly has

$$R_{\mu^{\boxplus k}}(z) = k R_\mu(z) \tag{1.3}$$

for all sufficiently small s, or equivalently

$$\kappa_n(\mu^{\boxplus k}) = k\kappa_n(\mu) \tag{1.4}$$

for all $n \geq 1$. One can normalize these free convolutions by defining the dilates $\lambda_*\mu$ of a probability measure μ by a scaling factor $\lambda > 0$ to be the pushforward of μ by the dilation $x \mapsto \lambda x$ (thus, if μ is the law of a random variable X, then $\lambda_*\mu$ is the law of λX). One easily verifies the scaling laws

$$G_{\lambda_*\mu}(z) = \lambda^{-1} G_\mu(z/\lambda) \tag{1.5}$$

for all z outside of the support of μ, and

$$R_{\lambda_*\mu}(s) = \lambda R_\mu(\lambda s)$$

for all sufficiently small s (or equivalently, $\kappa_n(\lambda_*\mu) = \lambda^n \kappa_n(\mu)$ for all $n \geq 1$); thus, one has

$$R_{k_*^{-1/2}\mu^{\boxplus k}}(s) = k^{1/2} R_\mu(k^{-1/2}s).$$

Using this relation, Voiculescu [31] established the *free central limit theorem*: if μ is a compactly supported probability measure of mean zero and variance one, then the normalized free convolutions $k_*^{-1/2}\mu^{\boxplus k}$ converge in the vague topology to the semicircular distribution μ_{sc}.

In [33], Voiculescu also introduced the *free entropy*

$$\chi(\mu) := \int_{\mathbb{R}} \int_{\mathbb{R}} \log|x - y| \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) + \frac{3}{4} + \frac{1}{2}\log 2\pi$$

and the *free Fisher information*[2]

$$\Phi(\mu) := \frac{4\pi^2}{3} \int_{\mathbb{R}} \left(\frac{\mathrm{d}\mu}{\mathrm{d}x}\right)^3 \mathrm{d}x \tag{1.6}$$

for compactly supported probability measures μ (with the convention that, if μ is not absolutely continuous, $\Phi(\mu) = +\infty$); the two concepts are related by the derivative

$$\Phi(\mu) = 2\frac{\mathrm{d}}{\mathrm{d}t}\chi(\mu \boxplus \sqrt{t}_*\mu_{\mathrm{sc}})\,|_{t=0}$$

[2]There appears to be some inconsistency in terms of normalization constants in the definition of Φ between (and within) Voiculescu's papers [33, 35]. In particular, there appears to be an unfortunate typo in the statement and proof of Lemma 3.2 of [33], in which a factor of $\pi^2/2$ was left off. Our choice of normalization in the definition of Φ is compatible with its definition via the L^2 norm of a free conjugate variable as in [35] and differs by a factor of $4\pi^2/3$ from the definition in [33]. If μ is the semicircular law with second moment equal to 1 as in Example 1.1, then its free Fisher information equals 1 in our normalization.

and the closely associated integral formula

$$\chi(\mu) = \frac{1}{2}\int_0^\infty \left(\frac{1}{1+t} - \Phi(\mu \boxplus \sqrt{t}_* \mu_{\mathrm{sc}})\right) \mathrm{d}t + \frac{1}{2}\log 2\pi e. \tag{1.7}$$

In [25], it was shown that these quantities were monotone with respect to normalized free convolution powers in the sense that

$$\chi((k+1)_*^{-1/2}\mu^{\boxplus k+1}) \geq \chi(k_*^{-1/2}\mu^{\boxplus k}) \tag{1.8}$$

and

$$\Phi((k+1)_*^{-1/2}\mu^{\boxplus k+1}) \leq \Phi(k_*^{-1/2}\mu^{\boxplus k}) \tag{1.9}$$

for all compactly supported μ, and all $k \geq 1$. This was the free analog of a corresponding result proven in [3] for the Shannon entropy and classical Fisher information, answering a question of Shannon [24].

As is customary, if X is a real noncommutative random variable with law μ, we write $G_X := G_\mu$, $R_X := R_\mu$, $\kappa_n(X) := \kappa_n(\mu)$, $\Phi(X) := \Phi(\mu)$, and $\chi(X) := \chi(\mu)$.

1.2. Fractional free convolution powers. Observe that the righthand sides of (1.3), (1.4) make sense for any real number k. This raises the question of whether one can define fractional powers $\mu^{\boxplus k}$ for non-integer choices of k. This is indeed true, according to the statement below.

Proposition 1.2 (Existence of fractional free convolution powers). *Let μ be a compactly supported probability measure on $\mathbb{R}$, and let $k \geq 1$ be real. Then, there exists a unique compactly supported probability measure $\mu^{\boxplus k}$ on $\mathbb{R}$ such that*

$$R_{\mu^{\boxplus k}}(s) = kR_\mu(s) \tag{1.10}$$

for all su ciently small s, or equivalently,

$$\kappa_n(\mu^{\boxplus k}) = k^n \kappa_n(\mu)$$

for all $n \geq 1$.

Thus, for instance, $\mu_{\mathrm{sc}}^{\boxplus k} = k_*^{1/2}\mu_{\mathrm{sc}}$ for any $k \geq 1$.

Proposition 1.2 was first established for su ciently large k by Bercovici and Voiculescu [9], and then for all $k \geq 1$ by Nica and Speicher [23]; a complex analysis proof using subordination was given by Belinschi-Bercovici [6, 7] and Huang [17]. (See also the recent paper [5] for further study of the subordination functions associated with these measures, [17], [38] for further regularity and support properties of the $\mu^{\boxplus k}$, and [2], [26] for an extension to the case when k is a completely positive map and μ takes values in a C^*-algebra.)

From (1.10) and the invertibility of the R-transform, we have the semigroup law

$$(\mu^{\boxplus k})^{\boxplus \ell} = \mu^{\boxplus k\ell} \tag{1.11}$$

for any real $k, \ell \geq 1$, and similarly

$$\mu^{\boxplus k} \boxplus \mu^{\boxplus \ell} = \mu^{\boxplus k+\ell}.$$

Thus, one can now view $k \mapsto \mu^{\boxplus k}$ as a continuous one-parameter semigroup. There are also connections between fractional free convolution powers and free *multiplicative* convolution (see [8] and below).

The proof of Proposition 1.2 by Nica and Speicher [23] also gave the following free probability interpretation of such powers. Let $(\mathcal{A}, \tau)$ be a noncommutative probability space (that is to say, a complex associative unital $*$-algebra $\mathcal{A}$ equipped with a unital tracial positive linear functional τ), and let $p \in \mathcal{A}$ be a self-adjoint projection of trace $1/k$ for some $k \geq 1$ (thus, $p^* = p^2 = p$ and $\tau(p) = 1/k$). Then, we can form another noncommutative probability space $(\mathcal{A}_p, \tau_p)$ by defining $\mathcal{A}_p = [p\mathcal{A}p]$ to be a copy[3]

$$\mathcal{A}_p := \{[pXp] : X \in \mathcal{A}\}$$

of $p\mathcal{A}p := \{pXp : X \in \mathcal{A}\}$, and

$$\tau_p([pXp]) := k\tau(pXp) = k\tau(pX) = k\tau(Xp) \tag{1.12}$$

for any $X \in \mathcal{A}$. It is not di cult to verify that $(\mathcal{A}_p, \tau_p)$ is a noncommutative probability space. We have a “minor map” or “compression map” $\pi \colon \mathcal{A} \to \mathcal{A}_p$ defined by $\pi(X) := [pXp]$; this map is $*$-linear, surjective, and maps the unit 1 of $\mathcal{A}$ to the unit $1 = [p]$ of $\mathcal{A}_p$. The minor map π is not an algebra homomorphism nor is it trace preserving, but one does at least have homomorphism-like identities

$$\pi(X)\pi(Y) = \pi(pXpYp) = \pi(XpYp) = \pi(pXpY) = \pi(XpY) \tag{1.13}$$

for any $X, Y \in \mathcal{A}$, and from (1.12) we have

$$\tau_p(\pi(X)) := k\tau(pXp) \tag{1.14}$$

for any $X \in \mathcal{A}$.

Example 1.3. Let k be a rational number $k = N/M > 1$, $\mathcal{A} = M_N(\mathbb{C})$ be the space of $N \times N$ matrices with trace $\tau(X) := (1/N)\operatorname{Tr}(X)$, and

$$p = \begin{pmatrix} I_M & 0_{M\times N-M} \\ 0_{N-M\times M} & 0_{N-M\times N-M} \end{pmatrix}$$

[3]Thus, for instance, $[pXp][pYp] = [pXppYp] = [p(XpY)p]$ and $[pXp] + [pYp] = [pXp + pYp] = [p(X+Y)p]$. The brackets [ ] are a formal symbol, which we introduce in order to distinguish the algebraic structures of $\mathcal{A}_p$ from that of $\mathcal{A}$. In particular, the unit $1 = [p]$ of $\mathcal{A}_p$ needs to be distinguished from the non-unit p of $\mathcal{A}$, and the invertibility of an element $[pXp]$ of $\mathcal{A}_p$ does not imply the invertibility of the corresponding element pXp of $\mathcal{A}$.

be the orthogonal projection to the span of the first M standard basis vectors. Then, $\mathcal{A}_p$ can be identified with $M_M(\mathbb{C})$ (with trace $\tau_p(X) := (1/M)\operatorname{Tr}(X)$). With this identification, $\pi(X)$ is the upper left $M \times M$ minor of X.

We then have the following interpretation of fractional free convolution powers as a normalized free minor process.

Proposition 1.4 (Fractional free convolution powers from free minors). *If $(\mathcal{A}, \tau)$ is a noncommutative probability space, $k \geq 1$ is real, p is a real projection of trace $1/k$, and $X \in \mathcal{A}$ has some law μ and is freely independent of p, then $k\pi(X)$ has law $\mu^{\boxplus k}$. Thus,*

$$R_{k\pi(X)}(s) = kR_X(s),$$

or equivalently,

$$R_{\pi(X)}(s) = R_X(s/k) \tag{1.15}$$

for all sufficiently small s; in terms of free cumulants, this becomes

$$\kappa_n(\pi(X)) = k^{1-n}\kappa_n(X)$$

for $n \geq 1$.

Proof. See [23, Corollary 1.14]. For the convenience of the reader, we also give a self-contained proof in Appendix A. □

Remark 1.5. By the asymptotic free independence of independent unitarily invariant large matrices (see appendix to [23]), one can also define $\mu^{\boxplus k}$ for any real $k \geq 1$ as the asymptotic empirical distribution of the $M \times M$ random matrix $kA_{M\times M}$ as $N \to \infty$, where A is an $N \times N$ bounded random Hermitian matrix, invariant under unitary conjugation, whose empirical law converges to μ, $M := \lceil N/k \rceil$, and $A_{M\times M}$ is the upper left $M \times M$ minor of A. There is a similar interpretation of fractional free convolution powers in terms of the asymptotic distribution of large random Young tableaux, drawn uniformly from all tableaux of a given shape (see [10]).

One can investigate the dynamic of fractional free convolution powers as follows. From (1.10), (1.2) one has

$$\frac{1}{G_{\mu^{\boxplus k}(z)}} + kR_\mu(G_{\mu^{\boxplus k}}(z)) = z \tag{1.16}$$

for all $k \geq 1$ ranging in a compact set and all sufficiently large z. In particular, from the inverse function theorem, $G_{\mu^{\boxplus k}}(z)$ varies smoothly in k, z in this regime. Applying the first-order differential operator

$$\partial_z G_{\mu^{\boxplus k}}(z)\,\partial_k - \partial_k G_{\mu^{\boxplus k}}(z)\,\partial_z,$$

which annihilates $G_{\mu^{\boxplus k}}(z)$ as well as any autonomous function of $G_{\mu^{\boxplus k}(z)}$, to both sides of (1.16), we conclude that

$$(\partial_z G_{\mu^{\boxplus k}}(z)) R_\mu(G_{\mu^{\boxplus k}}(z)) = -\partial_k G_{\mu^{\boxplus k}}(z),$$

which when combined with (1.16) to eliminate the $R_\mu(G_{\mu^{\boxplus k}}(z))$ factor yields the Burgers-type equation

$$(k\,\partial_k + z\,\partial_z) G_{\mu^{\boxplus k}}(z) = \frac{\partial_z G_{\mu^{\boxplus k}}(z)}{G_{\mu^{\boxplus k}}(z)} \tag{1.17}$$

for $k \geq 1$ in a fixed compact region and sufficiently large z. From (1.5), we have

$$G_{k_*^{-1/2}\mu^{\boxplus k}}(z) = k^{1/2} G_{\mu^{\boxplus k}}(k^{1/2} z),$$

so after some calculation we can also write this equation in renormalized form as

$$\left(k\,\partial_k + \frac{1}{2} z\,\partial_z\right) G_{k_*^{-1/2}\mu^{\boxplus k}}(z) = \frac{\partial_z G_{k_*^{-1/2}\mu^{\boxplus k}}(z)}{G_{k_*^{-1/2}\mu^{\boxplus k}}(z)} + \frac{1}{2} G_{k_*^{-1/2}\mu^{\boxplus k}}(z). \tag{1.18}$$

This in turn gives a differential equation for $k_*^{-1/2}\mu^{\boxplus k}$ (see (3.6)).

It is now natural to ask whether the properties of integer-free convolution powers $\mu^{\boxplus k}$, $k \in \mathbb{N}$ extend to the fractional counterparts $\mu^{\boxplus k}$, $k \in \mathbb{R}$. For instance, the fractional convolution power allows us to make sense of the law of central limit sums $Y_k := k^{-1/2} \sum_{j=1}^k X_j$ of free independent, identically distributed copies X_j of a centered bounded random variable X. If X has law μ, then Y_N has law $k_*^{-1/2}\mu^{\boxplus k}$. The free central limit theorem states that the law of Y_k converges to the semicircle law as $k \to \infty$ along positive integers. It is easy to see that the R-transform proof of the free central limit theorem (see, e.g., [37]) shows also that $k_*^{-1/2}\mu^{\boxplus k}$ converges to the semicircle law as $k \to \infty$ along the positive reals.

Now we turn to the monotonicity of free entropy and free Fisher information, which is the first main result of our paper.

Theorem 1.6 (Monotonicity of free entropy and free Fisher information). *Let μ be a compactly supported finite probability measure. Then, $\chi(k_*^{-1/2}\mu^{\boxplus k})$ is monotone non-decreasing and $\Phi(k_*^{-1/2}\mu^{\boxplus k})$ is monotone non-increasing in k for real $k \geq 1$.*

Specializing to the case of integer k, we recover the previous results (1.8) and (1.9).

We prove this theorem in Section 2. Our argument relies on the characterization of fractional free convolution powers in Proposition 1.4, together with the fundamental fact that free independence is preserved by taking (free) minors. This proof also allows for an extension to several variables (see Theorem 2.4). In fact, as was shown to us by David Jekel, by applying a similar argument to the classical entropy and Fisher information of random matrix models, the argument can be adapted to a microstate setting, allowing one to also prove monotonicity for Voiculescu's multivariable microstates free entropy introduced in [34] (see Appendix B). Our argument shows that equality in Theorem 1.6 only holds when μ is a rescaled version of semicircular measure μ_{sc} (see Proposition 2.6).

By computing all of the quantities that appear explicitly or implicitly in the proof given in Section 2, we were able to extract a complex analytic proof of Theorem 1.6 using the di erential equation (1.17), at least if one assumes additional regularity on the original measure μ (we present a streamlined (but somewhat unmotivated) version of this proof in Section 3).

The fact that the flow (1.17) enjoys some monotonicity properties suggests that it has an interpretation as a gradient flow. We were not able to obtain such an interpretation, but we instead were able to find a (formal) *Lagrangian* interpretation of this flow, when viewed in "Gelfand-Tsetlin coordinates". In particular, let μ be a compactly supported probability measure on $\mathbb{R}$, let Δ denote the "Gelfand-Tsetlin pyramid"

$$\Delta := \{(s, y) : 0 < s < 1,\ 0 < y < s\},$$

and for any $(s, y) \in \mathbb{R}$, let $\lambda(s, y)$ denote the real number for which

$$\mu^{\boxplus 1/s}((-\infty, \lambda(s, y)/s]) = y/s. \tag{1.19}$$

Under suitable non-degeneracy assumptions on μ, $\lambda(s, y)$ will be well defined and vary smoothly with s, y. This function $\lambda(s, y)$ has the following random matrix interpretation. Let N be a large natural number parameter, and let A be a random Hermitian $N \times N$ matrix, invariant under unitary conjugation, and with empirical spectral distribution converging to μ as $N \to \infty$. Then, the $\lceil yN \rceil^{\mathrm{th}}$ smallest eigenvalue of the $\lceil sN \rceil \times \lceil sN \rceil$ minor will be concentrated around $\lambda(s, y)$. In Section 4 we establish the following result.

Theorem 1.7 (Variational formulation). *Formally, λ is a critical point of the Lagrangian*

$$\int_{\Delta} L(\partial_s \lambda, \partial_y \lambda)\, \mathrm{d}s\, \mathrm{d}y \tag{1.20}$$

where the Lagrangian density L is given by the formula

$$L(\lambda_s, \lambda_y) := \log \lambda_y + \log \sin \pi \frac{\lambda_s}{\lambda_y}. \tag{1.21}$$

We do not have a satisfactory interpretation of this Lagrangian density L. In [22] it is shown that random Gelfand-Tsetlin patterns formed by taking eigenvalues of successive minors asymptotically have the law of the Boutillier bead process [11], so it seems reasonable to conjecture[4] that the Lagrangian density $L(\lambda_s, \lambda_y)$ is proportional to the entropy of this process (with density proportional to $1/\lambda_y$, and drift velocity proportional to λ_s/λ_y).

2. Proof of Monotonicity

We now prove Theorem 1.6. We will rely on two main tools. The first is the fact that free independence is preserved by taking free minors, as follows.

Lemma 2.1. *Let $(\mathcal{A}, \tau)$ be a noncommutative probability space, and let $p \in \mathcal{A}$ be a real projection. If $B_1, \ldots, B_n \in \mathcal{A}$ are unital algebras such that $B_1, \ldots, B_n, p$ are free in $\mathcal{A}$, then $\pi(B_1), \ldots, \pi(B_n)$ are free in $\mathcal{A}_p$.*

Proof. See [23, Corollary 1.12]. ◻

Next, we recall the notion of *free score* (also called free conjugate variable) from [35]. If $(\mathcal{A}, \tau)$ is a noncommutative probability space, $X \in \mathcal{A}$, and B is a unital subalgebra of $\mathcal{A}$, we define the *free score* $J(X : B)$ of X relative to B (if it exists) to be the unique element in the $L^2(\tau)$ closure of the algebra $\mathrm{Alg}(X, B)$ generated by X and B with the property that

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\tau(ZP(X + \varepsilon Z, Y_1, \ldots, Y_n))\big|_{\varepsilon=0} = \tau(J(X : B)P(X, Y_1, \ldots, Y_n)) \tag{2.1}$$

for any $Y_1, \ldots, Y_n \in B$ and any noncommutative polynomial $P(X, Y_1, \ldots, Y_n)$ in $n + 1$ variables, where Z is a noncommutative random variable of mean zero and variance one that is freely independent of X, B (such a variable always exists if one is willing to extend the noncommutative space $(\mathcal{A}, \tau)$.) An equivalent definition (see [35, Proposition 3.4]) is that

$$\tau \otimes \tau(\partial P(X, Y_1, \ldots, Y_n)) = \tau(J(X : B)P(X, Y_1, \ldots, Y_n)) \tag{2.2}$$

where $\partial \colon \mathrm{Alg}(X, B) \to L^2(\tau \otimes \tau)$ is the unique derivation such that $\partial X = 1 \otimes 1$ and $\partial Y = 0$ for all $Y \in B$ (see [35]). If B is the trivial algebra $\mathbb{C}$, we abbreviate $J(X : \mathbb{C})$ as $J(X)$. It is known that the free Fisher information $\Phi(X)$ is finite if and only if the score exists, in which case [35]

$$\Phi(X) = ||J(X)||^2_{L^2(\tau)} = \tau(J(X)^2); \tag{2.3}$$

[4] Note added in proof: the recent calculations of local entropy (or "surface-tension") of the bead process in [29] (see also [19]) seem to strongly support this conjecture. We thank Istvan Prause for these references. Furthermore, it was pointed out to us by Vadim Gorin (private communication) that the random Gelfand-Tsetlin process is a continuous version of a random lozenge tiling [15], for which a variational description was provided in [13], and that the calculation in [29] can be viewed as a careful evaluation of the continuum limit of the theory in [13]. A very similar conjecture in the context of random Young tableaux has recently been proposed in [14].

indeed, this can be viewed as the "true" definition of the free Fisher information. Specializing (2.1) to the case $P = 1$, we see that the score, if it exists, is always trace-free:

$$\tau(J(X:B)) = 0. \tag{2.4}$$

We have the following basic fact from [35].

Lemma 2.2 (Free extensions do not affect free score). *Let $(\mathcal{A}, \tau)$ be a noncommutative probability space, let B, B' be unital subalgebras of $\mathcal{A}$, and $X \in \mathcal{A}$ be such that X, B are free from B'. The score $J(X : B)$ exists if and only if the score $J(X : B, B')$ exists, and the two scores are equal:*

$$J(X:B) = J(X:B,B').$$

Here, we use B, B' to denote the algebras generated by B and B'.

Proof. See [35, Proposition 3.6]. □

We now come to a basic identity.

Proposition 2.3 (Free score and minors). *Let $(\mathcal{A}, \tau)$ be a noncommutative probability space, let $p \in \mathcal{A}$ be a real projection of trace k^{-1} for some $k \geq 1$, let $X \in \mathcal{A}$, and let B be a unital subalgebra of $\mathcal{A}$. Assume that X, B are free of p and that the free score $J(X : B)$ exists. Then, the free score $J(\pi(X) : \pi(B))$ exists and is equal to*

$$J(\pi(X) : \pi(B)) = k\mathbb{E}(\pi(J(X:B))|\pi(X), \pi(B))$$

where $\mathbb{E}(\cdot|\pi(X), \pi(B))$ denotes the orthogonal projection (or conditional expectation) in $L^2(\tau_p)$ to the subalgebra of $\mathcal{A}_p$ generated by $\pi(X)$ and $\pi(B)$.

Proof. Let Z be a noncommutative random variable in $\mathcal{A}$ of mean zero and variance 1 that is free from X, B; such a variable exists after extending $\mathcal{A}$ if necessary. From Lemma 2.1, $k^{1/2}\pi(Z) \in \mathcal{A}_p$ has mean zero and variance 1, and is free from $\pi(X), \pi(B)$. By the definition of free score, it thus suffices to establish the identity

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}&\tau_p(k^{1/2}\pi(Z)P(\pi(X) + k^{1/2}\varepsilon\pi(Z), \pi(B)))\big|_{\varepsilon=0} \\ &= \tau_p(k\mathbb{E}(\pi(J(X:B))|\pi(X), \pi(B))P(\pi(X), \pi(B)))\end{aligned}$$

for any polynomial $P(\pi(X), \pi(B))$. By the chain rule we may cancel the factors of $k^{1/2}, k$, and as $P(\pi(X), \pi(B))$ lies in the range of the orthogonal projection $\mathbb{E}(\cdot|\pi(X), \pi(B))$ we may delete the projection; thus, we now need to show

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}&\tau_p(\pi(Z)P(\pi(X) + \varepsilon\pi(Z), \pi(B)))\big|_{\varepsilon=0} \\ &= \tau_p(\pi(J(X:B))P(\pi(X), \pi(B))).\end{aligned}$$

By using the definition of τ_p and π and the idempotent nature of p, this is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\tau(\pi(Z)P(p(X+\varepsilon Z)p, pBp))\big|_{\varepsilon=0} \\ = \tau(J(X:B)P(pXp, pBp)).$$

By Lemma 2.2, $J(X : B, p)$ exists and is equal to $J(X : B)$. Applying the definition of free score to the polynomial $P(p(X+\varepsilon Z)p, pBp)$, we obtain the claim. $\square$

Specializing this proposition to the case when $B = \mathbb{C}$, we conclude that (if the free score $J(X)$ exists) $J(\pi(X)) = k\mathbb{E}(\pi(J(X))|\pi(X))$, and hence by Pythagoras's theorem,

$$\Phi(\pi(X)) = ||J(\pi(X))||^2_{L^2(\tau_p)} \leq k^2||\pi(J(X))||^2_{L^2(\tau_p)}.$$

As $J(X)$ lies in the closure of the algebra generated by X, it is free of p; thus, by (1.14) and free independence,

$$||\pi(J(X))||^2_{L^2(\tau_p)} = k\tau(pJ(X)pJ(X)p) = k^{-1}\tau(J(X)^2).$$

We conclude the inequality $\Phi(\pi(X)) \leq k\Phi(X)$. Using the easily verified scaling

$$\Phi(\lambda X) = \lambda^{-2}\Phi(X) \tag{2.5}$$

for any $\lambda > 0$, we conclude that

$$\Phi(k^{1/2}\pi(X)) \leq \Phi(X) \tag{2.6}$$

whenever $J(X)$ exists. Clearly, this inequality also holds when $J(X)$ does not exist, since the righthand side is infinite. We thus have

$$\Phi(k_*^{-1/2}\mu^{\boxplus k}) \leq \Phi(\mu) \tag{2.7}$$

for any $k \geq 1$ and any compactly supported μ. Rescaling using (1.11), (2.5), we obtain the non-increasing nature of $\Phi(\chi(k_*^{-1/2}\mu^{\boxplus k})$. To obtain the corresponding monotonicity for free entropy, we use (1.7), (2.5) to compute

$$\begin{aligned}\chi(k_*^{-1/2}\mu^{\boxplus k}) &= \frac{1}{2}\int_0^\infty \left(\frac{1}{1+t} - \Phi(k_*^{-1/2}\mu^{\boxplus k} \boxplus \sqrt{t}_*\mu_{\mathrm{sc}})\right)\,\mathrm{d}t + \frac{1}{2}\log 2\pi e \\ &= \frac{1}{2}\int_0^\infty \left(\frac{1}{1+t} - \Phi(k_*^{-1/2}(\mu \boxplus \sqrt{t}_*\mu_{\mathrm{sc}})^{\boxplus k})\right)\,\mathrm{d}t + \frac{1}{2}\log 2\pi e,\end{aligned}$$

and the non-increasing nature of $\chi(k_*^{-1/2}\mu^{\boxplus k})$ then follows from the non-increasing nature of $\Phi(k_*^{-1/2}(\mu \boxplus \sqrt{t}_*\mu_{\mathrm{sc}})^{\boxplus k}))$ for each $t \geq 0$.

The above argument can also be generalized to obtain analogous monotonicity properties for the (non-microstate) free entropy and free Fisher information of several variables. We recall from [35] that the *relative free Fisher information* $\Phi^*(X : B)$ of a noncommutative real random variable $X \in \mathcal{A}$ relative to an algebra B is given by the formula

$$\Phi^*(X : B) = \|J(X : B)\|_{L^2(\tau)}^2 = \tau(J(X : B)^2),$$

and the non-microstate free Fisher information $\Phi^*(X_1, \dots, X_n)$ of a finite number of noncommutative real random variables $X_1, \dots, X_n \in \mathcal{A}$ is given by the formula

$$\Phi^*(X_1, \dots, X_n) := \sum_{i=1}^{n} \Phi^*(X_i : X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n). \tag{2.8}$$

The corresponding non-microstate free entropy $\chi^*(X_1, \dots, X_n)$ is then defined as

$$\chi^*(X_1, \dots, X_n) = \frac{1}{2} \int_0^\infty \left(\frac{n}{1+t} - \Phi^*(X_1 + t^{1/2} Z_1, \dots, X_n + t^{1/2} Z_n) \right) \mathrm{d}t + \frac{n}{2} \log 2\pi e$$

where $Z_1, \dots, Z_n$ are semicircular elements that are free from each other and from $X_1, \dots, X_n$.

Theorem 2.4 (Monotonicity for several variables). *If $X_1, \dots, X_n \in \mathcal{A}$, $k \geq 1$, and p is a real projection of trace $1/k$ that is free from $X_1, \dots, X_n$, one has*

$$\Phi^*(k^{1/2}\pi(X_1), \dots, k^{1/2}\pi(X_n)) \leq \Phi^*(X_1, \dots, X_n)$$

and

$$\chi^*(k^{1/2}\pi(X_1), \dots, k^{1/2}\pi(X_n)) \geq \chi^*(X_1, \dots, X_n).$$

We comment that an easy rescaling gives the equivalent forms

$$\Phi^*(\pi(X_1), \dots, \pi(X_n)) \leq k\Phi^*(X_1, \dots, X_n)$$

and

$$\chi^*(\pi(X_1), \dots, \pi(X_n)) \geq \chi^*(X_1, \dots, X_n) - \frac{n}{2} \log k.$$

of these inequalities.

Proof. It su ces to prove the former inequality, as the latter follows by repeating the previous arguments. From (2.8), it su ces to show that

$$\Phi^*(k^{1/2}\pi(X_i) : \pi(X_1), \dots, \pi(X_{i-1}), \pi(X_{i+1}), \dots, \pi(X_n)) \leq \Phi^*(X_i : X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$$

for each $i = 1, \ldots, n$. Let B be the algebra generated by $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$; then, we can rewrite this inequality as

$$k^{-1}||J(\pi(X_i) : \pi(X_1), \ldots, \pi(X_{i-1}), \pi(X_{i+1}), \ldots, \pi(X_n))||^2_{L^2(\tau_p)} \\ \leq ||J(X_i : B)||^2_{L^2(\tau)}.$$

From Proposition 2.3 and Pythagoras's theorem, we see that if $J(X_i : B)$ exists, then so does $J(\pi(X_i) : \pi(B))$ and

$$||J(\pi(X_i) : \pi(B))||^2_{L^2(\tau_p)} \leq k^2||\pi(J(X : B))||^2_{L^2(\tau_p)} = k||J(X : B)||^2_{L^2(\tau)},$$

where, as before, we use the fact that $J(X : B)$ is in the closure of the algebra generated by $X_1, \ldots, X_n$, and is hence free of p.

The algebra B' generated by

$$\pi(X_1), \ldots, \pi(X_{i-1}), \quad \pi(X_{i+1}), \ldots, \pi(X_n)$$

is a subalgebra of $\pi(B)$; hence, the score $J(\pi(X_i) : B')$ exists and is a projection of $J(\pi(X_i) : \pi(B))$. By a further application of Pythagoras, we conclude that

$$||J(\pi(X_i) : \pi(X_1), \ldots, \pi(X_{i-1}), \pi(X_{i+1}), \ldots, \pi(X_n))||^2_{L^2(\tau_p)} \\ \leq k||J(X : B)||^2_{L^2(\tau)},$$

and the claim follows. □

Remark 2.5. Appendix B establishes monotonicity of entropy for n-tuples for the so-called microstates free entropy χ, introduced by Voiculescu in [34].

Returning to the case of a single variable, we can analyze the above proof of monotonicity further to extract when equality occurs.

Proposition 2.6 (Characterization of equality). *Let μ be a compactly supported real probability measure with $\Phi(\mu) < \infty$, and let $k > 1$. If*

$$\Phi(k_*^{-1/2}\mu^{\boxplus k}) = \Phi(\mu),$$

then μ is the law of $\alpha + \beta u$ for some semicircular element u, real α, and $\beta > 0$.

Conversely, it is easy to see that if μ is the law of $\alpha + \beta u$ for a semicircular u, then $k_*^{-1/2}\mu^{\boxplus k}$ is the law of $k^{1/2}\alpha + \beta u$, so that $\Phi(k_*^{-1/2}\mu^{\boxplus k}) = \Phi(\mu)$. Using the representation (1.7) we see that we also have an analogous claim with the free Fisher information Φ replaced by the free entropy χ.

Proof. By translating μ (which does not a ect the free Fisher information of μ or $k_*^{-1/2}\mu^{\boxplus k}$), we may assume μ has mean zero. We can also assume μ is not a point mass as the free Fisher information is infinite in that case. Inspecting the proof of (2.7), we must have

$$\tau_p((k\mathbb{E}(\pi(J(X))|\pi(X)))^2) \le k^2\tau_p(\pi(J(X))^2),$$

and thus $\pi(J(X))$ lies in the L^2 closure of the algebra generated by $\pi(X)$. In particular, these two variables commute, so that

$$\tau_p(\pi(J(X))\pi(J(X))\pi(X)\pi(X)) = \tau_p(\pi(J(X))\pi(X)\pi(J(X))\pi(X))$$

(note that both sides are finite by Cauchy-Schwarz); by (1.12), we thus have

$$\tau(pJ(X)pJ(X)pXpX) = \tau(pJ(X)pXpJ(X)pX). \tag{2.9}$$

The variables $X, J(X)$ are free of p (since $J(X)$ lies in the closure of the algebra generated by X), and have trace zero by hypothesis and (2.4). Splitting p into the trace $1/k$ and the trace-free part $p' := p - 1/k$, we obtain 2^4 terms, but from free independence the only terms that survive are those that involve either zero or two copies of p', and in the latter case the p' terms need to be separated from each other cyclically by two of the $X, J(X)$ factors. In other words, we have

$$\begin{aligned}
&\tau(pJ(X)pJ(X)pXpX)\\
&\quad = k^{-4}\tau(J(X)^2X^2) + k^{-2}\tau(p'J(X)^2p'X^2)\\
&\qquad + k^{-2}\tau(J(X)p'J(X)Xp'X),
\end{aligned}$$

and similarly,

$$\begin{aligned}
&\tau(pJ(X)pXpJ(X)pX)\\
&\quad = k^{-4}\tau(J(X)XJ(X)X) + k^{-2}\tau(p'J(X)Xp'J(X)X)\\
&\qquad + k^{-2}\tau(J(X)p'XJ(X)p'X).
\end{aligned}$$

Applying these identities to (2.9) and noting that $J(X)$ commutes with X, we conclude that

$$\tau(p'J(X)^2p'X^2) = \tau(p'J(X)Xp'J(X)X).$$

From free independence, we see that

$$\tau(p'J(X)^2p'X^2) = \tau((p')^2)\tau(J(X)^2)\tau(X^2),$$

and similarly,

$$\tau(p'J(X)Xp'J(X)X) = \tau((p')^2)\tau(J(X)X)\tau(J(X)X).$$

Thus, we have

$$\tau(J(X)^2)\tau(X^2) = \tau(J(X)X)\tau(J(X)X),$$

which by the converse to Cauchy-Schwarz applied to the $L^2(\tau)$ inner product implies that $J(X)$ is a scalar multiple of X. To finish the proof, we can either invoke the equality case of the free Stam inequality in [35], or argue as follows. The identity $J(X) = \alpha X$ for a scalar α implies that, if μ is the law of X,

$$\begin{aligned}\int \frac{x}{z-x}\,\mathrm{d}\mu &= \tau(X(z-X)^{-1})\\ &= \alpha\tau\otimes\tau(\partial(z-X)^{-1})\\ &= \alpha\iint \frac{1/(z-s)-1/(z-t)}{s-t}\,\mathrm{d}\mu(s)\,\mathrm{d}\mu(t)\\ &= \alpha\iint \frac{1}{(z-s)(z-t)}\,\mathrm{d}\mu(s)\,\mathrm{d}\mu(t)\\ &= \alpha G_\mu(z)^2,\end{aligned}$$

where $G_\mu(z) = \int (1/(z-s))\,\mathrm{d}\mu(s)$ is the Cauchy transform. Using the identity

$$\int \frac{x}{z-x}\,\mathrm{d}\mu(x) = \int \frac{x-z}{z-x}\,\mathrm{d}\mu(x) + \int \frac{z}{z-x}\,\mathrm{d}\mu(x) = -1 + zG_\mu(z),$$

we deduce that $-1 + zG_\mu(z) = \alpha G_\mu^2(z)$. Solving this quadratic equation for G_μ (recalling that G_μ maps the upper half-plane to the lower half-plane and that μ is a probability measure) shows that μ is a scalar multiple of the semicircle law. ◻

It remains an interesting open problem to obtain an analogous characterization of equality in Theorem 2.4.

3. Complex Analytic Proof

We now give a direct proof of Theorem 1.6 using the di erential equation (1.18). To avoid technicalities we will work at a somewhat formal level, ignoring some questions of convergence and regularity, or justifying operations such as integration by parts, although we will still need to be careful when handling the limiting contribution of singular integrals involving kernels such as $1/(z-w)$ or $1/(z-w)^2$ when z, w are close. We will also assume that the measures $k_*^{-1/2}\mu^{\boxplus k}$ take the absolutely continuous form

$$d(k_*^{-1/2}\mu^{\boxplus k}) = f_k(x)\,\mathrm{d}x$$

for $k \geq 1$ and $x \in \mathbb{R}$, where f_k is compactly supported in x for each k and is assumed to obey su cient regularity[5] in k, x to justify the manipulations in the

[5]It is likely that these regularity hypotheses can be removed by a limiting argument to recover Theorem 1.6 in full generality. For instance, one can take advantage of the fact that if $\mathrm{d}\nu_\varepsilon = (1/\pi)(\varepsilon/(x^2+\varepsilon^2))\,\mathrm{d}x$ is the Cauchy distribution with parameter $\varepsilon > 0$, then $\mu * \nu_\varepsilon = \mu \boxplus \nu_\varepsilon$

sequel. We abbreviate

$$G_k(z) := G_{k_*^{-1/2}\mu^{\boxplus k}}(z) = \int_{\mathbb{R}} \frac{f_k(x)}{z-x}\,\mathrm{d}x. \tag{3.1}$$

As is well known, the limiting values

$$G_k(y+i0^\pm) := \lim_{\varepsilon\to 0^+} G_k(y \pm i\varepsilon)$$

for either choice of sign $\pm$ are then given (for su ciently regular f) by the Plemelj formulae

$$G_k(y+i0^\pm) = \pi H f_k(y) \mp \pi i f_k(y), \tag{3.2}$$

where

$$Hf(y) := \mathrm{p.v.}\frac{1}{\pi}\int_{\mathbb{R}} \frac{f(x)}{y-x}\,\mathrm{d}x \tag{3.3}$$

is the Hilbert transform of f. We recall some basic identities about this Hilbert transform.

Lemma 3.1 (Hilbert transform identities). *If $f : \mathbb{R} \to \mathbb{R}$ is compactly supported and su ciently regular, then one has the identities*

$$\int_{\mathbb{R}} f(y)Hf(y)\,\mathrm{d}y = 0,$$
$$\int_{\mathbb{R}} f(y)(Hf(y))^2\,\mathrm{d}y = \frac{1}{3}\int_{\mathbb{R}} f(y)^3\,\mathrm{d}y,$$
$$H(fHf) = \frac{(Hf)^2 - f^2}{2}.$$

Proof. Setting $G(z) := \int_{\mathbb{R}} (f(x)/(z-x))\,\mathrm{d}x$, by contour shifting we have

$$\int_{\mathbb{R}} G(y+i0^+)^2\,\mathrm{d}y = 0 \quad \text{and} \quad \int_{\mathbb{R}} G(y+i0^+)^3\,\mathrm{d}y = 0.$$

Substituting the Plemelj formula $G(y+i0^+) = \pi Hf - \pi i f$, and taking imaginary parts of both identities, we obtain the first two claims. For the final claim, we square (3.2) to conclude that

$$G_f(y+i0^+)^2 = \pi^2(Hf^2 - f^2) - 2\pi i f Hf$$

and compare this function against the function

$$G_{2\pi fHf}(y+i0^+) = 2\pi^2 H(fHf) - 2\pi i f Hf,$$

---

and $(\mu * \nu_\varepsilon)^{\boxplus k} = \mu^{\boxplus k} * \nu_{k\varepsilon}$, so one can apply the arguments in this section to the smooth measure $\mu * \nu_\varepsilon$ (after carefully taking into account that this measure is no longer compactly supported), and then taking limits as $\varepsilon \to 0$. We leave the details to the interested reader.

to conclude that the two holomorphic functions G_f^2, $G_{2\pi f Hf}$ (that both vanish at infinity) have identical imaginary parts on the half-plane, and are thus completely identical, giving the claim. □

Remark 3.2. From the identity $(y^n - x^n)/(y-x) = \sum_{j=0}^{n-1} x^j y^{n-1-j}$ for $n \geq 0$, we see that

$$\begin{aligned}\int_{\mathbb{R}} Hf(y) y^n f(y)\,\mathrm{d}y &= \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} f(x)f(y)\frac{y^n - x^n}{y-x}\,\mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{2\pi}\sum_{j=0}^{n-1}\left(\int_{\mathbb{R}} f(x)x^j\,\mathrm{d}x\right)\left(\int_{\mathbb{R}} f(y)y^{n-j-1}\,\mathrm{d}y\right).\end{aligned}$$

Thus, if X is a random variable with law $\mathrm{d}\mu = f(x)\,\mathrm{d}x$ for a compactly supported and su ciently regular f, then on comparing the above identity with (2.2) we see that the free score $J(X)$ is given by the formula $J(X) = 2\pi Hf(X)$, and thus from (2.3), we have

$$\Phi(X) = 4\pi^2 \int_{\mathbb{R}} f(y) Hf(y)^2\,\mathrm{d}y.$$

Lemma 3.1 shows that this formula is compatible with (1.6).

We abbreviate $f = f_1$ and $G = G_1$, and introduce the biholomorphic kernel $K(z,w)$ for $z, w \in \mathbb{C}\setminus\mathbb{R}$ by the formula

$$K(z,w) := \frac{1}{G(z)G(w)}\left(\frac{G(z)-G(w)}{z-w} + G(z)G(w)\right)^2, \tag{3.4}$$

noting that there is a removable singularity on the diagonal $z = w$. This kernel K emerged after lengthy but rather opaque calculations involving the quantities appearing in the previous section; it would be desirable to have a conceptual interpretation of this expression.

We can now derive Theorem 1.6 from the following three facts.

Proposition 3.3. *Formally at least, we have the following claims:*

(i) *We have*

$$\partial_k \Phi(k_*^{-1/2}\mu^{\boxplus k})\big|_{k=1} = \frac{8\pi^2}{3}\int_{\mathbb{R}} f^3\,\mathrm{d}x + 4\pi\int_{\mathbb{R}}\frac{Hf\,\partial_x f - f\,\partial_x Hf}{(Hf)^2 + f^2} f^2\,\mathrm{d}x.$$

(ii) *We have*

$$\begin{aligned}\lim_{\varepsilon\to 0^+}\sum_{\alpha,\beta\in\{-1,+1\}}\int_{\mathbb{R}}\int_{\mathbb{R}} f(x)f(y)K(x_{\alpha\varepsilon}, y_{\beta\varepsilon})\,\mathrm{d}x\,\mathrm{d}y \\ = -\frac{8\pi^2}{3}\int_{\mathbb{R}} f^3\,\mathrm{d}x - 4\pi\int_{\mathbb{R}}\frac{Hf\,\partial_x f - f\,\partial_x Hf}{(Hf)^2+f^2} f^2\,\mathrm{d}x,\end{aligned}\tag{3.5}$$

where $x_{\alpha\varepsilon} := x + i\alpha\varepsilon$ *and* $y_{\beta\varepsilon} := y + i\beta\varepsilon$.

(iii) *The kernel* $K(z, \bar{w})$ *is positive semi-definite; thus,*

$$\sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} K(z_j, \overline{z_k}) \geq 0$$

for all complex numbers $z_1, \ldots, z_n \in \mathbb{C} \setminus \mathbb{R}$ *and* $c_1, \ldots, c_n \in \mathbb{C}$.

Indeed, from (iii) we have that

$$\sum_{\alpha,\beta \in \{-1,+1\}} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x) f(y) K(x_{\alpha\varepsilon}, y_{\beta\varepsilon}) \,\mathrm{d}x \,\mathrm{d}y = \sum_{\alpha,\beta \in \{-1,+1\}} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x) f(y) K(x + i\alpha\varepsilon, \overline{y_{\beta\varepsilon}}) \,\mathrm{d}x \,\mathrm{d}y$$

is non-negative for any $\varepsilon > 0$. Meanwhile, from (i), (ii) we have

$$\lim_{\varepsilon \to 0^+} \sum_{\alpha,\beta \in \{-1,+1\}} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x) f(y) K(x_{\alpha\varepsilon}, y_{\beta\varepsilon}) \,\mathrm{d}x \,\mathrm{d}y = -\partial_k \Phi(k_*^{-1/2} \mu^{\boxplus k}) \,|_{k=1}$$

and hence we get $\partial_k \Phi(k_*^{-1/2} \mu^{\boxplus k}) \,|_{k=1} \leq 0$. From (1.11), (1.9) we then have $\partial_k \Phi(k_*^{-1/2} \mu^{\boxplus k}) \leq 0$ for all $k \geq 1$, giving the non-increasing nature of $\Phi(k_*^{-1/2} \mu^{\boxplus k})$; then, the non-decreasing nature of $\chi(k_*^{-1/2} \mu^{\boxplus k})$ follows from (1.7) as in the previous section.

It remains to establish the three claims in Proposition 3.3.

We begin with (i). From (1.6) and the chain rule we have

$$\partial_k \Phi(k_*^{-1/2} \mu^{\boxplus k})_{k=1} = 4\pi^2 \int_{\mathbb{R}} f_1(x)^2 \, \partial_k f_k(x) \,|_{k=1} \,\mathrm{d}x.$$

On the other hand, applying (1.18) at $z = x + i0^+$ and using (3.2) and the Cauchy-Riemann equations, we have

$$\left(k\,\partial_k + \frac{1}{2} x\,\partial_x\right)(\pi H f_k(x) - i\pi f_k(x)) = \frac{\partial_x (\pi H f_k(x) - i\pi f_k(x))}{\pi H f_k(x) - i\pi f_k(x)} + \frac{1}{2}(\pi H f_k(x) - i\pi f_k(x)),$$

which, upon taking imaginary parts, gives an integral di erential equation for f_k:

$$\left(k\,\partial_k + \frac{1}{2} x\,\partial_x\right) f_k = \frac{1}{\pi} \frac{H f_k \,\partial_x f_k - f_k \,\partial_x H f_k}{(H f_k)^2 + f_k^2} + \frac{1}{2} f_k. \tag{3.6}$$

Multiplying by f_k^2 and integrating, we obtain the claim (i) after a routine integration by parts.

We now skip ahead to (iii). The Schur product theorem asserts that the pointwise product of positive semi-definite kernels is again positive semi-definite. Since the rank one kernel $1/(G(z)G(\bar{w}))$ is clearly positive semi-definite, it thus su ces from (3.4) to show that the kernel

$$\frac{G(z)-G(\bar{w})}{z-\bar{w}}+G(z)G(\bar{w}) \tag{3.7}$$

is negative semi-definite. But from (3.1) and the identities $\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$ and $-\frac{1}{(z-x)(\bar{w}-x)} = \frac{1/(z-x)-1/(\bar{w}-x)}{z-\bar{w}}$, we see after a brief calculation that[6]

$$\begin{aligned}
-\int_{\mathbb{R}} f(x)\left(\frac{1}{z-x}-G(z)\right)\left(\frac{1}{\bar{w}-x}-G(\bar{w})\right)\mathrm{d}x& \\
=\frac{G(z)-G(\bar{w})}{z-\bar{w}}+G(z)G(\bar{w}).&
\end{aligned}$$

Since $f(x)$ is non-negative and the rank-one kernels

$$\left(\frac{1}{x-z}-G(z)\right)\left(\frac{1}{x-\bar{w}}-G(\bar{w})\right)$$

are positive semi-definite, the claim (iii) follows.

It remains to establish the identity (ii), which is the lengthiest calculation. We expand the lefthand side of (3.5) as $A_2+2A_1+A_0$, where

$$\begin{aligned}
A_2 &:= \lim_{\varepsilon\to 0^+}\sum_{\alpha,\beta\in\{-1,+1\}}\int_{\mathbb{R}}\int_{\mathbb{R}} f(x)f(y)\frac{(G(x_{\alpha\varepsilon})-G(y_{\beta\varepsilon}))^2}{G(x_{\alpha\varepsilon})G(y_{\beta\varepsilon})(x_{\alpha\varepsilon}-y_{\beta\varepsilon})^2}\,\mathrm{d}x\,\mathrm{d}y\\
A_1 &:= \lim_{\varepsilon\to 0^+}\sum_{\alpha,\beta\in\{-1,+1\}}\int_{\mathbb{R}}\int_{\mathbb{R}} f(x)f(y)\frac{G(x_{\alpha\varepsilon})-G(y_{\beta\varepsilon})}{x_{\alpha\varepsilon}-y_{\beta\varepsilon}}\,\mathrm{d}x\,\mathrm{d}y\\
A_0 &:= \lim_{\varepsilon\to 0^+}\sum_{\alpha,\beta\in\{-1,+1\}}\int_{\mathbb{R}}\int_{\mathbb{R}} f(x)f(y)G(x_{\alpha\varepsilon})G(y_{\beta\varepsilon})\,\mathrm{d}x\,\mathrm{d}y.
\end{aligned}$$

The quantity A_0 is easiest to compute, as it factorizes as

$$\left|\sum_{\pm}\int_{\mathbb{R}} f(x)G(x+i0^{\pm})\,\mathrm{d}x\right|^2.$$

Applying (3.2) and Lemma 3.1 we conclude that $A_0=0$.

[6]In other words, the quantity (3.7) is the negative of the covariance of $(z-X)^{-1}$ and $(\bar{w}-X)^{-1}$, where X is a random variable with law μ.

Now we turn to A_1. In order to compute the limit $\varepsilon \to 0^+$ it will be convenient to use integration by parts to replace the divergent-looking factor $1/(x_{\alpha\varepsilon} - y_{\beta\varepsilon})$ with a tamer singularity. More precisely, we write

$$\frac{1}{x_{\alpha\varepsilon} - y_{\beta\varepsilon}} = \partial_x \operatorname{Log}(x_{\alpha\varepsilon} - y_{\beta\varepsilon})$$

where we define the $\operatorname{Log} z$ away from the branch cut $(-\infty, 0)$ to be the branch of the complex logarithm with imaginary part in $(-\pi, \pi)$, and on the branch cut $(-\infty, 0)$ we define the averaged limiting value

$$\operatorname{Log}(-x) := \frac{1}{2}(\operatorname{Log}(-x + i0^+) + \operatorname{Log}(-x + i0^-)) = \log|x|.$$

The above identity breaks down when $x_{\alpha\varepsilon} - y_{\beta\varepsilon}$ vanishes, but this will not cause di culty because of the vanishing of the numerator $G(x_{\alpha\varepsilon}) - G(y_{\beta\varepsilon})$ in this case. Integrating by parts, we conclude that

$$\begin{aligned} A_1 = -\lim_{\varepsilon\to 0^+} \sum_{\alpha,\beta\in\{-1,+1\}} \int_{\mathbb{R}}\int_{\mathbb{R}} &\partial_x (f(x)f(y)(G(x_{\alpha\varepsilon}) - G(y_{\beta\varepsilon})) \\ &\times \operatorname{Log}(x_{\alpha\varepsilon} - y_{\beta\varepsilon})\, \mathrm{d}x\, \mathrm{d}y. \end{aligned}$$

As $\varepsilon \to 0^+$, the quantity $\operatorname{Log}(x_{\alpha\varepsilon} - y_{\beta\varepsilon})$ converges[7] to $\log|x-y| + i\pi 1_{y>x}\dfrac{\alpha-\beta}{2}$ for $x \neq y$, while from (3.2), $G(x_{\alpha\varepsilon}) - G(y_{\beta\varepsilon})$ converges to

$$\pi(Hf(x) - Hf(y) - i\alpha f(x) + i\beta f(y)).$$

For f su ciently regular, we conclude that

$$\sum_{\alpha,\beta\in\{-1,+1\}} \partial_x (f(x)f(y)(G(x_{\alpha\varepsilon}) - G(y_{\beta\varepsilon}))) \operatorname{Log}(x_{\alpha\varepsilon} - y_{\beta\varepsilon})$$

converges to

$$\begin{aligned} &4\pi\, \partial_x (f(x)f(y)(Hf(x) - Hf(y))) \log|x-y| \\ &\quad + 2\pi^2\, \partial_x (f(x)^2 f(y)) 1_{y>x} + 2\pi^2\, \partial_x (f(x)f(y)^2) 1_{y>x}, \end{aligned}$$

and hence

$$\begin{aligned} A_1 = -4\pi &\int_{\mathbb{R}}\int_{\mathbb{R}} \partial_x (f(x)f(y)(Hf(x) - Hf(y))) \log|x-y|\, \mathrm{d}x\, \mathrm{d}y \\ &- 2\pi^2 \int_{\mathbb{R}}\int_{\mathbb{R}} \partial_x (f(x)^2 f(y) + f(x)f(y)^2) 1_{y>x}\, \mathrm{d}x\, \mathrm{d}y. \end{aligned}$$

[7] Here, the indicator function $1_{y>x}$ is defined to equal 1 when $y > x$ and 0 otherwise.

Integrating by parts, we conclude that

$$A_1 = 4\pi \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{f(x)f(y)(Hf(x) - Hf(y))}{x - y}\, \mathrm{d}x\, \mathrm{d}y - 4\pi^2 \int_{\mathbb{R}} f^3\, \mathrm{d}x.$$

By symmetry and (3.3), we have

$$\begin{aligned}\int_{\mathbb{R}} \int_{\mathbb{R}} &\frac{f(x)f(y)(Hf(x) - Hf(y))}{x - y}\, \mathrm{d}x\, \mathrm{d}y \\ &= \pi \int_{\mathbb{R}} f(x)Hf(x)Hf(x)\, \mathrm{d}x + \pi \int_{\mathbb{R}} Hf(y)f(y)Hf(y)\, \mathrm{d}y,\end{aligned}$$

and hence by Lemma 3.1 and a brief calculation,

$$A_1 = -\frac{4\pi^2}{3} \int_{\mathbb{R}} f^3\, \mathrm{d}x.$$

We can compute A_2 in a similar fashion, writing

$$\frac{1}{(x_{\alpha\varepsilon} - y_{\beta\varepsilon})^2} = -\partial_x^2 \operatorname{Log}(x_{\alpha\varepsilon} - y_{\beta\varepsilon})$$

and integrating by parts twice to obtain

$$\begin{aligned}A_2 = -\lim_{\varepsilon \to 0^+} \sum_{\alpha,\beta \in \{-1,+1\}} \int_{\mathbb{R}} \int_{\mathbb{R}} \partial_x^2 &\left(f(x)f(y) \frac{(G(x_{\alpha\varepsilon}) - G(y_{\beta\varepsilon}))^2}{G(x_{\alpha\varepsilon})G(y_{\beta\varepsilon})} \right) \\ &\times \operatorname{Log}(x_{\alpha\varepsilon} - y_{\beta\varepsilon})\, \mathrm{d}x\, \mathrm{d}y.\end{aligned}$$

We expand

$$\begin{aligned}\frac{(G(x_{\alpha\varepsilon}) - G(y_{\beta\varepsilon}))^2}{G(x_{\alpha\varepsilon})G(y_{\beta\varepsilon})} &= \frac{G(x_{\alpha\varepsilon})}{G(y_{\beta\varepsilon})} - 2 + \frac{G(y_{\beta\varepsilon})}{G(x_{\alpha\varepsilon})} \\ &= \frac{G(x_{\alpha\varepsilon})\overline{G(y_{\beta\varepsilon})}}{|G(y_{\beta\varepsilon})|^2} - 2 + \frac{G(y_{\beta\varepsilon})\overline{G(x_{\alpha\varepsilon})}}{|G(x_{\alpha\varepsilon})|^2},\end{aligned}$$

and hence by (3.2) this quantity converges to

$$\begin{aligned}&\frac{(Hf(x) - i\alpha f(x))(Hf(y) + i\beta f(y))}{Hf(y)^2 + f(y)^2} - 2 \\ &\quad + \frac{(Hf(y) - i\beta f(y))(Hf(x) + i\alpha f(x))}{Hf(x)^2 + f(x)^2} \quad \text{as } \varepsilon \to 0^+.\end{aligned}$$

We can then evaluate A_2 as before (for f su ciently regular) as

$$A_2 = -4\int_{\mathbb{R}}\int_{\mathbb{R}} \partial_x^2\left(f(x)f(y)\left(\frac{Hf(x)Hf(y)}{Hf(y)^2+f(y)^2} - 2 + \frac{Hf(x)Hf(y)}{Hf(x)^2+f(x)^2}\right)\right) \times \log|x-y|\,\mathrm{d}x\,\mathrm{d}y$$
$$- 2\pi\int_{\mathbb{R}}\int_{\mathbb{R}} \partial_x^2\left(f(x)f(y)\left(\frac{f(x)Hf(y)}{Hf(y)^2+f(y)^2} - \frac{f(x)Hf(y)}{Hf(x)^2+f(x)^2}\right)\right) \times 1_{y>x}\,\mathrm{d}x\,\mathrm{d}y$$
$$- 2\pi\int_{\mathbb{R}}\int_{\mathbb{R}} \partial_x^2\left(f(x)f(y)\left(\frac{f(y)Hf(x)}{Hf(y)^2+f(y)^2} - \frac{f(y)Hf(x)}{Hf(x)^2+f(x)^2}\right)\right) \times 1_{y>x}\,\mathrm{d}x\,\mathrm{d}y.$$

From the fundamental theorem of calculus we have

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \partial_x^2\left(f(x)f(y)\frac{f(x)Hf(y)}{Hf(y)^2+f(y)^2}\right)1_{y>x}\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}} \partial_y(f(y)^2)\frac{f(y)Hf(y)}{Hf(y)^2+f(y)^2}\,\mathrm{d}y.$$

Using the distributional identity $\partial_x^2 1_{y>x} = \partial_y^2 1_{y>x}$ and integrating by parts repeatedly, we also have

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \partial_x^2\left(f(x)f(y)\frac{f(x)Hf(y)}{Hf(x)^2+f(x)^2}\right)1_{y>x}\,\mathrm{d}x\,\mathrm{d}y$$
$$= \int_{\mathbb{R}}\int_{\mathbb{R}} \partial_y^2\left(f(x)f(y)\frac{f(x)Hf(y)}{Hf(x)^2+f(x)^2}\right)1_{y>x}\,\mathrm{d}x\,\mathrm{d}y$$
$$= -\int_{\mathbb{R}} \partial_x(f(x)Hf(x))\frac{f(x)^2}{Hf(x)^2+f(x)^2}\,\mathrm{d}x.$$

Similar computations give

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \partial_x^2\left(f(x)f(y)\frac{f(y)Hf(x)}{Hf(y)^2+f(y)^2}\right)1_{y>x}\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}} \partial_y(f(y)Hf(y))\frac{f(y)^2}{Hf(y)^2+f(y)^2}\,\mathrm{d}y$$

and

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \partial_x^2\left(f(x)f(y)\frac{f(y)Hf(x)}{Hf(x)^2+f(x)^2}\right)1_{y>x}\,\mathrm{d}x\,\mathrm{d}y = -\int_{\mathbb{R}} \partial_x(f(x)^2)\frac{f(x)Hf(x)}{Hf(x)^2+f(x)^2}\,\mathrm{d}x.$$

Next, we integrate by parts, then use Lemma 3.1 and the fact that H commutes with derivatives to compute

$$\begin{aligned}
&\int_{\mathbb{R}}\int_{\mathbb{R}} \partial_x^2\left(f(x)f(y)\frac{Hf(x)Hf(y)}{Hf(y)^2+f(y)^2}\right)\log|x-y|\,\mathrm{d}x\,\mathrm{d}y\\
&\quad= -\int_{\mathbb{R}}\left(\text{p.v.}\int_{\mathbb{R}} \partial_x(f(x)f(y))\frac{Hf(x)Hf(y)}{Hf(y)^2+f(y)^2}\,\frac{\mathrm{d}x}{x-y}\right)\mathrm{d}y\\
&\quad= \pi\int_{\mathbb{R}} H\,\partial_y(fHf)(y)\frac{f(y)Hf(y)}{Hf(y)^2+f(y)^2}\,\mathrm{d}y\\
&\quad= \frac{\pi}{2}\int_{\mathbb{R}} \partial_y((Hf)^2-f^2)(y)\frac{f(y)Hf(y)}{Hf(y)^2+f(y)^2}\,\mathrm{d}y.
\end{aligned}$$

From $\partial_x^2\log|x-y| = \partial_y^2\log|x-y|$, integration by parts, and symmetry, we then have

$$\begin{aligned}
&\int_{\mathbb{R}}\int_{\mathbb{R}} \partial_x^2\left(f(x)f(y)\frac{Hf(x)Hf(y)}{Hf(x)^2+f(x)^2}\right)\log|x-y|\,\mathrm{d}x\,\mathrm{d}y\\
&\quad= \int_{\mathbb{R}}\int_{\mathbb{R}} \partial_y^2\left(f(x)f(y)\frac{Hf(x)Hf(y)}{Hf(x)^2+f(x)^2}\right)\log|x-y|\,\mathrm{d}x\,\mathrm{d}y\\
&\quad= \frac{\pi}{2}\int_{\mathbb{R}} \partial_x((Hf)^2-f^2)(x)\frac{f(x)Hf(x)}{Hf(x)^2+f(x)^2}\,\mathrm{d}x.
\end{aligned}$$

Finally,

$$\begin{aligned}
&\int_{\mathbb{R}}\int_{\mathbb{R}} \partial_x^2(f(x)f(y))\log|x-y|\,\mathrm{d}x\,\mathrm{d}y\\
&\quad= -\int_{\mathbb{R}}\left(\text{p.v.}\int_{\mathbb{R}} \partial_x(f(x)f(y))\,\frac{\mathrm{d}x}{x-y}\right)\mathrm{d}y\\
&\quad= \pi\int_{\mathbb{R}} H\,\partial_y f(y)f(y)\,\mathrm{d}y.
\end{aligned}$$

Putting all this together, we conclude that

$$\begin{aligned}
A_2 = 2\pi\int_{\mathbb{R}} &-\partial_x((Hf)^2-f^2)\frac{fHf}{(Hf)^2+f^2}\\
&+ 4(H\,\partial_x f)f - \partial_x((Hf)^2-f^2)\frac{fHf}{(Hf)^2+f^2}\,\mathrm{d}x\\
&- 2\pi\int_{\mathbb{R}} \partial_x(f^2)\frac{fHf}{(Hf)^2+f^2} + \partial_x(fHf)\frac{f^2}{(Hf)^2+f^2}\,\mathrm{d}x\\
&- 2\pi\int_{\mathbb{R}} \partial_x(f^2)\frac{fHf}{(Hf)^2+f^2} + \partial_x(fHf)\frac{f^2}{(Hf)^2+f^2}\,\mathrm{d}x,
\end{aligned}$$

which, upon applying the Leibniz rule, the commutativity of H and ∂_x, and collecting terms, simplifies to

$$A_2 = -4\pi \int_{\mathbb{R}} \frac{Hf\,\partial_x f - f\,\partial_x Hf}{(Hf)^2 + f^2} f^2 \,\mathrm{d}x,$$

and the claim (ii) follows.

4. Variational Formulation

We now prove Theorem 1.7. Our calculations here will be completely formal. Similar calculations appear in the recent paper [20, Section 5] in the context of studying random Young tableaux from a variational perspective[8].

We assume that the measures $\mu^{\boxplus k}$ are absolutely continuous with

$$\mathrm{d}\mu^{\boxplus k} = f_k(x)\,\mathrm{d}x$$

for $k \geq 1$. Applying (3.2) at $x + i0^+$ together with (1.17), we conclude that

$$(k\,\partial_k + x\,\partial_x)(\pi H f_k - \pi i f_k) = \frac{\partial_x (Hf_k - if_k)}{Hf_k - if_k} = \partial_x \log(Hf_k - if_k),$$

and hence on taking real and imaginary parts, we have

$$(k\,\partial_k + x\,\partial_x)Hf_k = \frac{1}{\pi}\,\partial_x \log((Hf_k)^2 + f_k^2)^{1/2}$$

and

$$(k\,\partial_k + x\,\partial_x)f_k = \frac{1}{\pi}\,\partial_x \arctan \frac{f_k}{Hf_k}$$

(where we use the branch of arctan taking values in $[0, \pi]$) and thus by the change of variables $k = 1/s$ and abbreviating $f := f_{1/s}$,

$$(-s\,\partial_s + x\,\partial_x)Hf = \frac{1}{\pi}\,\partial_x \log((Hf)^2 + f^2)^{1/2} \tag{4.1}$$

and

$$(-s\,\partial_s + x\,\partial_x)f = \frac{1}{\pi}\,\partial_x \arctan \frac{f}{Hf} \tag{4.2}$$

for $0 < s < 1$. We comment that this latter equation was also formally derived in [27] in the context of the derivative process, which is an averaged version of the minor process as established in [21, Lemma 1.16].

[8]The authors thank Istvan Prause for this reference.

Meanwhile, for $(s, y) \in \Delta$, we have from (1.19) that

$$\int_{-\infty}^{\lambda/s} f(x)\,\mathrm{d}x = \frac{y}{s}, \tag{4.3}$$

where we abbreviate $\lambda = \lambda(s, y)$ and $f = f_{1/s}$. If we di erentiate this in y using the fundamental theorem of calculus, we see that

$$\frac{\partial_y \lambda}{s} f(\lambda/s) = \frac{1}{s};$$

thus,

$$f(\lambda/s) = \frac{1}{\partial_y \lambda}. \tag{4.4}$$

If instead we di erentiate in s, we conclude that

$$\left(\frac{\partial_s \lambda}{s} - \frac{\lambda}{s^2}\right) f(\lambda/s) + \int_{-\infty}^{\lambda/s} \partial_s f(x)\,\mathrm{d}x = -\frac{y}{s^2},$$

and thus by (4.4) and multiplying by s, we have

$$\frac{\partial_s \lambda}{\partial_y \lambda} - \frac{\lambda}{s\,\partial_y \lambda} + \int_{-\infty}^{\lambda/s} s\,\partial_s f(x)\,\mathrm{d}x = -\frac{y}{s}.$$

By (4.2) we have

$$s\,\partial_s f = x\,\partial_x f - \frac{1}{\pi}\partial_x \arctan \frac{f}{Hf},$$

and hence by integration by parts and (4.3), (4.4),

$$\begin{aligned}\int_{-\infty}^{\lambda/s} s\,\partial_s f(x)\,\mathrm{d}x &= \frac{\lambda}{s} f(\lambda/s) - \frac{1}{\pi}\arctan\frac{f(\lambda/s)}{Hf(\lambda/s)} - \int_{-\infty}^{\lambda/s} f(x)\,\mathrm{d}x \\ &= \frac{\lambda}{s\,\partial_y \lambda} - \frac{1}{\pi}\arctan\frac{f(\lambda/s)}{Hf(\lambda/s)} - \frac{y}{s},\end{aligned}$$

and thus

$$\frac{\partial_s \lambda}{\partial_y \lambda} = \frac{1}{\pi}\arctan\frac{f(\lambda/s)}{Hf(\lambda/s)}.$$

We comment that this gives the pointwise inequalities $0 \le \partial_s \lambda \le \partial_y \lambda$, which in the random matrix formulation corresponds to the Cauchy interlacing inequalities. We rewrite this equation using (4.4) as

$$Hf(\lambda/s) = \frac{\cot(\pi\,\partial_s \lambda/\partial_y \lambda)}{\partial_y \lambda}, \tag{4.5}$$

and hence

$$\log((Hf)^2 + f^2)^{1/2}(\lambda/s) = \log \frac{\operatorname{cosec}(\pi\, \partial_s\lambda/\partial_y\lambda)}{\partial_y\lambda}.$$

Differentiating in y using the chain rule, we conclude

$$\frac{\partial_y\lambda}{s}(\partial_x \log((Hf)^2 + f^2)^{1/2})(\lambda/s) = \partial_y \log \frac{\operatorname{cosec}(\pi\, \partial_s\lambda/\partial_y\lambda)}{\partial_y\lambda},$$

and similarly by differentiating (4.5) in y, s we have

$$\frac{\partial_y\lambda}{s}(\partial_x Hf)(\lambda/s) = \partial_y \frac{\cot(\pi\, \partial_s\lambda/\partial_y\lambda)}{\partial_y\lambda}$$

and

$$(\partial_s Hf)(\lambda/s) + \left(\frac{\partial_s\lambda}{s} - \frac{\lambda}{s^2}\right)(\partial_x Hf)(\lambda/s) = \partial_s \frac{\cot(\pi\, \partial_s\lambda/\partial_y\lambda)}{\partial_y\lambda},$$

so that

$$(\partial_s Hf)(\lambda/s) = \left(\partial_s - \frac{\partial_s\lambda}{\partial_y\lambda}\,\partial_y + \frac{\lambda}{s\partial_y\lambda}\,\partial_y\right)\frac{\cot(\pi\, \partial_s\lambda/\partial_y\lambda)}{\partial_y\lambda}.$$

Inserting these identities into (4.1) evaluated at λ/s, we obtain a differential equation for λ in the variables s, y:

$$-s\left(\partial_s - \frac{\partial_s\lambda}{\partial_y\lambda}\,\partial_y\right)\frac{\cot(\pi\, \partial_s\lambda/\partial_y\lambda)}{\partial_y\lambda} = \frac{1}{\pi}\frac{s}{\partial_y\lambda}\,\partial_y \log \frac{\operatorname{cosec}(\pi\, \partial_s\lambda/\partial_y\lambda)}{\partial_y\lambda}.$$

Multiplying by $-\pi\, \partial_y\lambda/s$, we obtain

$$(\partial_y\lambda\, \partial_s - \partial_s\lambda\, \partial_y)\frac{\pi\cot(\pi\, \partial_s\lambda/\partial_y\lambda)}{\partial_y\lambda} = \partial_y\left(\log \partial_y\lambda + \log\sin\left(\pi\frac{\partial_s\lambda}{\partial_y\lambda}\right)\right)$$

which we write in divergence form using (1.21) as

$$\begin{aligned}\partial_s &\left(\partial_y\lambda\frac{\pi\cot(\pi\, \partial_s\lambda/\partial_y\lambda)}{\partial_y\lambda}\right) + \\ &+ \partial_y\left(-\partial_s\lambda\frac{\pi\cot(\pi\, \partial_s\lambda/\partial_y\lambda)}{\partial_y\lambda} - L(\partial_s\lambda, \partial_y\lambda)\right) = 0.\end{aligned}$$

Since the partial derivatives of

$$L(\lambda_s, \lambda_y) := \log \lambda_y + \log \sin \left(\pi \frac{\lambda_s}{\lambda_y} \right)$$

are given by

$$L_{\lambda_s} = \frac{\pi}{\lambda_y} \cot \left(\pi \frac{\lambda_s}{\lambda_y} \right)$$

and

$$L_{\lambda_y} = \frac{1}{\lambda_y} - \frac{\pi \lambda_s}{\lambda_y^2} \cot \left(\pi \frac{\lambda_s}{\lambda_y} \right),$$

we can rewrite the above equation as

$$\partial_s (\partial_y \lambda L_{\lambda_s}(\partial_s \lambda, \partial_y \lambda)) + \partial_y (\partial_y \lambda L_{\lambda_y}(\partial_s \lambda, \partial_y \lambda) - L(\partial_s \lambda, \partial_y \lambda)) = 0.$$

From the chain rule we have

$$\partial_y L(\partial_s \lambda, \partial_y \lambda) = (\partial_y \, \partial_y \lambda) L_{\lambda_y}(\partial_s \lambda, \partial_y \lambda) + (\partial_s \, \partial_y \lambda) L_{\lambda_s}(\partial_s \lambda, \partial_y \lambda);$$

inserting this into the previous equation and using the product rule and then cancelling the $\partial_y \lambda$ factor, we conclude that

$$\partial_s L_{\lambda_s}(\partial_s \lambda, \partial_y \lambda) + \partial_y L_{\lambda_y}(\partial_s \lambda, \partial_y \lambda) = 0,$$

which is the Euler-Lagrange equation for the Lagrangian (1.20), and the claim follows.

Appendix A. Fractional Free Convolution Powers from the Minor Process

In this appendix we prove Proposition 1.4. Let the hypotheses be as in that proposition; our task is to establish (1.15). We follow the arguments from [30, Section 2.5.4]. Using the GNS construction we may assume that $\mathcal{A}$ is a von Neumann algebra of bounded operators.

We begin with some algebraic identities. For any noncommutative variable E of operator norm less than 1, define the transform

$$\Psi(E) := (1 - E)^{-1} - 1 = E + E^2 + E^3 + \cdots .$$

Lemma A.1 (Algebraic identities).

(i) *If $Z \in \mathcal{A}$ is su ciently small (in operator norm), then*

$$(1 + \pi(Z))^{-1} = \pi((1 + pZ)^{-1}).$$

(ii) *If $Y \in \mathcal{A}$, and $E \in \mathcal{A}$ is su ciently small (in operator norm) depending on Y, then $\pi(Y\Psi(E)) = \Psi(E_{pY})$, where*

$$E_{pY} := \pi(1 - (1 - E)(1 - (1 - pY)E)^{-1}). \tag{A.1}$$

Proof. If Z is small enough, then $1 + pZ$ is invertible by Neumann series, and by (1.13) we have

$$\begin{aligned} \pi((1 + pZ)^{-1})(1 + \pi(Z)) &= \pi((1 + pZ)^{-1}p(1 + Z)p) \\ &= \pi((1 + pZ)^{-1}(1 + pZ)p) \\ &= \pi(p) \\ &= 1, \end{aligned}$$

giving (i). For (ii), we apply (i) with $Z := Y\Psi(E)$ to conclude that

$$(1 + \pi(Y\Psi(E)))^{-1} = \pi((1 + pY\Psi(E))^{-1}).$$

Since $1 + pY\Psi(E) = (1 - (1 - pY)E)(1 - E)^{-1}$, we see from (A.1) that

$$\pi((1 + pY\Psi(E))^{-1}) = 1 - E_{pY},$$

and the claim (ii) then follows after some rearranging. $\square$

Now set $Y = k$. From (A.1) and Neumann series we have

$$E_{kp} = [pEp] - \sum_{n=1}^{\infty} [p((1 - kp)E)^n p] - [pE((1 - kp)E)^n p],$$

when E is su ciently small in operator norm. As $1 - kp$ has trace zero, we conclude on taking traces that $\tau_p(E_{kp}) = 0$ whenever E has trace zero, is su ciently small in operator norm, and is freely independent from p.

This has the following consequence. If z is su ciently large and $s = G_\mu(z)$, then from (1.1) we have $s = \tau((z - X)^{-1})$, and thus

$$(z - X)^{-1} = s(1 - E(s)) \tag{A.2}$$

for some trace zero element $E(s) \in \mathcal{A}$, which will be small when z is large. Since X is freely independent of p, $E(s)$ is also. Meanwhile, from (1.2) one has $R_\mu(s) + \frac{1}{s} = z$, which when combined with (A.2) and rearranging gives

$$X = R_\mu(s) - \frac{1}{s}\Psi(E(s))$$

for all su ciently small s. Applying $k\pi$, we conclude that

$$k\pi(X) = kR_\mu(s) - \frac{1}{s}\pi(k\Psi(E(s))).$$

Applying Lemma A.1 (ii), and (1.12), we conclude that

$$k\pi(X) = kR_\mu(s) - \frac{1}{s}\Psi(E_{kp}(s))$$

where $E_{kp}(s) \in \mathcal{A}_p$ obeys (1.12). If we set $z' := kR_\mu(s) + 1/s$, we can rearrange this as $(z' - k\pi(X))^{-1} = s(1 - E_{kp}(s))$, and then upon taking traces, we conclude

$$s = \tau_p((z' - k\pi(X))^{-1}) = G_{k\pi(X)}(z).$$

From (1.2) we then conclude that $R_{k\pi(X)}(s) = kR_\mu(s)$ for all su ciently small s, giving the claim (1.15).

Appendix B. Monotonicity for Microstates Free Entropy

by David Jekel

In this section, we adapt the free probability proof of Theorem 2.4 to the microstates setting to obtain an analog of that theorem for Voiculescu's *microstates free entropy* χ, introduced in [34]. The main result is as follows.

Theorem B.1 (Monotonicity of microstate free entropy). *Let $k \in [1, \infty)$. Let $(\mathcal{A}, \tau)$ be a noncommutative probability space, let $X \in \mathcal{A}_{\mathrm{sa}}^n$ (i.e., X is a tuple $(X_1, \ldots, X_n)$ of self-adjoint elements of $\mathcal{A}$), and let p be a projection of trace $1/k$ freely independent of X. Let $\pi \colon \mathcal{A} \to [p\mathcal{A}p]$ be the compression map, and let $\Pi := k^{1/2}\pi$ be the normalized compression. Then, $\chi(\Pi(X)) \geq \chi(X)$.*

The first step in the proof is to reformulate χ in terms of the classical entropy of random matrix approximations of X. The second step is to apply a similar argument as in Section 2 for the *classical* entropy and score functions, which results in an approximate version of (2.7) for the minors of the random matrix models.

We first set up all the notation that we need. We begin by recalling various classical information theory notions in the general context of random variables taking values in finite-dimensional inner product spaces[9] H.

Definition B.2 (Classical information theory concepts). Let H be a finite-dimensional inner product space, with inner product $\langle u, v\rangle_H$ and norm $\|u\|_H$. We let ν_H be the Haar measure canonically associated with H (thus ν assigns unit mass to the unit cube generated by any orthonormal basis in H). We have the following:

(i) If X is square integrable, then the *total variance* $\mathrm{Var}_H(X)$ is given by $\mathrm{Var}_H(X) := \mathbb{E}\|X - \mathbb{E}X\|_H^2$.

[9] All inner product spaces here will be over the reals.

(ii) If X is a (classical) random variable taking values in H with absolutely continuous law $\mathrm{d}\mu = \rho\,\mathrm{d}\nu_H$, the *classical* (*di erential*) *entropy* $h_H(X)$ of X is given by $-\int_H \rho \log \rho \,\mathrm{d}\nu_H$. If there is no density ρ, the entropy is defined to equal $-\infty$. We also write $h_H(\mu)$ for $h_H(X)$.

(iii) A *standard Gaussian random variable* in H is a Gaussian variable Z_H of mean zero and identity covariance matrix in the sense that

$$\mathbb{E}\langle u, Z_H\rangle_H \langle Z_H, v\rangle_H = \langle u, v\rangle_H \quad \text{for all } u, v \in H;$$

equivalently, Z_H has law $(2\pi)^{-\dim(H)/2} e^{-\|u\|_H^2/2}\,\mathrm{d}\nu_H$.

(iv) If X is a random variable taking values in H, then a random variable $J_H(X)$ is said to be a *classical score* of X (relative to the inner product H) if it lies in the L^2 closure of the algebra generated by X, and

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathbb{E}\langle f(X+\varepsilon Z_H), Z_H\rangle_H \big|_{\varepsilon=0} = \mathbb{E}\langle J_H(X), f(X)\rangle_H \tag{B.1}$$

for any $f \in C_c^\infty(H; H)$, where Z_H is a standard Gaussian variable in H (classically) independent of X (cf. (2.1)). Note that if the classical score exists, it is unique.

(v) The *classical Fisher information* of X is $\mathcal{I}_H(X) := \mathbb{E}\|J_H(X)\|_H^2$ if a classical score $J_H(X)$ exists, and $\mathcal{I}_H(X) = +\infty$ otherwise.

Example B.3. If H is a standard Euclidean space $\mathbb{R}^d$ and if X has a C^1 probability density ρ, then the classical score is given explicitly by $J_{\mathbb{R}^d}(X) = -\nabla\rho(X)/\rho(X)$ provided the latter is in L^2. The classical Fisher information is then equal to $\mathcal{I}_{\mathbb{R}^d}(X) = \int_{\mathbb{R}^d} |\nabla\rho|^2/\rho$.

Example B.4. If H is a d-dimensional Hilbert space, Z_H is a standard Gaussian variable in H and $t > 0$; then,

$$\mathrm{Var}_H(t^{1/2}Z_H) = td, \qquad h_H(t^{1/2}Z_H) = \frac{d}{2}\log(2\pi e t),$$
$$J_H(t^{1/2}Z_H) = t^{-1/2}Z_H, \qquad \mathcal{I}_H(t^{1/2}Z_H) = \frac{d}{t}.$$

Thus, we see that with this "standard" choice of normalization, most quantities scale linearly with the dimension d. Later on we shall switch to a "microstate" choice of normalization that is better suited for passing to the free probability limit $d \to \infty$.

We now recall some standard properties of the above notions.

Lemma B.5 (Standard classical information theory facts). *Let H be a finite-dimensional inner product space of some dimension d, with canonical Haar measure ν_H. Let X be a random variable taking values in H with law μ, and let Z_H be a standard Gaussian random variable in H classically independent of X.*

(i) (*Entropy controlled by variance*) *If X has finite variance, then*

$$-\infty < h_H(X) \leq \frac{d}{2}\log(2\pi e \operatorname{Var}_H(X)/d). \tag{B.2}$$

In particular, each multiple $t^{1/2}Z_H$ of Z_H maximizes the entropy amongst all variables of the same variance.

(ii) (*Entropy controlled by partition*) *Let $(S_j)_{j=1}^{\infty}$ be a measurable partition of H. Then,*

$$h_H(\mu) \leq \sum_{j=0}^{\infty} \mu(S_j)\log\nu_H(S_j) - \sum_{j=0}^{\infty}\mu(S_j)\log\mu(S_j).$$

(iii) (*Shannon inequality*) *If Y is a random variable in H classically independent of X, then $h_H(X+Y) \geq h_H(X), h_H(Y)$.*

(iv) (*Stein identity*) *If $t > 0$, then the score $J_H(X + t^{1/2}Z_H)$ exists and is given by*

$$J_H(X + t^{1/2}Z_H) = \mathbb{E}[t^{-1/2}Z_H | X + t^{1/2}Z_H]. \tag{B.3}$$

In particular, the Fisher information $\mathcal{I}_H(X + t^{1/2}Z_H)$ is finite.

(v) (*de Bruijn identity*) *If $0 < t_0 < t_1$, we have the identity*

$$h_H(X + t_1^{1/2}Z_H) - h_H(X + t_0^{1/2}Z_H) = \frac{1}{2}\int_{t_0}^{t_1} \mathcal{I}_H(X + t^{1/2}Z_H)\,\mathrm{d}t. \tag{B.4}$$

Proof. By using an orthonormal basis one can identify H with a standard Euclidean space $\mathbb{R}^d$. The facts (i), (iv), (v) are then well known and can be found in, for instance, [28], and (iii) is similarly well known [24]. Now, we prove (ii). If μ does not have a density, then $h_H(\mu) = -\infty$ and hence the claim is trivially true. Assume that μ has a density ρ. Then,

$$h_H(\mu) = -\sum_{j=0}^{\infty}\int_{S_j} \rho\log\rho\,\mathrm{d}\nu_H.$$

We apply Jensen's inequality to the concave function $-t\log t$ and the probability measure that is the pushforward by ρ of the uniform distribution on S_j, and thus obtain

$$\begin{aligned}\frac{1}{\nu_H(S_j)}\int_{S_j} -\rho\log\rho\,\mathrm{d}x &\leq -\left(\frac{1}{\nu_H(S_j)}\int_{S_j}\rho\,\mathrm{d}\nu_H\right)\log\left(\frac{1}{\nu_H(S_j)}\int_{S_j}\rho\,\mathrm{d}\nu_H\right) \\ &= -\frac{\mu(S_j)}{\nu_H(S_j)}\log\frac{\mu(S_j)}{\nu_H(S_j)},\end{aligned}$$

which produces the desired estimate. $\square$

We will primarily work in the inner product space $M_N(\mathbb{C})^n_{\mathrm{sa}}$ of n-tuples $X = (X_1, \dots, X_N)$ of $N \times N$ Hermitian matrices, with inner product

$$\langle X, Y \rangle_{M_N(\mathbb{C})^n_{\mathrm{sa}}} := \sum_{j=1}^{n} \mathrm{tr}_N(X_j Y_j)$$

defined using the normalized trace $\mathrm{tr}_N := \frac{1}{N}\,\mathrm{Tr}$. Thus, in particular we have the normalized Frobenius norms

$$\|X\|^2_{M_N(\mathbb{C})^n_{\mathrm{sa}}} = \sum_{j=1}^{n} \mathrm{tr}_N(X_j^2) = \frac{1}{N}\sum_{j=1}^{n} \mathrm{Tr}(X_j^2).$$

This is an nN^2-dimensional inner product space. If $Z_{M_N(\mathbb{C})^n_{\mathrm{sa}}}$ is a standard Gaussian random variable in $M_N(\mathbb{C})^n_{\mathrm{sa}}$, then $Z_{M_N(\mathbb{C})^n_{\mathrm{sa}}}$ is an ensemble of n (classically) independent matrices, with each entry having variance N, for a total variance of $\mathrm{Var}_{M_N(\mathbb{C})^n_{\mathrm{sa}}}(Z) = nN^2$. To facilitate taking limits as $N \to \infty$, it is convenient to introduce the normalized Gaussian variable

$$Z^{(N)} := \frac{1}{N} Z_{M_N(\mathbb{C})^n_{\mathrm{sa}}}.$$

Thus, $Z^{(N)}$ is an ensemble of n (classically) independent GUE matrices, with each entry having variance $1/N$, converging to an n-tuple of freely independent semicircular random variables as $N \to \infty$ [32]; we refer to such random variables $Z^{(N)}$ as *GUE tuples* in $M_N(\mathbb{C})^n_{\mathrm{sa}}$. One easily computes the total variance

$$\mathrm{Var}_{M_N(\mathbb{C})^n_{\mathrm{sa}}}(t^{1/2} Z^{(N)}) = tn,$$

classical entropy

$$h_{M_N(\mathbb{C})^n_{\mathrm{sa}}}(t^{1/2} Z^{(N)}) = \frac{nN^2}{2}\log(2\pi e t) - nN^2 \log N,$$

classical score

$$J_{M_N(\mathbb{C})^n_{\mathrm{sa}}}(t^{1/2} Z^{(N)}) = N^2 t^{-1/2} Z^{(N)},$$

and classical Fisher information

$$\mathcal{I}_{M_N(\mathbb{C})^n_{\mathrm{sa}}}(t^{1/2} Z^{(N)}) = \frac{nN^4}{t}$$

of multiples $t^{1/2} Z^{(N)}$ of GUE tuples for $t > 0$. Note that most of the quantities on the righthand side depend on the matrix dimension N, which is undesirable for the purposes of extracting a meaningful limit as $N \to \infty$. To facilitate the process of taking such a limit, we therefore introduce the *normalized classical entropy*

$$h^{(N)}(X) := \frac{1}{N^2} h_{M_N(\mathbb{C})^n_{\mathrm{sa}}}(X) + n \log N,$$

the *normalized classical score*

$$J^{(N)}(X) := \frac{1}{N^2} J_{M_N(\mathbb{C})^n_{\mathrm{sa}}}(X),$$

and the *normalized classical Fisher information*

$$\mathcal{I}^{(N)}(X) := \mathbb{E}||J^{(N)}(X)||_H^2 = \frac{1}{N^4} \mathcal{I}_{M_N(\mathbb{C})^n_{\mathrm{sa}}}(X)$$

while leaving the variance unchanged:

$$\mathrm{Var}^{(N)}(X) := \mathrm{Var}_{M_N(\mathbb{C})^n_{\mathrm{sa}}}(X).$$

Thus, for instance, we have

$$\mathrm{Var}^{(N)}(t^{1/2} Z^{(N)}) = tn, \tag{B.5a}$$

$$h^{(N)}(t^{1/2} Z^{(N)}) = \frac{n}{2} \log(2\pi e t), \tag{B.5b}$$

$$J^{(N)}(t^{1/2} Z^{(N)}) = t^{-1/2} Z^{(N)}, \tag{B.5c}$$

$$\mathcal{I}^{(N)}(t^{1/2} Z^{(N)}) = \frac{n}{t}. \tag{B.5d}$$

Comparing this with Example B.4, we see that these normalizations have lowered the "e ective dimension" of $M_N(\mathbb{C})^n_{\mathrm{sa}}$ from nN^2 to n. With these "microstate" normalizations, the definition (B.1) of the classical score becomes

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} \mathbb{E}\langle f(X + \varepsilon Z^{(N)}), Z^{(N)} \rangle_{M_N(\mathbb{C})^n_{\mathrm{sa}}} \big|_{\varepsilon = 0} = \mathbb{E}\langle J^{(N)}(X), f(X) \rangle_{M_N(\mathbb{C})^n_{\mathrm{sa}}}, \tag{B.6}$$

the relationship (B.2) between classical entropy and variance becomes

$$-\infty < h^{(N)}(X) \leq \frac{n}{2} \log(2\pi e \, \mathrm{Var}^{(N)}(X)/n),$$

the Stein identity (B.3) becomes

$$J^{(N)}(X + t^{1/2} Z^{(N)}) = \mathbb{E}[t^{-1/2} Z^{(N)} | X + t^{1/2} Z^{(N)}], \tag{B.7}$$

and the de Bruijn identity (B.4) becomes

$$
\begin{aligned}
\text{(B.8)} \qquad & h^{(N)}(X + t_1^{1/2} Z^{(N)}) - h^{(N)}(X + t_0^{1/2} Z^{(N)}) \\
&= \frac{1}{2} \int_{t_0}^{t_1} \mathcal{I}^{(N)}(X + t^{1/2} Z^{(N)})\, \mathrm{d}t.
\end{aligned}
$$

Note how there are no longer any factors of N appearing explicitly in these assertions (other than in the superscripts and subscripts). The reader is invited to verify that these identities and inequalities are compatible with (B.5). (See, e.g., [18, Section 16] for further explanation of these normalizations.)

We now introduce the definitions necessary to define microstate entropy.

Definition B.6 (***Microstates free entropy, cf.*** [34, Section 2.1]). Let $n \geq 1$ and $R > 0$, let $(\mathcal{A}, \tau)$ be a noncommutative probability space, and let $X \in \mathcal{A}_{\mathrm{sa}}^n$ be an n-tuple of self-adjoint elements with operator norm $\|X\|_{\mathrm{op}} := \max_j \|X_j\|_{\mathrm{op}} \leq R$.

(i) Let $\mathbb{C}\langle x_1, \ldots, x_n \rangle$ be the $*$-algebra of noncommutative polynomials in formal self-adjoint variables $x_1, \ldots, x_n$. We define $\Sigma_{n,R}$ as the space of tracial positive linear functionals $\lambda : \mathbb{C}\langle x_1, \ldots, x_n \rangle \to \mathbb{C}$ such that for all $i_1, \ldots, i_\ell \in \{1, \ldots, n\}$, we have $|\lambda(x_{i_1} \ldots x_{i_\ell})| \leq R^\ell$. We equip $\Sigma_{n,R}$ with the weak-$*$ topology.

(ii) We define the *noncommutative law of* X in $\Sigma_{n,R}$ to be the linear functional $\lambda_X \in \Sigma_{n,R}$ defined by

$$\lambda_X(p) := \tau(p(X)).$$

In particular, in the case where $\mathcal{A}$ is $M_N(\mathbb{C})$ and τ is the normalized trace $\mathrm{tr}_N := (1/N)\,\mathrm{Tr}$, we have $\lambda_Y(p) = \mathrm{tr}_N(p(Y))$ for any $p \in \mathbb{C}\langle x_1, \ldots, x_n \rangle$ and any n-tuple $(Y_1, \ldots, Y_n) \in M_N(\mathbb{C})_{\mathrm{sa}}^n$ of self-adjoint matrices in $M_N(\mathbb{C})$.

(iii) For an open set $\mathcal{U} \subseteq \Sigma_{n,R}$, we define the *microstate space*[10]

$$\Gamma_R^{(N)}(\mathcal{U}) := \{Y \in M_N(\mathbb{C})_{\mathrm{sa}}^n : \lambda_Y \in \mathcal{U}\}.$$

(iv) For $\lambda \in \Sigma_{n,R}$, we define[11]

$$\chi_R(\lambda) := \inf_{\mathcal{U} \ni \lambda} \limsup_{N \to \infty} \frac{1}{N^2} (\log \nu_{M_N(\mathbb{C})_{\mathrm{sa}}^n}(\Gamma_R^{(N)}(\mathcal{U})) + n \log N),$$

where the infimum is taken over all neighborhoods $\mathcal{U}$ of λ in $\Sigma_{n,R}$.

(v) We define $\chi(\lambda) := \sup_{R' \geq R} \chi_{R'}(\lambda)$. If $(\mathcal{A}, \tau)$ is a noncommutative probability space and $X \in \mathcal{A}_{\mathrm{sa}}^n$, then we also define $\chi(X) := \chi(\lambda_X)$.

---

[10] The condition $Y \in \Gamma_R^{(N)}(\mathcal{U})$ entails that $\lambda_Y \in \Sigma_{n,R}$, and hence $\|Y\|_{\mathrm{op}} \leq R$.

[11] The corresponding definition in [34] uses $\frac{1}{2} n \log N$ instead of $n \log N$, but this is due to the use of the un-normalized trace Tr instead of the normalized trace tr_N to define the Haar measure $\nu_{M_N(\mathbb{C})_{\mathrm{sa}}^n}$.

The next proposition expresses the microstate entropy χ in terms of the normalized classical entropies $h^{(N)}$ introduced previously.

Proposition B.7 (Random matrix interpretation of microstates free entropy). *Let X be an n-tuple of self-adjoint noncommutative random variables from $(\mathcal{A}, \tau)$. Then, $\chi(X)$ is the supremum of*

$$\limsup_{\ell \to \infty} h^{(N_\ell)}(X^{(\ell)})$$

over all sequences of natural numbers $(N_\ell)_{\ell \in \mathbb{N}}$ tending to ∞, and all sequences random variables $(X^{(\ell)})_{\ell \in \mathbb{N}}$ from $M_{N_\ell}(\mathbb{C})_{\mathrm{sa}}^n$ satisfying the following conditions:

(1) *$\lambda_{X^{(N_\ell)}}$ converges in probability to λ_X.*

(2) *For some $R > 0$, we have $\limsup_{\ell \to \infty} \|X^{(\ell)}\|_{\mathrm{op}} \leq R$ in probability, where $\|X\|_{\mathrm{op}}$ denotes the supremum of the operator norms $\|X_i\|_{\mathrm{op}}$ of the components $X_1, \ldots, X_n$ of X.*

(3) *There exist some constants $C > 0$ and $K > 0$ such that*

$$P(\|X^{(\ell)}\|_{M_{N_\ell}(\mathbb{C})_{\mathrm{sa}}^n} \geq C + \delta) \leq e^{-KN_\ell^2\delta^2} \quad \text{for all } \delta > 0. \tag{B.9}$$

Furthermore, the supremum (if it is $> -\infty$) is witnessed by random matrices which are uniformly bounded in operator norm and unitarily invariant in distribution.

Proof. First, let $(X^{(\ell)})_{\ell \in \mathbb{N}}$ be a sequence of random matrices as described above, and let $\mu^{(\ell)}$ be the associated probability measure. Fix $R' > R$. Let $\mathcal{U}$ be a neighborhood of the noncommutative law of X in $\Sigma_{n,R'}$. We apply Lemma B.5 (ii) with the partition $S_j^{(\ell)}$, $j \geq 0$, of $M_N(\mathbb{C})_{\mathrm{sa}}^n$ defined by

$$\begin{aligned} S_0^{(\ell)} &:= \Gamma_{R'}^{(N_\ell)}(\mathcal{U}), \\ S_1^{(\ell)} &:= B(0, C+1) \setminus \Gamma_{R'}^{(N_\ell)}(\mathcal{U}), \\ S_j^{(\ell)} &:= B(0, C+j) \setminus B(0, C+j-1) \quad \text{for } j \geq 2, \end{aligned}$$

where $B(0, r)$ denotes the ball of radius r in $M_N(\mathbb{C})_{\mathrm{sa}}^n$, to obtain

$$h^{(N_\ell)}(X^{(N_\ell)}) \leq \sum_{j=0}^{\infty} H_j^{(\ell)},$$

where

$$\begin{aligned} H_j^{(\ell)} := \mu^{(\ell)}(S_j^{(\ell)}) &\left(\frac{1}{N_\ell^2} \log \nu_{M_{N_\ell}(\mathbb{C})_{\mathrm{sa}}^n}(S_j^{(\ell)}) + n \log N \right) \\ &- \mu^{(\ell)}(S_j^{(\ell)}) \frac{1}{N_\ell^2} \log \mu^{(\ell)}(S_j^{(\ell)}). \end{aligned}$$

We have

$$H_0^{(\ell)} = \mu^{(\ell)}(\Gamma_{R'}^{(N_\ell)}(\mathcal{U}))\left(\frac{1}{N_\ell^2}\log \nu_{M_{N_\ell}(\mathbb{C})^n_{\mathrm{sa}}}(\Gamma_{R'}^{(N_\ell)}(\mathcal{U})) + n\log N_\ell\right) - \mu^{(\ell)}(S_0^{(\ell)})\frac{1}{N_\ell^2}\log \mu^{(\ell)}(S_0^{(\ell)}).$$

As $\ell \to \infty$, the second term on the righthand side goes to zero (bounding $-t\log t \le 1/e$ for any $t > 0$). From Definition B.6, we thus see that for any $\varepsilon > 0$ one can find $\mathcal{U}$ for which

$$\limsup_{\ell\to\infty} H_0^{(\ell)} \le \chi(X) + \varepsilon.$$

Next, to estimate $H_1^{(\ell)}$, we observe from a routine application of Stirling's formula (identifying the inner product space $M_{N_\ell}(\mathbb{C})^n_{\mathrm{sa}}$ with a standard nN_ℓ^2-dimensional Euclidean space) that

$$\frac{1}{N_\ell^2}\log \nu_{M_{N_\ell}(\mathbb{C})^n_{\mathrm{sa}}}(B(0,r)) = -n\log N_\ell + n\log r + O(n) \tag{B.10}$$

for any $r > 0$. By the inequality $\nu_{M_{N_\ell}(\mathbb{C})^n_{\mathrm{sa}}}(S_1^{(\ell)}) \le \nu_{M_{N_\ell}(\mathbb{C})^n_{\mathrm{sa}}}(B(0,C+1))$, we conclude that

$$\limsup_{\ell\to\infty} H_1^{(\ell)} \le 0.$$

For the terms $j \ge 2$, we see from (B.10), (B.9), and the fact that $-t\log t$ is increasing for $t \le 1/e$ that

$$\limsup_{\ell\to\infty}\sum_{j=2}^{\infty} H_j^{(\ell)} \le \limsup_{\ell\to\infty}\sum_{j=2}^{\infty} e^{-KN_\ell^2(j-1)^2}(n\log j + O(n) + K(j-1)^2) = 0.$$

Putting all these bounds together, and sending ε to zero, we conclude that

$$\limsup_{\ell\to\infty} h^{(N_\ell)}(X^{(\ell)}) \le \chi(X).$$

Hence, the supremum of the lim sup of classical entropies is thus less than or equal to $\chi(X)$.

For opposite inequality, assume without loss of generality that $\chi(X) > -\infty$ since otherwise the inequality is trivial. Fix $R > \|X\|_{\mathrm{op}}$. Let $(\mathcal{U}_\ell)_{\ell\in\mathbb{N}}$ be a sequence of nested neighborhoods of λ_X in $\Sigma_{n,R}$ shrinking to λ_X as $\ell \to \infty$. For each ℓ, choose a number N_ℓ such that

$$\frac{1}{N_\ell^2}\log \nu_{M_{N_\ell}(\mathbb{C})^n_{\mathrm{sa}}}(\Gamma_R^{(N_\ell)}(\mathcal{U}_\ell)) + n\log N_\ell > \limsup_{N\to\infty}\left(\frac{1}{N^2}\log \nu_{M_N(\mathbb{C})^n_{\mathrm{sa}}}(\Gamma_{R'}^{(N)}(\mathcal{U})) + n\log N\right) - \frac{1}{\ell}.$$

We can arrange that $N_{\ell+1} > N_\ell$ and hence $N_\ell \to \infty$. Define $\mu^{(\ell)}$ to be the uniform measure on $\Gamma_R^{(N_\ell)}(\mathcal{U}_\ell)$, and let $X^{(\ell)}$ be a random matrix tuple with distribution $\mu^{(\ell)}$. Then,

$$h^{(N_\ell)}(\mu^{(\ell)}) = \frac{1}{N_\ell^2} \log \nu_{M_{N_\ell}(\mathbb{C})^n_{\mathrm{sa}}}(\Gamma_R^{(N_\ell)}(\mathcal{U}_\ell)) + n \log N_\ell.$$

Hence,

$$\begin{aligned}
&\limsup_{N\to\infty} h^{(N_\ell)}(\mu^{(\ell)}) \\
&\qquad \geq \limsup_{\ell\to\infty} \left(\limsup_{N\to\infty} \left(\frac{1}{N^2} \log \nu_{M_N(\mathbb{C})^n_{\mathrm{sa}}}(\Gamma_{R'}^{(N)}(\mathcal{U})) + n \log N \right) - \frac{1}{\ell} \right) \\
&\qquad = \chi_R(\mu) = \chi(\mu),
\end{aligned}$$

where the last equality follows from [34, Proposition 2.4]. Moreover, it is clear that this choice of random matrix models is unitarily invariant and bounded in operator norm. ◻

Remark B.8. Note that, if $\|X^{(\ell)}\|_{M_N(\mathbb{C})^n_{\mathrm{sa}}} \leq C$, assumption (3) is trivially satisfied. It is also true of any random matrix models which satisfy Herbst's concentration inequality with a suitable normalization depending on the dimension N_ℓ (which in turn follows from a normalized log-Sobolev inequality). In particular, this applies when $X^{(\ell)} = t^{1/2} Z^{(\ell)}$ for a GUE tuple $Z^{(\ell)}$ from $M_{N_\ell}(\mathbb{C})^d_{\mathrm{sa}}$ and any fixed $t > 0$. See [16] and [1, Section 4.4.2]. Herbst's concentration inequality also implies that (2) holds for some R by [18, Lemma 11.5.2].

Remark B.9. Compare Proposition B.7 to the more explicit connections between microstates free entropy and classical entropy that occur for special random matrix models in [34] and [18, Proposition 16.1.4].

Remark B.10. It was pointed out to us by Ben Hayes (private communication) that a similar idea to Proposition B.7 has already been used in the context of sofic entropy. Bowen expressed the entropy of algebraic actions of residually finite groups as the supremum of the limits of classical entropies of certain measures on the model spaces (finitary approximations) (see Definition 4 and Theorem 4.1 in [12]). Similarly, Austin used this approach to define a version of sofic entropy in a more general context [4].

Now, we give an analog of Proposition 2.3.

Lemma B.11 (Classical score and minors). *Let X be a random element of $M_N(\mathbb{C})^n_{\mathrm{sa}}$ with finite classical Fisher information (in particular, the normalized classical score $J^{(N)}(X)$ exists). Let $1 \leq M \leq N$, and let $\pi^{(N,M)} : M_N(\mathbb{C})^n_{\mathrm{sa}} \to M_M(\mathbb{C})^n_{\mathrm{sa}}$ be* *the compression map that sends a tuple* $(X_1, \dots, X_n)$ *in* $M_N(\mathbb{C})^n_{\mathrm{sa}}$ *to the tuple consisting of the upper left* $M \times M$ *minors of* $X_1, \dots, X_n$. *Define the normalized compression*

$$\Pi^{(N,M)} := \frac{N^{1/2}}{M^{1/2}} \pi^{(N,M)}.$$

Then, $\Pi^{(N,M)}(X)$ *has a normalized classical score in* $M_M(\mathbb{C})^n_{\mathrm{sa}}$ *given by*

$$J^{(M)}(\Pi^{(N,M)}(X)) = \mathbb{E}[\Pi^{(N,M)}(J^{(N)}(X)) | \Pi^{(N,M)}(X)].$$

Proof. Let $Z^{(N)}$ be a GUE tuple in $M_N(\mathbb{C})^n_{\mathrm{sa}}$ (classically) independent of X. Then, it is easy to see that $\Pi^{(N,M)}(Z^{(N)})$ is a GUE tuple in $M_M(\mathbb{C})^n_{\mathrm{sa}}$ (classically) independent of $\Pi^{(N,M)}(X)$. By (B.6), it su ces to show that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\varepsilon} \mathbb{E}\langle f(\Pi^{(N,M)}(X) + \varepsilon \Pi^{(N,M)}(Z^{(N)})), \Pi^{(N,M)}(Z^{(N)}) \rangle_{M_M(\mathbb{C})^n_{\mathrm{sa}}} |_{\varepsilon=0} \\ = \mathbb{E}\langle \mathbb{E}[\Pi^{(N,M)}(J^{(N)}(X)) | \Pi^{(N,M)}(X)] f(\Pi^{(N,M)}(X)) \rangle_{M_M(\mathbb{C})^n_{\mathrm{sa}}}\end{aligned}$$

for any smooth $f \colon M_M(\mathbb{C})^n_{\mathrm{sa}} \to M_M(\mathbb{C})^n_{\mathrm{sa}}$. We can remove the conditional expectation on the righthand side, thus reducing to

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\varepsilon} \mathbb{E}\langle f(\Pi^{(N,M)}(X) + \varepsilon \Pi^{(N,M)}(Z^{(N)})), \Pi^{(N,M)}(Z^{(N)}) \rangle_{M_M(\mathbb{C})^n_{\mathrm{sa}}} |_{\varepsilon=0} \\ = \mathbb{E}\langle \Pi^{(N,M)}(J^{(N)}(X)), f(\Pi^{(N,M)}(X)) \rangle_{M_M(\mathbb{C})^n_{\mathrm{sa}}}.\end{aligned}$$

By embedding $M_M(\mathbb{C})^n_{\mathrm{sa}}$ into $M_N(\mathbb{C})^n_{\mathrm{sa}}$ by padding zero entries to the $M \times M$ matrices to create $N \times N$ matrices, this simplifies further to

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\varepsilon} \mathbb{E}\langle f(\Pi^{(N,M)}(X + \varepsilon Z^{(N)})), Z^{(N)} \rangle_{M_N(\mathbb{C})^n_{\mathrm{sa}}} |_{\varepsilon=0} \\ = \mathbb{E}\langle J^{(N)}(X), f(\Pi^{(N,M)}(X)) \rangle_{M_N(\mathbb{C})^n_{\mathrm{sa}}}.\end{aligned}$$

But this follows from (B.6). ◻

As a consequence we can establish a classical analog of (2.7), except that there is an error coming from the diagonal elements of the matrix (which will end up going to zero in the limit as $N \to \infty$).

Corollary B.12 (Approximate monotonicity of normalized classical Fisher information). *Let the notation and hypotheses be as in Lemma B.11. If the distribution of* X *is additionally invariant under unitary conjugation, one has*

$$\mathcal{I}^{(M)}(\Pi^{(N,M)}(X)) \leq \mathcal{I}^{(N)}(X) + \frac{N}{M} \mathbb{E}\| \operatorname{diag}(J^{(N)}(X)) \|^2_{M_N(\mathbb{C})^n_{\mathrm{sa}}}, \tag{B.11}$$

where $\operatorname{diag}(A) = (\operatorname{diag}(A_1), \dots, \operatorname{diag}(A_n))$ *is the orthogonal projection onto the space of diagonal matrices of a tuple* $A \in M_N(\mathbb{C})^n_{\mathrm{sa}}$.

Proof. From Lemma B.11, one has

$$\mathcal{I}^{(M)}(\Pi^{(N,M)}(X)) = \mathbb{E}||\mathbb{E}[\Pi^{(N,M)}(J^{(N)}(X))|\Pi^{(N,M)}(X)]||^2_{M_M(\mathbb{C})^n_{\mathrm{sa}}}$$
$$\leq \mathbb{E}||\Pi^{(N,M)}(J^{(N)}(X))||^2_{M_M(\mathbb{C})^n_{\mathrm{sa}}}.$$

Now, let σ be a random permutation matrix in $U(N)$, drawn using Haar measure, (classically) independent of X. From the unitary invariance of X we then have

$$\mathbb{E}||\Pi^{(N,M)}(J^{(N)}(X))||^2_{M_M(\mathbb{C})^n_{\mathrm{sa}}} = \mathbb{E}_X\mathbb{E}_\sigma||\Pi^{(N,M)}(\sigma J^{(N)}(X)\sigma^{-1})||^2_{M_M(\mathbb{C})^n_{\mathrm{sa}}}$$

where we use $\mathbb{E}_\sigma$ to denote taking expectation just over σ, and $\mathbb{E}_X$ to denote taking expectation over the variable X (which is independent of σ). For any (deterministic) tuple $A = (A_1, \ldots, A_n)$ in $M_N(\mathbb{C})^n_{\mathrm{sa}}$, with $a_{k,ij}$ denoting the ij entry of A_k, direct computation shows that

$$||A||^2_{M_N(\mathbb{C})^n_{\mathrm{sa}}} = \frac{1}{N}\sum_{k=1}^n \sum_{1\leq i,j\leq N} |a_{k,ij}|^2$$

and

$$\mathbb{E}_\sigma||\Pi^{(N,M)}(\sigma A\sigma^{-1})||^2_{M_M(\mathbb{C})^n_{\mathrm{sa}}} = \frac{1}{N}\frac{N(M-1)}{M(N-1)}\sum_{k=1}^n \sum_{1\leq i,j\leq N: i\neq j} |a_{k,ij}|^2$$
$$+ \frac{1}{N}\frac{N}{M}\sum_{k=1}^n\sum_{i=1}^N |a_{k,ii}|^2$$

and thus (since $N(M-1) \leq M(N-1)$)

$$\mathbb{E}_\sigma||\Pi^{(N,M)}(\sigma A\sigma^{-1})||^2_{M_M(\mathbb{C})^n_{\mathrm{sa}}} \leq ||A||^2_{M_N(\mathbb{C})^n_{\mathrm{sa}}} + \frac{N}{M}||\operatorname{diag}(A)||^2_{M_N(\mathbb{C})^n_{\mathrm{sa}}}.$$

Replacing $A = J^{(N)}(X)$ for each possible value of X and then applying the expectation $\mathbb{E}_X$, we conclude

$$\mathbb{E}||\Pi^{(N,M)}(J^{(N)}(X))||^2_{M_M(\mathbb{C})^n_{\mathrm{sa}}}$$
$$\leq \mathbb{E}||J^{(N)}||^2_{M_N(\mathbb{C})^n_{\mathrm{sa}}} + \frac{N}{M}\mathbb{E}||\operatorname{diag}(J^{(N)}(X))||^2_{M_M(\mathbb{C})^n_{\mathrm{sa}}},$$

and the claim follows. □

We now integrate this to obtain the following result.

Lemma B.13 (Approximate monotonicity of normalized classical entropy di erences). *Let X be a random element of $M_N(\mathbb{C})^n_{\mathrm{sa}}$ with finite variance and unitarily invariant distribution. Let $Z^{(N)}$ be a* GUE *tuple in $M_N(\mathbb{C})^n_{\mathrm{sa}}$ (classically) independent of X, and set $X_t := X + t^{1/2}Z^{(N)}$. Let $1 \leq M \leq N$, and let $\Pi^{(N,M)}$ be the* *normalized compression operator from Lemma B.11. Then, for* $0 < t_0 < t_1$,

$$h^{(M)}(\Pi^{(N,M)}(X_{t_1})) - h^{(M)}(\Pi^{(N,M)}(X_{t_0})) \\ \leq h^{(N)}(X_{t_1}) - h^{(N)}(X_{t_0}) + \frac{n}{2M}\log\frac{t_1}{t_0}.$$

Proof. Note that

$$\Pi^{(N,M)}(X_t) = \Pi^{(N,M)}(X) + t^{1/2}\Pi^{(N,M)}(Z^{(N)})$$

and that $\Pi^{(N,M)}(Z^{(N)})$ is a GUE tuple in $M_M(\mathbb{C})^n_{\text{sa}}$ classically independent of $\Pi^{(N,M)}(X)$. Hence, by two applications of (B.8), it su ces to establish the inequality

$$\mathcal{I}^{(M)}(\Pi^{(N,M)}(X_t)) \leq \mathcal{I}^{(N)}(X_t) + \frac{n}{Mt}$$

for all $t > 0$. By (B.11), it su ces to show that

$$\mathbb{E}\|\operatorname{diag}(J^{(N)}(X_t))\|^2_{M_N(\mathbb{C})^n_{\text{sa}}} \leq \frac{n}{Nt}.$$

By (B.7), we have $J^{(N)}(X_t) = \mathbb{E}[t^{-1/2}Z^{(N)}|X_t]$. In particular,

$$\mathbb{E}\|\operatorname{diag}(J^{(N)}(X_t))\|^2_{M_N(\mathbb{C})^n_{\text{sa}}} \leq \mathbb{E}\|t^{-1/2}\operatorname{diag}(Z^{(N)})\|^2_{M_N(\mathbb{C})^n_{\text{sa}}} = \frac{n}{Nt}$$

as required. □

Proof of Theorem B.1. Consider a self-adjoint n-tuple X from $(\mathcal{A},\tau)$ and a freely independent projection p of trace $1/k$ in $\mathcal{A}$. By enlarging $(\mathcal{A},\tau)$ if necessary assume it contains a tuple $Z = (Z_1,\dots,Z_n)$ of semicircular variables freely independent of each other and of X and p. Set $X_t := X + t^{1/2}Z$ for every $t \geq 0$, and let Π be the normalized compression $\Pi := k^{1/2}\pi$.

We can assume without loss of generality that $\chi(X) > -\infty$ since otherwise the inequality is trivial. By Proposition B.7, there exists a sequence of integers N_ℓ tending to ∞ and random $N_\ell \times N_\ell$ matrix tuples $X^{(\ell)}$ with $\|X^{(\ell)}\| \leq R$ and $\lambda_{X^{(\ell)}} \to \mu$ in probability, such that

$$\chi(X) = \limsup_{\ell\to\infty} h^{(N_\ell)}(X^{(\ell)}).$$

Let $M_\ell := \lceil N_\ell/k\rceil$, so that $M_\ell/N_\ell \to 1/k$ as $\ell \to \infty$. Let $Z^{(\ell)}$ be a GUE tuple in $M_{N_\ell}(\mathbb{C})^n_{\text{sa}}$ (classically) independent of $X^{(\ell)}$, and set $X^{(\ell)}_t := X^{(\ell)} + t^{1/2}Z^{(\ell)}$ for $t \geq 0$. Let $\Pi^{(N_\ell,M_\ell)}$ be the normalized compression operator from Lemma B.11, and let $P^{(\ell)} \in M_{N_\ell}(\mathbb{C})$ be the orthogonal projection matrix onto the span of the first M_ℓ basis vectors.

It is a standard result in random matrix theory (see, e.g., Theorem 2.2 in [32], Section 5.5 in [1]) that $\lambda_{Z^{(\ell)}}$ converges almost surely to λ_Z. We also have that $\lambda_{X^{(\ell)}} \to \lambda_X$ in probability, and $\lambda_{P^{(\ell)}} \to \lambda_p$. Because of the independence and unitary invariance of $X^{(\ell)}$ and $Z^{(\ell)}$, Voiculescu's asymptotic freeness theory [32], [36] implies that $\lambda_{(X^{(\ell)},Z^{(\ell)},P^{(\ell)})} \to \lambda_{(X,Z,p)}$ in probability. In particular, this implies that $\lambda_{\Pi^{(\ell)}(X_t^{(\ell)})} \to \lambda_{\Pi(X_t)}$ in probability.

Note that for any $t_0 > 0$, the matrix models $\Pi^{(N_\ell,M_\ell)}(X_{t_0}^{(\ell)})$ satisfy the hypotheses of Proposition B.7. The tail bound hypothesis (3) follows because $X^{(\ell)}$ is bounded in operator norm and because $Z^{(\ell)}$ satisfies these tail bounds using known concentration inequalities as explained in Remark B.8. Thus, by Proposition B.7,

$$\chi(\Pi(X_{t_0})) \geq \limsup_{\ell\to\infty} h^{(M_\ell)}(\Pi^{(N_\ell,M_\ell)}(X_{t_0}^{(\ell)})).$$

By Lemma B.13, for any $0 < t_0 < t_1$, we have

$$\begin{aligned} h^{(M_\ell)}(\Pi^{(N_\ell,M_\ell)}(X_{t_0}^{(\ell)})) - h^{(M_\ell)}(\Pi^{(N_\ell,M_\ell)}(X_{t_1}^{(\ell)})) \\ \geq h^{(N_\ell)}(X_{t_0}^{(\ell)}) - h^{(N_\ell)}(X_{t_1}^{(\ell)}) - \frac{n}{2M_\ell}\log\frac{t_1}{t_0}. \end{aligned} \tag{B.12}$$

Using Lemma B.5 (iii), and (B.5), we have

$$\begin{aligned} h^{(M_\ell)}(\Pi^{(N_\ell,M_\ell)}(X_{t_1}^{(\ell)})) &= h^{(M_\ell)}(\Pi^{(N_\ell,M_\ell)}(X^{(\ell)}) + t_1^{1/2}\Pi^{(N_\ell,M_\ell)}(Z^{(\ell)})) \\ &\geq h^{(M_\ell)}(t_1^{1/2}\Pi^{(N_\ell,M_\ell)}(Z^{(\ell)})) \\ &= \frac{n}{2}\log(2\pi e t_1). \end{aligned}$$

and similarly,

$$h^{(N_\ell)}(X_{t_0}^{(\ell)}) \geq h^{(N_\ell)}(X^{(\ell)}).$$

Finally, from Lemma B.5 (i) we have

$$h^{(N_\ell)}(X_{t_1}^{(\ell)}) \leq \frac{n}{2}\log(2\pi e(\mathrm{Var}^{(N_\ell)}(X^{(\ell)})/n + t_1)).$$

Substituting these estimates into (B.12) and collecting terms, we obtain

$$\begin{aligned} h^{(M_\ell)}(\Pi^{(N_\ell,M_\ell)}(X_{t_0}^{(\ell)})) \geq h^{(N_\ell)}(X^{(\ell)}) - n\log\left[2\pi e\left(1 + \frac{\mathrm{Var}^{(N_\ell)}(X^{(\ell)})}{nt_1}\right)\right] \\ - \frac{n}{2M_\ell}\log\frac{t_1}{t_0}. \end{aligned}$$

Taking the lim sup as $\ell \to \infty$, we conclude

$$\chi(\Pi(X_{t_0})) \geq \chi(X) - n \log\left[2\pi e\left(1 + \frac{\mathrm{Var}(X)}{nt_1}\right)\right],$$

where

$$\mathrm{Var}(X) := \sum_{j=1}^{n} \|X - \tau(X)\|_\tau^2,$$

and we have observed that $\mathrm{Var}^{(N_\ell)}(X^{(\ell)}) \to \mathrm{Var}(X)$ because $X^{(\ell)}$ is bounded in operator norm and $\lambda_{X^{(\ell)}} \to \lambda_X$ in probability. Taking limits as $t_1 \to \infty$, we obtain $\chi(\Pi(X_{t_0})) \geq \chi(X)$. Finally, note that $\Pi(X_{t_0})$ is bounded in operator norm by some constant R', and converges in noncommutative law to $\Pi(X)$; hence, using the upper-semicontinuity of χ on $\Sigma_{n,R'}$ established in [34, Proposition 2.6], we take the limit of the above inequality as $t_0 \to 0$ to conclude

$$\chi(\Pi(X)) \geq \chi(X)$$

as required. ◻

2.1. Acknowledgements. The authors were partially supported by the National Science Foundation (grant nos. DMS-1762360 and DMS-1764034, respectively). The second author was also supported by a Simons Investigator Award. This project was initiated during the IPAM program for Quantitative Linear Algebra in 2018. Thanks are due to Vadim Gorin, Istvan Prause, and Stefan Steinerberger for providing recent relevant references, and to David Jekel for providing Appendix B (David Jekel was supported by NSF grant no. DMS-2002826). Finally, much gratitude to the anonymous referee for careful reading of the manuscript and several useful suggestions and corrections.

REFERENCES

[1] G. W. ANDERSON, A. GUIONNET, and O. ZEITOUNI, *An Introduction to Random Matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010. MR2760897

[2] M. V. ANSHELEVICH, S. T. BELINSCHI, M. FEVRIER, and A. NICA, *Convolution powers in the operator-valued framework*, Trans. Amer. Math. Soc. **365** (2013), no. 4, 2063–2097. `http://dx.doi.org/10.1090/S0002-9947-2012-05736-9`. MR3009653

[3] S. ARTSTEIN-AVIDAN, K. M. BALL, F. BARTHE, and A. NAOR, *Solution of Shannon's problem on the monotonicity of entropy*, J. Amer. Math. Soc. **17** (2004), no. 4, 975–982. `http://dx.doi.org/10.1090/S0894-0347-04-00459-X`. MR2083473

[4] T. AUSTIN, *Additivity properties of sofic entropy and measures on model spaces*, Forum Math. Sigma **4** (2016), Paper No. e25, 79. `http://dx.doi.org/10.1017/fms.2016.18`. MR3542515

[5] S. T. BELINSCHI, *Some geometric properties of the subordination function associated to an operator-valued free convolution semigroup*, Complex Anal. Oper. Theory **13** (2019), no. 1, 61–84. `http://dx.doi.org/10.1007/s11785-017-0688-y`. MR3905583

[6] S. T. BELINSCHI and H. BERCOVICI, *Atoms and regularity for measures in a partially defined free convolution semigroup*, Math. Z. **248** (2004), no. 4, 665–674. `http://dx.doi.org/10.1007/s00209-004-0671-y`. MR2103535

[7] ———, *Partially defined semigroups relative to multiplicative free convolution*, Int. Math. Res. Not. **2** (2005), 65–101. http://dx.doi.org/10.1155/IMRN.2005.65. MR2128863

[8] S. T. Belinschi and A. Nica, *On a remarkable semigroup of homomorphisms with respect to free multiplicative convolution*, Indiana Univ. Math. J. **57** (2008), no. 4, 1679–1713. http://dx.doi.org/10.1512/iumj.2008.57.3285. MR2440877

[9] H. Bercovici and D. V. Voiculescu, *Superconvergence to the central limit and failure of the Cramér theorem for free random variables*, Probab. Theory Related Fields **103** (1995), no. 2, 215–222. http://dx.doi.org/10.1007/BF01204215. MR1355057

[10] P. Biane, *Representations of unitary groups and free convolution*, Publ. Res. Inst. Math. Sci. **31** (1995), no. 1, 63–79. http://dx.doi.org/10.2977/prims/1195164791. MR1317523

[11] C. Boutillier, *The bead model and limit behaviors of dimer models*, Ann. Probab. **37** (2009), no. 1, 107–142. http://dx.doi.org/10.1214/08-AOP398. MR2489161

[12] L. Bowen, *Entropy for expansive algebraic actions of residually finite groups*, Ergodic Theory Dynam. Systems **31** (2011), no. 3, 703–718. http://dx.doi.org/10.1017/S0143385710000179. MR2794944

[13] H. Cohn, R. Kenyon, and J. Propp, *A variational principle for domino tilings*, J. Amer. Math. Soc. **14** (2001), no. 2, 297–346. http://dx.doi.org/10.1090/S0894-0347-00-00355-6. MR1815214

[14] A. Gordenko, *Limit shapes of large skew Young tableaux and a modification of the TASEP process* (September, 2020), preprint. https://doi.org/10.48550/arXiv.2009.10480.

[15] V. Gorin, *Lectures on Random Lozenge Tilings*, Cambridge Studies in Advanced Mathematics, vol. 193, Cambridge University Press, Cambridge, 2021. MR4299268

[16] A. Guionnet and O. Zeitouni, *Concentration of the spectral measure for large matrices*, Electron. Comm. Probab. **5** (2000), 119–136. http://dx.doi.org/10.1214/ECP.v5-1026. MR1781846

[17] H.-W. Huang, *Supports of measures in a free additive convolution semigroup*, Int. Math. Res. Not. IMRN **12** (2015), 4269–4292. http://dx.doi.org/10.1093/imrn/rnu064. MR3356753

[18] D. Jekel, *Evolution equations in noncommutative probability*, Ph.D. thesis, University of California, Los Angeles, 2020.

[19] S. G. G. Johnston and N. O'Connell, *Scaling limits for non-intersecting polymers and Whittaker measures*, J. Stat. Phys. **179** (2020), no. 2, 354–407. http://dx.doi.org/10.1007/s10955-020-02534-y. MR4091564

[20] R. Kenyon and I. Prause, *Gradient variational problems in* $\mathbb{R}^2$ (June, 2020), preprint. https://doi.org/10.48550/arXiv.2006.01219.

[21] D. Marcus, D. Spielman, and N. Srivastava, *Finite free convolutions of polynomials* (April, 2015), preprint, available at http://arxiv.org/abs/arXiv:1504.00350.

[22] A. P. Metcalfe, *Universality properties of Gelfand-Tsetlin patterns*, Probab. Theory Related Fields **155** (2013), no. 1–2, 303–346. http://dx.doi.org/10.1007/s00440-011-0399-7. MR3010400

[23] A. Nica and R. Speicher, *On the multiplication of free N-tuples of noncommutative random variables*, Amer. J. Math. **118** (1996), no. 4, 799–837. http://dx.doi.org/10.1353/ajm.1996.0034. MR1400060

[24] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, Ill., 1949. MR0032134

[25] D. Y. Shlyakhtenko, *A free analogue of Shannon's problem on monotonicity of entropy*, Adv. Math. **208** (2007), no. 2, 824–833. http://dx.doi.org/10.1016/j.aim.2006.03.014. MR2304337

[26] ———, *On operator-valued free convolution powers*, Indiana Univ. Math. J. **62** (2013), no. 1, 91–97. http://dx.doi.org/10.1512/iumj.2013.62.4863. MR3158503

[27] S. Steinerberger, *A nonlocal transport equation describing roots of polynomials under di erentiation*, Proc. Amer. Math. Soc. **147** (2019), no. 11, 4733–4744. http://dx.doi.org/10.1090/proc/14699. MR4011508

[28] A. J. STAM, *Some inequalities satisfied by the quantities of information of Fisher and Shannon*, Information and Control **2** (1959), no. 2, 101–112. http://dx.doi.org/10.1016/S0019-9958(59)90348-1. MR109101

[29] W. SUN, *Dimer model, bead model and standard Young tableaux: Finite cases and limit shapes* (April, 2018), preprint. https://doi.org/10.48550/arXiv.1802.08750.

[30] T. TAO, *Topics in Random Matrix Theory*, Graduate Studies in Mathematics, vol. 132, American Mathematical Society, Providence, RI, 2012. http://dx.doi.org/10.1090/gsm/132. MR2906465

[31] D. V. VOICULESCU, *Addition of certain noncommuting random variables*, J. Funct. Anal. **66** (1986), no. 3, 323–346. http://dx.doi.org/10.1016/0022-1236(86)90062-5. MR839105

[32] ______, *Limit laws for random matrices and free products*, Invent. Math. **104** (1991), no. 1, 201–220. http://dx.doi.org/10.1007/BF01245072. MR1094052

[33] ______, *The analogues of entropy and of Fisher's information measure in free probability theory. I*, Comm. Math. Phys. **155** (1993), no. 1, 71–92. http://dx.doi.org/10.1007/BF02100050. MR1228526

[34] ______, *The analogues of entropy and of Fisher's information measure in free probability theory. II*, Invent. Math. **118** (1994), no. 3, 411–440. http://dx.doi.org/10.1007/BF01231539. MR1296352

[35] ______, *The analogues of entropy and of Fisher's information measure in free probability theory. V: Noncommutative Hilbert transforms*, Invent. Math. **132** (1998), no. 1, 189–227. http://dx.doi.org/10.1007/s002220050222. MR1618636

[36] ______, *A strengthened asymptotic freeness result for random matrices with applications to free entropy*, Internat. Math. Res. Notices **1** (1998), 41–63. http://dx.doi.org/10.1155/S107379289800004X. MR1601878

[37] D. V. VOICULESCU, K. J. DYKEMA, and A. NICA, *Free Random Variables: A Noncommutative Probability Approach to Free Products with Applications to Random Matrices, Operator Algebras and Harmonic Analysis on Free Groups*, CRM Monograph Series, vol. 1, American Mathematical Society, Providence, RI, 1992. http://dx.doi.org/10.1090/crmm/001. MR1217253

[38] J. D. WILLIAMS, *On the Hausdor continuity of free Lèvy processes and free convolution semigroups*, J. Math. Anal. Appl. **459** (2018), no. 1, 604–613. http://dx.doi.org/10.1016/j.jmaa.2016.09.062. MR3730459

DIMITRI SHLYAKHTENKO & TERENCE TAO:
Department of Mathematics
University of California Los Angeles,
Los Angeles, CA 90095-1555
USA
E-MAIL: shlyakht@math.ucla.edu
E-MAIL: tao@math.ucla.edu

DAVID JEKEL:
University of California
San Diego, Department of Mathematics
La Jolla, CA 92093-0112
USA
E-MAIL: djekel@ucsd.edu

2010 MATHEMATICS SUBJECT CLASSIFICATION: 46L54, 15B52.

Received: September 3, 2020.