

# FREE STEIN KERNELS AND AN IMPROVEMENT OF THE FREE LOGARITHMIC SOBOLEV INEQUALITY

MAX FATHI<sup>•</sup> AND BRENT NELSON<sup>◦</sup>

ABSTRACT. We introduce a free version of the Stein kernel, relative to a semicircular law. We use it to obtain a free counterpart of the HSI inequality of Ledoux, Peccatti and Nourdin, which is an improvement of the free logarithmic Sobolev inequality of Biane and Speicher, as well as a rate of convergence in the (multivariate) entropic free Central Limit Theorem. We also compute the free Stein kernels for several relevant families of self-adjoint operators.

## INTRODUCTION

**Entropy.** The classical notion of entropy was first introduced by Boltzmann in his work on the kinetic theory of gases, and since then has played an ubiquitous role in many fields, such as statistical physics, information theory and mathematics. The classical entropy takes the form

$$H(\nu) := \int f \log f dx \quad \nu = f dx.$$

Following the usual convention in probability, this is the opposite of the physical entropy. Often, we are interested in a relative entropy with respect to a non-negative measure  $\mu = e^{-V} dx$ , which takes the form

$$H_\mu(\nu) = \int \frac{d\nu}{d\mu} \log \left( \frac{d\nu}{d\mu} \right) d\mu = H \left( \frac{d\nu}{dx} \right) + \int V d\nu.$$

We would like to point out that physicists would call this quantity a free energy (the second term plays the role of an internal energy, and we omitted the temperature, since it will play no role in this work and can be absorbed in the potential). When the reference measure is the Lebesgue measure, the relative entropy coincides with the (simple) entropy. The relative entropy notably appears in Sanov's theorem as the large deviations rate function for sums of i.i.d. random variables.

In [37], Voiculescu introduced a notion of free entropy  $\chi$  which plays the role of entropy in free probability theory. Most notably, it is monotone along the free central limit theorem, and the associated free energy/relative entropy appears as the large deviations rate function for the empirical spectrum of large random matrices. It also detects freeness: an  $n$ -tuple  $(x_1, \dots, x_n)$  of non-commutative random variables with  $\sum \chi(x_i) > -\infty$  is free iff it satisfies  $\chi(x_1, \dots, x_n) = \sum \chi(x_i)$ .

The microstates free entropy relative to a potential  $V$  can then be defined as

$$\chi(x_1, \dots, x_n | V) = \tau(V(x_1, \dots, x_n)) - \chi(x_1, \dots, x_n).$$

As in the classical case, we should think of this quantity as playing the role of a free energy.

Several of the uses of the entropy in classical probability are consequences of its property of being monotone along the heat flow. However, for technical reasons, it is hard to study the analogous property for the microstates free entropy (which would be monotonicity along convolutions with a free semicircular random variable). It is not even clear whether the microstates free entropy can be differentiated along the free heat flow. To get around this difficulty, Voiculescu introduced a second notion of entropy, called the non-microstates free entropy. This second notion of entropy was specifically constructed so that the Fisher information could be easily introduced as the derivative of the entropy along the free heat flow. Since this work will mainly deal with logarithmic Sobolev inequalities, for which this property is essential, we shall mostly deal with this second notion of entropy. As is standard, the non-microstates free entropy (relative to  $V$ ) shall be denoted by  $\chi^*$  (resp.  $\chi^*(x_1, \dots, x_n | V)$ ).

---

• Research supported by NSF FRG grant DMS-1361122.  
◦ Research supported by NSF grants DMS-1502822.

In dimension one, the two notions of entropy coincide, and have an explicit expression (as we shall see in Section 4), but for  $n$ -tuples of non-commutative random variables with  $n \geq 2$  this is an open problem.

**Logarithmic Sobolev inequalities.** The classical logarithmic Sobolev inequality (LSI for short)

$$H_\mu(\nu) \leq \frac{1}{2\rho} \int \frac{|\nabla f|^2}{f} d\mu; \quad \nu = f\mu.$$

was introduced by Gross in [17], where it was shown to be equivalent to hypercontractivity of the semigroup of a diffusion process naturally associated with  $\mu$ , a remarkable smoothing property that states that the action of the semigroup sends the space  $L^p(\mu)$  into  $L^q(\mu)$  for any  $q > p > 1$  after a finite time. This inequality can be thought of as a kind of Sobolev inequality, where instead of embedding the Sobolev space  $H^1$  into  $L^p$  for some  $p > 2$ , we embed it into the Orlicz space  $L^2 \log L$ . It has played an important role in many aspects of stochastic analysis: it is used to study long-time behavior for diffusion processes, it implies Gaussian concentration for the measure  $\mu$  and has also been used to study large-scale behavior of interacting particle systems and large random matrices. It also implies Talagrand's transport-entropy inequality [30]. We refer to [1] for more about its uses.

A fundamental property of this inequality is that, unlike for the usual Sobolev inequalities, the dimension plays no role: when it holds for a metric-measure space, then it holds for products of the space, with the same constant. This property is the source of many interesting applications in infinite-dimensional analysis, such as dimension-free concentration bounds.

When the reference measure is the standard Gaussian measure on  $\mathbb{R}^n$ , it is known that the LSI holds with sharp constant  $\rho = 1$ , and moreover cases of equality are explicitly known, and are standard Gaussians with arbitrary mean [8, 9]. A classical theorem of Bakry and Émery [2] states that if the reference measure on  $\mathbb{R}^n$  is of the form  $\mu = e^{-V} dx$  with  $\text{Hess } V \geq \rho Id$  for some  $\rho > 0$ , then it satisfies a LSI with constant  $\rho$ .

In the free case, the LSI (if it holds) takes the form

$$E(X|V) - E(Y|V) \leq \frac{1}{2\rho} \Phi^*(X|V)$$

where  $E \in \{\chi, \chi^*\}$ ,  $V$  is some reference potential,  $X$  is an  $n$ -tuple of free random variables,  $Y$  is another  $n$ -tuple that minimizes the free entropy relative to  $V$ , and  $\Phi^*(X|V)$  is the non-microstates free Fisher information of  $X$  relative to the potential  $V$ . This inequality was obtained by Biane and Speicher [3] when the reference potential is quadratic and  $Y$  is therefore an  $n$ -tuple of free semicircular random variables.

We shall still manage to obtain a LSI involving the microstates free entropy in the situation where the potential is uniformly convex. However, the improved versions we shall discuss in the next subsection rely on semigroup arguments, and we shall only establish them for the non-microstates free entropy.

Unlike in the classical setting, we do not know whether the free LSI automatically implies the free Talagrand inequality (which has been studied in [19] for example), except in dimension one [22]. The main obstacle is that the existence of free monotone transport is not yet well-understood and examples are quite rare.

**Improved functional inequalities.** In the past decade, there has been a lot of interest in improved or quantitative functional inequalities. Informally, the situation is as follows: given a sharp functional inequality of the form  $F(f) \leq G(f)$  for which the cases of equality are explicitly known, can one *improve* the functional inequality by taking into account how far  $f$  is from the set  $\mathcal{F}$  of functions for which equality holds. Typical improvements that have been considered take the form

$$\delta(f) = G(f) - F(f) \geq c d(f, \mathcal{F})^\alpha$$

where  $d$  is a well-chosen distance on the set of functions considered. Examples of inequalities for which such improvements have been obtained include isoperimetric inequalities [16], the Brunn-Minkowski inequality [14], Sobolev inequalities [11] and many others. In the last few years, there has been several partial results for the Gaussian logarithmic Sobolev inequality [4, 5, 15, 21, 24] but they are not entirely satisfactory. The problem of obtaining a deficit estimate in terms of a simple distance that is valid for all functions and behaves well in high dimension is still open at the time of writing.

One of the results presented in this work is an adaptation of a result of [4] on deficit estimates for the LSI to the free case. We show that for a  $n$ -tuple of non-commutative random variables with variance smaller than  $n$ , the deficit in the free LSI (with respect to a  $n$ -tuple of free semicircular laws) is bounded from below by a function of the free Fisher information.

Another type of improvement to functional inequalities one can seek is to take into account another functional  $H$  and seek a functional inequality of the form  $F(f) \leq c(G(f), H(f))$  for some function  $c$  satisfying  $c(x, y) \leq x \forall y$ . Informally, this type of improvement consists of using extra information (the size of  $H$ ) to learn more about how  $F$  and  $G$  relate. An improvement of this form for the Gaussian logarithmic Sobolev inequality, the so-called HSI inequality, has been obtained by Ledoux, Nourdin and Peccatti in [23], involving the Stein deficit (see 1.7):

$$H_\gamma(\nu) \leq \frac{S(\nu|\gamma)^2}{2} \log \left( 1 + \frac{I_\gamma(\nu)}{S(\nu|\gamma)^2} \right) \quad (1)$$

for all centered measures  $\nu$ , where  $H_\gamma$  stands for the relative entropy with respect to the Gaussian measure, and  $I_\gamma$  for the Fisher information with respect to  $\gamma$ . Since  $\log(1+x) < x$  whenever  $x > 0$ , this is indeed a strict improvement of the classical LSI. This inequality has some applications to the study of long-time behavior of the Ornstein-Uhlenbeck semigroup, and can be used to yield rates of convergence in the entropic central limit theorem.

One of the results of this work is to obtain a free analogue of this HSI inequality, involving a non-commutative analogue of the Stein discrepancy. We shall use this inequality to establish a rate of convergence in entropy for the multivariate free Central Limit Theorem.

Other free functional inequalities for which improvements could be investigated include the one-dimensional Brunn-Minkowski inequality of [22] and the one-dimensional transport-entropy inequality of [19].

**Acknowledgements.** We would like to thank Qiang Zeng for some helpful discussions we had in the early stages of this paper. We would also like to thank Dan-Virgil Voiculescu for his suggestions and encouragement.

## 1. DEFINITIONS AND NOTATION

In this section,  $M$  is a von Neumann algebra with a faithful normal state  $\varphi$ , and  $X = (x_1, \dots, x_n) \in M^n$  is an  $n$ -tuple of self-adjoint operators. We recall several free probabilistic quantities associated to  $X$  and also define the quantities which appear in the improved free LSI.

**1.1. Norms.** Recall that via the GNS construction,  $\varphi$  defines an inner product on  $M$ :

$$\langle x, y \rangle_\varphi := \varphi(y^*x), \quad x, y \in M.$$

The completion of  $M$  with respect to the induced norm  $\|\cdot\|_\varphi$  is denoted  $L^2(M, \varphi)$ . When the state  $\varphi$  is clear from the context, we will abuse the notation and write  $\|\cdot\|_2$  and  $L^2(M)$ .

For the von Neumann tensor product  $M \bar{\otimes} M^{op}$  with faithful normal state  $\varphi \otimes \varphi^{op}$ , we can consider the Hilbert space  $L^2(M \bar{\otimes} M^{op}, \varphi \otimes \varphi^{op})$ . This space can be identified with  $\text{HS}(L^2(M))$  (the Hilbert-Schmidt operators on  $L^2(M)$ ) via the following map

$$x \otimes y^{op} \mapsto \langle y, \cdot \rangle_2 x, \quad x, y \in M.$$

In light of this identification, we will write  $\langle \cdot, \cdot \rangle_{\text{HS}}$  for  $\langle \cdot, \cdot \rangle_{\varphi \otimes \varphi^{op}}$ , when  $\varphi$  is clear from the context.

For the  $n$ -tuple  $X$  we define  $\|X\| := \max_j \|x_j\|$ . We will also write  $C^*(X)$  and  $W^*(X)$  for the  $C^*$ -algebra and von Neumann algebra generated by  $x_1, \dots, x_n$ , respectively. When  $\varphi$  is clear, we write  $\text{Var}(X)$  for  $\sum_{j=1}^n \varphi(x_j^* x_j)$ . Given another  $n$ -tuple  $Y = (y_1, \dots, y_n) \in M^n$ , we write

$$\langle X, Y \rangle_\varphi = \sum_{j=1}^n \langle x_j, y_j \rangle_\varphi.$$

When  $\varphi$  is clear we write  $\langle X, Y \rangle_2$ .

For  $A, B \in L^2(M_n(M \bar{\otimes} M^{op}), (\varphi \otimes \varphi^{op}) \circ \text{Tr})$  (the  $n \times n$  matrices with entries in  $L^2(M \bar{\otimes} M^{op}, \varphi \otimes \varphi^{op})$ ), we write

$$\langle A, B \rangle_{(\varphi \otimes \varphi^{op}) \circ \text{Tr}} = \sum_{j,k=1}^n \langle [A]_{jk}, [B]_{jk} \rangle_{\varphi \otimes \varphi^{op}}.$$

Again, when  $\varphi$  is clear we simply write  $\langle A, B \rangle_{\text{HS}}$ .

**1.2. Formal power series and non-commutative derivations.** We let  $t_1, \dots, t_n$  be non-commuting self-adjoint indeterminates, collected as the  $n$ -tuple  $T = (t_1, \dots, t_n)$ . We let  $\mathcal{P}$  denote the set of polynomials in these indeterminates. For a polynomial  $p \in \mathcal{P}$  and a monomial  $m$ , we let  $c_m(p) \in \mathbb{C}$  denote the coefficient of  $m$  in  $p$ . After [18], for each  $R > 0$  we define the following norm

$$\|p\|_R = \sum_m |c_m(p)| R^{\deg(m)},$$

where the (finite) sum is over all monomials appearing in  $p$ . The completion of  $\mathcal{P}$  with respect to the  $\|\cdot\|_R$ -norm, a Banach algebra, is denoted  $\mathcal{P}^{(R)}$  and can be thought of power series with radius of convergence at least  $R$ . Given an  $n$ -tuple of power series  $F = (f_1, \dots, f_n) \in (\mathcal{P}^{(R)})^n$ , we define  $\|F\|_R := \max_j \|f_j\|_R$ .

After [39], for each  $j = 1, \dots, n$  the  $j$ -th cyclic derivative  $\mathcal{D}_j$  is defined on monomials  $m \in \mathcal{P}$  by

$$\mathcal{D}_j m = \sum_{m=at_j b} ba,$$

and extended linearly to all of  $\mathcal{P}$ . Then for  $p \in \mathcal{P}$ ,

$$\mathcal{D}p = (\mathcal{D}_1 p, \dots, \mathcal{D}_n p)$$

is the cyclic derivative of  $p$ . Similarly, for each  $j = 1, \dots, n$ , the  $j$ -th free difference quotient  $\partial$  is defined on monomials  $m \in \mathcal{P}$  by

$$\partial_j m = \sum_{m=at_j b} a \otimes b^{op},$$

and extended linearly to all of  $\mathcal{P}$ . For  $P = (p_1, \dots, p_n) \in \mathcal{P}^n$ , we define

$$\mathcal{J}P = \begin{pmatrix} \partial_1 p_1 & \partial_2 p_1 & \cdots & \partial_n p_1 \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 p_n & \partial_n p_n & \cdots & \partial_n p_n \end{pmatrix} \in M_n(\mathcal{P} \otimes \mathcal{P}^{op}).$$

By [28, Corollary 3.8],  $\mathcal{J}\mathcal{D}$  extends to  $\mathcal{P}^{(R)}$ , for any  $R > 0$ , and is valued in the Banach algebra  $M_n(\mathcal{P} \hat{\otimes}_R \mathcal{P}^{op})$ . Here  $\mathcal{P} \hat{\otimes}_R \mathcal{P}^{op}$  is the completion of the algebraic tensor product  $\mathcal{P}^{(R)} \otimes (\mathcal{P}^{(R)})^{op}$  with respect to the projective tensor norm (see [28, Section 3.1] for further details).

Given  $p \in \mathcal{P}$ , we will write  $p(X)$  for the image of  $p$  under the map defined by  $t_{i_1} \cdots t_{i_d} \mapsto x_{i_1} \cdots x_{i_d}$ . For any  $R \geq \|X\|$ , this map extends to  $\mathcal{P}^{(R)}$ , on which we use the same notation. We let  $\mathcal{P}(X)^{(R)}$  denote the image of this extension, which we note lies in  $C^*(X)$ .

Similarly, given another  $n$ -tuple  $Y = (y_1, \dots, y_n) \in M^n$  and  $\eta \in \mathcal{P} \otimes \mathcal{P}^{op}$ , we write  $\eta[X, Y]$  for the image of  $\eta$  under the map  $t_{i_1} \cdots t_{i_d} \otimes t_{j_1} \cdots t_{j_e} \mapsto x_{i_1} \cdots x_{i_d} \otimes y_{j_1} \cdots y_{j_e}$ . When  $Y = X$ , we simply write  $\eta[X]$ . For  $R \geq \|X\|, \|Y\|$ , this map extends to  $\mathcal{P} \hat{\otimes}_R \mathcal{P}^{op}$ , on which we use the same notation. We let  $\mathcal{P}(X) \hat{\otimes}_R \mathcal{P}(Y)$  denote the image of this extension, which we note lies in  $C^*(X) \otimes_{\min} C^*(Y)^{op}$ .

**1.3. Relative free Fisher information.** Recall from [39] that the conjugate variables to  $X$  are vectors  $\xi_{x_1}, \dots, \xi_{x_n} \in L^2(W^*(X), \varphi)$  satisfying for each  $j = 1, \dots, n$  and each  $p \in \mathcal{P}$

$$\langle \xi_{x_j}, p(X) \rangle_2 = \langle 1 \otimes 1^{op}, [\partial_j p](X) \rangle_{\text{HS}}. \quad (2)$$

The free Fisher information for  $x_1, \dots, x_n$  is defined in [39] as the quantity

$$\Phi^*(x_1, \dots, x_n) = \sum_{j=1}^n \|\xi_{x_j}\|_2^2,$$

when the conjugate variables  $\xi_{x_1}, \dots, \xi_{x_n}$  to  $X$  exist, and as  $+\infty$  otherwise. We will also denote this quantity by  $\Phi^*(X)$ .

**Definition 1.1.** Given  $\mu \in \mathbb{R}$  and  $\sigma > 0$ , by a free  $(\mu, \sigma^2)$ -semicircular  $n$ -tuple we shall mean an  $n$ -tuple  $(s_1, \dots, s_n)$  of semicircular operators  $s_1, \dots, s_n$  that are freely independent from each other and satisfy  $\varphi(s_j) = \mu$  and  $\varphi(s_j^2) = \sigma^2$  for each  $j = 1, \dots, n$ .

The conjugate variables to a free  $(0, 1)$ -semicircular  $n$ -tuple  $S = (s_1, \dots, s_n)$  are simply  $s_1, \dots, s_n$ . Hence  $\Phi^*(S) = \text{Var}(S) = n$ . More generally, for  $\rho > 0$ , the conjugate variables to  $\frac{1}{\sqrt{\rho}}S$  (a free  $(0, \rho^{-1})$ -semicircular  $n$ -tuple) are  $\sqrt{\rho}s_1, \dots, \sqrt{\rho}s_n$  so that  $\Phi^*(\frac{1}{\sqrt{\rho}}S) = n\rho$ .

**Remark 1.2.** Recall that if the conjugate variables to  $X$  exist and  $R > \|X\|$ , then by [26, Theorem 2.5] and [13, Lemma 37] the map  $\mathcal{P}^{(R)} \ni f \mapsto f(X)$  is injective. Consequently,  $\mathcal{D}_j$ ,  $\mathcal{D}$ , and  $\partial_j$  induce derivations on  $\mathcal{P}(X)^{(R)}$ , which we denote  $\mathcal{D}_{x_j}$ ,  $\mathcal{D}_X$ , and  $\partial_{x_j}$ , respectively. Similarly,  $\mathcal{J}$  induces a derivation valued in  $M_n(\mathcal{P}(X) \hat{\otimes}_R \mathcal{P}(X))$ , which we denote by  $\mathcal{J}_X$ .

For  $V \in \mathcal{P}^{(R)}$ ,  $R > \|X\|$ , we say that the joint law of  $X$  with respect to  $\varphi$  is a *free Gibbs state with potential*  $V$  if for each  $j = 1, \dots, n$  and each  $p \in \mathcal{P}$

$$\langle [\mathcal{D}_j V](X), p(X) \rangle_2 = \langle (1 \otimes 1^{op}, [\partial_j p](X)) \rangle_{\text{HS}}.$$

That is, if the conjugate variables to  $X$  are given by  $[\mathcal{D}_1 V](X), \dots, [\mathcal{D}_n V](X)$ . Equivalently, the following equation holds for all  $P \in \mathcal{P}^n$ :

$$\langle [\mathcal{D}V](X), P(X) \rangle_2 = \langle (1 \otimes 1^{op}) \otimes I_n, [\mathcal{J}P](X) \rangle_{\text{HS}},$$

where  $I_n \in M_n(\mathbb{C})$  is the  $n \times n$  identity matrix. For some simple examples, consider for each  $\rho > 0$  the potential

$$V_\rho := \frac{\rho}{2} \sum_{j=1}^n t_j^2 \in \mathcal{P}.$$

Then the joint law of a free  $(0, \rho^{-1})$ -semicircular  $n$ -tuple is a free Gibbs state with potential  $V_\rho$ .

**Definition 1.3.** For  $X = (x_1, \dots, x_n) \in M^n$ , let  $V \in \mathcal{P}^{(R)}$  for  $R > \|X\|$ . We define the *free Fisher information of  $x_1, \dots, x_n$  relative to  $V$*  to be the quantity

$$\Phi^*(x_1, \dots, x_n | V) := \sum_{j=1}^n \|\xi_{x_j} - [\mathcal{D}_j V](X)\|_2^2$$

if the conjugate variables  $\xi_{x_1}, \dots, \xi_{x_n}$  to  $X$  exist, and as  $+\infty$  otherwise. We also denote this quantity by  $\Phi^*(X | V)$ . In this context, we refer to  $V$  as a *potential*.

The quantity  $\Phi^*(X | V)$  is meant to measure how close the conjugate variables  $\xi_{x_j}$  are to  $[\mathcal{D}_j V](X)$ , and so measures in some sense how close the joint law of  $X$  is to a free Gibbs state with potential  $V$ .

**1.4. Relative non-microstates free entropy.** Let  $S = (s_1, \dots, s_n)$  be a free  $(0, 1)$ -semicircular  $n$ -tuple, free from  $x_1, \dots, x_n$ . Then the *non-microstates free entropy of  $x_1, \dots, x_n$*  is defined in [39] to be the quantity

$$\chi^*(x_1, \dots, x_n) = \frac{1}{2} \int_0^\infty \left( \frac{n}{1+t} - \Phi^*(X + \sqrt{t}S) \right) dt + \frac{n}{2} \log(2\pi e),$$

which we also denote by  $\chi^*(X)$ .

If  $X$  is a free  $(0, \rho^{-1})$ -semicircular family, then  $X + \sqrt{t}S$  is a free  $(0, \rho^{-1} + t)$ -semicircular family and therefore  $\Phi^*(X + \sqrt{t}S) = n(\rho^{-1} + t)^{-1}$ . From this it is easy to compute  $\chi^*(X) = \frac{n}{2} \log(2\pi e \rho^{-1})$ .

**Definition 1.4.** For  $X = (x_1, \dots, x_n) \in M^n$ , let  $V \in \mathcal{P}^{(R)}$  for  $R \geq \|X\|$ , we define the *non-microstates free entropy of  $x_1, \dots, x_n$  relative to  $V$*  to be the quantity

$$\chi^*(x_1, \dots, x_n | V) := \varphi(V(X)) - \chi^*(x_1, \dots, x_n),$$

which we will also denote by  $\chi^*(X | V)$ . In this context, we refer to  $V$  as a *potential*.

Since  $\chi^*(\cdot)$  is maximized (for fixed variance) by an  $n$ -tuple of free semicircular operators, it is easy to see that  $\chi^*(\cdot | V_\rho)$  is minimized by a free  $(0, \rho^{-1})$ -semicircular  $n$ -tuple.

**1.5. Relative microstates free entropy.** For each  $k \in \mathbb{N}$ , let  $M_k^{sa}$  denote the space of  $k \times k$  self-adjoint matrices. For  $l \in \mathbb{N}$  and  $\epsilon > 0$  consider the set

$$\Gamma(X; k, l, \epsilon) = \{Y \in (M_k^{sa})^n : |\frac{1}{n} \text{Tr}(p(Y)) - \varphi(p(X))| < \epsilon, \forall p \in \mathcal{P} \text{ with } \deg(p) \leq l\}.$$

The *microstates free entropy of  $x_1, \dots, x_n$*  is then defined in [37] as the quantity

$$\chi(x_1, \dots, x_n) = \inf_{\epsilon, l} \limsup_{k \rightarrow \infty} \frac{1}{k^2} \log(\text{Vol}(\Gamma(X; k, l, \epsilon))) + \frac{n}{2} \log k,$$

which we will also denote by  $\chi(X)$ .

**Definition 1.5.** For  $X = (x_1, \dots, x_n) \in M^n$ , let  $V \in \mathcal{P}^{(R)}$  for  $R \geq \|X\|$ , we define the *microstates free entropy of  $x_1, \dots, x_n$  relative to  $V$*  to be the quantity

$$\chi(x_1, \dots, x_n | V) := \varphi(V(X)) - \chi(x_1, \dots, x_n),$$

which we will also denote by  $\chi(X | V)$ . In this context, we refer to  $V$  as a *potential*.

The functional  $\chi(\cdot | V)$  has been previously considered in [40, Section 3.7] and [18, Theorem 5.1], where the following lemma was observed.

**Lemma 1.6.** *Let  $\varphi = \tau$  be a trace. For  $V = V^* \in \mathcal{P}^{(R)}$ , suppose  $Y = (y_1, \dots, y_n) \in M^n$  satisfies  $\|Y\| < R$  and minimizes  $\chi(\cdot | V)$ . Then joint law of  $Y$  is a free Gibbs state with potential  $V$ .*

*Proof.* Let  $P = (p_1, \dots, p_n) \in \mathcal{P}^n$  be an  $n$ -tuple of polynomials. By [38, Proposition 1.3], we have

$$\frac{d}{dt} \chi(y_1 + tp_1(Y), \dots, y_n + tp_n(Y)) |_{t=0} = \sum_{j=1}^n \tau \otimes \tau^{op}([\partial_j p_j](Y)) = (\tau \otimes \tau^{op}) \circ \text{Tr}([\mathcal{J}P](Y)).$$

It is also easy to compute that

$$\frac{d}{dt} \tau(V(y_1 + tp_1(Y), \dots, y_n + tp_n(Y))) |_{t=0} = \sum_{j=1}^n \tau([\partial_j V](Y) \# p_j(Y)) = \tau([\mathcal{D}V](Y) \cdot P(Y)),$$

where  $(a \otimes b^{op}) \# c = acb$ . So, by virtue of minimizing the microstates free entropy relative to  $V$  we have

$$0 = \frac{d}{dt} \chi(y_1 + tp_1(Y), \dots, y_n + tp_n(Y) | V) |_{t=0} = \tau([\mathcal{D}V](Y) \cdot P(Y)) - (\tau \otimes \tau^{op}) \circ \text{Tr}([\mathcal{J}P](Y));$$

that is, the joint law of  $y_1, \dots, y_n$  is a free Gibbs state with potential  $V$ .  $\square$

**1.6. Free Stein Discrepancy.** In this section, we introduce a free analogue of the Stein discrepancy. In commutative probability, the Stein discrepancy is a way to measure how far away a probability measure is from a given reference measure by looking at how badly it violates a set of relations that characterize the reference measure. These relations are often defined via an integration by parts formula the reference measure satisfies. For the Gaussian measure, the formula usually employed is

$$\int x \cdot \nabla f d\gamma = \int \Delta f d\gamma \quad \forall f \in C_c^\infty(\mathbb{R}^n). \quad (3)$$

In general, one finds such an integration by parts formula by characterizing the reference measure  $\mu$  as the unique invariant probability measure of a well-chosen reversible Markov process with generator  $\mathcal{L}$ , and then using the relation  $\int \mathcal{L}f d\mu = 0 \quad \forall f$ . For the Gaussian measure, we get the above relation when using the Ornstein-Uhlenbeck flow as such a Markov process.

**Definition 1.7.** Let  $\nu$  be a centered probability measure on  $\mathbb{R}^n$ . The Stein kernel of  $\nu$  with respect to the Gaussian measure  $\gamma$  on  $\mathbb{R}^n$  is the matrix-valued function  $\tau : \mathbb{R}^n \rightarrow S_n(\mathbb{R})$  such that

$$\int x \cdot \nabla f d\nu = \int \langle \tau, \text{Hess } f \rangle_{HS} d\nu. \quad (4)$$

The Stein discrepancy of  $\nu$  with respect to  $\gamma$  is then given by

$$S(\nu | \gamma) := \left( \int \|\tau - Id\|_{HS}^2 d\nu \right)^{1/2}$$

if the Stein kernel exists, and  $+\infty$  if not.

The Stein kernel is a way to reformulate the integration by parts formula. It is easy to see that a probability measure satisfies (3) for all  $f$  if and only if its Stein kernel is the identity. We also note that for a Stein kernel with respect to the Gaussian to exist,  $\nu$  must be centered. We refer to [23] and the references therein for more about the Stein method in classical probability.

We now introduce the non-commutative analogues of the Stein kernel and Stein discrepancy:

**Definition 1.8.** For  $X = (x_1, \dots, x_n) \in M^n$ , let  $V \in \mathcal{P}^{(R)}$  for  $R > \|X\|$ . An element

$$A \in L^2(M_n(M \bar{\otimes} M^{op}), (\varphi \otimes \varphi^{op}) \circ \text{Tr})$$

is said to be a *free Stein kernel for  $X$  relative to  $V$  (with respect to  $\varphi$ )* if

$$\langle [\mathcal{D}V](X), P(X) \rangle_\varphi = \langle A, [\mathcal{J}P](X) \rangle_{(\varphi \otimes \varphi^{op}) \circ \text{Tr}}, \quad \forall P \in \mathcal{P}^n. \quad (5)$$

The *free Stein discrepancy of  $X$  relative to  $V$*  is defined as the quantity

$$\Sigma^*(X | V) = \inf_A \|A - (1 \otimes 1^{op}) \otimes I_n\|_{(\varphi \otimes \varphi^{op}) \circ \text{Tr}},$$

where the infimum is taken over the set of Stein kernels of  $X$  relative to  $V$ .

Observe that (5) is not the exact analogue of (4). Indeed, the exact analogue would be

$$\langle [\mathcal{D}V](X), [\mathcal{D}p](X) \rangle_\varphi = \langle A, [\mathcal{J}\mathcal{D}p](X) \rangle_{(\varphi \otimes \varphi^{op}) \circ \text{Tr}}, \quad \forall p \in \mathcal{P},$$

which is a weaker condition. Fortunately, all of the examples in Section 3 satisfy our stronger condition and therefore offer applications of the free HSI. We use this stronger condition in the proof of Lemma 2.4, but the weaker condition would suffice if a conjecture of Voiculescu receives even a partially affirmative answer (see Remark 2.5).

We note that since  $[\mathcal{J}(\mathcal{P}^n)](X)$  need not be dense in  $L^2(M_n(M \bar{\otimes} M^{op}), (\varphi \otimes \varphi^{op}) \circ \text{Tr})$ , the free Stein kernel may not be unique. Nevertheless, so long as at least one free Stein kernel exists, then its projection onto the  $L^2$ -closure of  $[\mathcal{J}(\mathcal{P}^n)](X)$  is also a free Stein kernel which attains the infimum in the free Stein discrepancy. Furthermore, even the Stein kernel of a probability measure  $\nu$  on  $\mathbb{R}^d$  need not be unique for  $d > 1$ : the space of Hessians is not dense.

When  $\Phi^*(X) < +\infty$ , Remark 1.2 implies that (5) is equivalent to  $A \in \text{dom}(\mathcal{J}_X^*)$  with  $\mathcal{J}_X^*(A) = [\mathcal{D}V](X)$ . Also note that when the potential  $V$  is self-adjoint and  $\varphi$  is a trace, equation (5) implies  $A^\dagger = A$ , where  $[A^\dagger]_{ij} = [A]_{ij}^\dagger$  and  $(a \otimes b^{op})^\dagger = b^* \otimes (a^*)^{op}$  for  $a, b \in M$ .

Just as with the free Fisher information relative to some potential  $V$ , the free Stein discrepancy relative to  $V$  measures in some sense how close the  $n$ -tuple is to having as its joint law a free Gibbs state with potential  $V$ . Indeed, by the definition of a free Gibbs state,  $X$  has such a joint distribution if and only if  $(1 \otimes 1^{op}) \otimes I_n$  is a free Stein kernel for  $X$  and consequently  $\Sigma^*(X | V) = 0$ . In particular, we have  $\Sigma^*(\frac{1}{\sqrt{\rho}}S | V_\rho) = 0$ . We will denote  $(1 \otimes 1^{op}) \otimes I_n$  by  $1$  when the context precludes any confusion.

## 2. AN IMPROVED (NON-MICROSTATES) FREE LOG-SOBOLEV INEQUALITY

We now fix a von Neumann algebra  $M$  with a faithful normal trace  $\tau$  and an  $n$ -tuple  $X = (x_1, \dots, x_n)$  of self-adjoint operators. Provided they exist, the conjugate variables to  $X$  will always be denoted  $\xi_1, \dots, \xi_n$ . Let  $S = (s_1, \dots, s_n) \in M^n$  be a free  $(0, 1)$ -semicircular  $n$ -tuple, free from  $x_1, \dots, x_n$ . Fix  $\rho > 0$  and recall the potential

$$V_\rho = \frac{\rho}{2} \sum_{j=1}^n t_j^2.$$

For each  $t \geq 0$  and  $j = 1, \dots, n$ , define  $x_j(t) := e^{-t}x_j + \sqrt{1 - e^{-2t}}\frac{1}{\sqrt{\rho}}s_j$ , collected as the  $n$ -tuple  $X(t) = (x_1(t), \dots, x_n(t))$ . Let  $\mathcal{E}_t: M \rightarrow W^*(X(t))$  denote the conditional expectation. By [39, Proposition 3.7], the conjugate variables to  $X(t)$  always exist, and if we denote them by  $\xi_1(t), \dots, \xi_n(t)$  then for each  $j = 1, \dots, n$

$$\xi_j(t) = \frac{\sqrt{\rho}}{\sqrt{1 - e^{-2t}}}\mathcal{E}_t(s_j).$$

Moreover, if the conjugate variables to  $X$  exist then we have

$$\xi_j(t) = e^t\mathcal{E}_t(\xi_j).$$

We collect the conjugate variables to  $X(t)$  as the  $n$ -tuple  $\Xi(t) = (\xi_1(t), \dots, \xi_n(t))$ .

**2.1. De Bruijn's formula and a free log-Sobolev inequality.** The following lemma gives the non-commutative version of de Bruijn's formula.

**Lemma 2.1.** *For  $\rho > 0$  and  $X(t) = (x_1(t), \dots, x_n(t))$  as above,*

$$\frac{d}{dt}\chi^*(X(t) | V_\rho) = -\frac{1}{\rho}\Phi^*(X(t) | V_\rho).$$

*Proof.* We first note that because  $\mathcal{D}_j V_\rho = \rho t_j$ , we have

$$\begin{aligned}\Phi^*(X(t) | V_\rho) &= \sum_{j=1}^n \|\xi_j(t) - \rho x_j(t)\|_2^2 \\ &= \sum_{j=1}^n \|\xi_j(t)\|_2^2 - 2\rho + \rho^2(e^{-2t}\|x_j\|_2^2 + (1 - e^{-2t})\rho^{-1}) \\ &= \Phi^*(X(t)) - n\rho + \rho^2 e^{-2t} \text{Var}(X) - n\rho e^{-2t}.\end{aligned}$$

Next we compute

$$\frac{d}{dt}\tau(V_\rho(X(t))) = \frac{d}{dt}\left[\frac{\rho}{2}e^{-2t}\text{Var}(X) + \frac{n}{2}(1 - e^{-2t})\right] = -\rho e^{-2t}\text{Var}(X) + ne^{-2t}.$$

By [39, Proposition 7.8],

$$\chi^*(X(t)) = -nt + \chi^*\left(X + \sqrt{\rho^{-1}(e^{2t} - 1)}S\right).$$

Let  $r(t) = \rho^{-1}(e^{2t} - 1)$ . Then by [39, Proposition 7.5] we have

$$\begin{aligned}\frac{d}{dt}\chi^*(X(t)) &= -n + \frac{1}{2}\Phi^*(X + \sqrt{r(t)}S)\frac{dr(t)}{dt} \\ &= -n + \rho^{-1}e^{2t}\Phi^*\left(X + \sqrt{e^{2t} - 1}\frac{1}{\sqrt{\rho}}S\right) \\ &= -n + \rho^{-1}\Phi^*(X(t)).\end{aligned}$$

Thus the claimed equality holds. □

**Lemma 2.2.** *For  $\rho > 0$  and  $X(t) = (x_1(t), \dots, x_n(t))$  as above,*

$$\Phi^*(X(t) | V_\rho) \leq e^{-2t}\Phi^*(X | V_\rho).$$

*Proof.* If  $\Phi^*(X) = +\infty$ , then there is nothing to prove. So assume  $\Phi^*(X) < +\infty$  and let  $\xi_1, \dots, \xi_n$  denote the conjugate variables. Recall that for each  $j = 1, \dots, n$

$$\xi_j(t) = e^t \mathcal{E}_t(\xi_j) = \frac{\sqrt{\rho}}{\sqrt{1 - e^{-2t}}} \mathcal{E}_t(s_j).$$

As observed in Lemma 2.1,

$$\Phi^*(X(t) | V_\alpha) = \sum_{j=1}^n \|\xi_j(t)\|_2^2 - \rho(1 + e^{-2t}) + \rho^2 e^{-2t} \|x_j\|_2^2$$

Now, for each  $j = 1, \dots, n$  we have

$$\begin{aligned}\|\xi_j(t)\|_2^2 &= \left\| e^{-2t} e^t \mathcal{E}_t(\xi_j) + (1 - e^{-2t}) \frac{\sqrt{\rho}}{\sqrt{1 - e^{-2t}}} \mathcal{E}_t(s_j) \right\|_2^2 \\ &= \left\| \mathcal{E}_t\left(e^{-t}\xi_j + \sqrt{\rho(1 - e^{-2t})}s_j\right) \right\|_2^2 \\ &\leq \left\| e^{-t}\xi_j + \sqrt{\rho(1 - e^{-2t})}s_j \right\|_2^2 \\ &= e^{-2t}\|\xi_j\|_2^2 + \rho(1 - e^{-2t}).\end{aligned}$$



So, resuming our computation from above we see

$$\begin{aligned}\Phi^*(X(t) | V_\rho) &\leq \sum_{j=1}^n e^{-2t} \|\xi_j\|^2 + \rho(1 - e^{-2t}) - \rho(1 + e^{-2t}) + \rho^2 e^{-2t} \|x_j\|_2^2 \\ &= e^{-2t} \left( \sum_{j=1}^n \|\xi_j\|_2^2 - 2\rho + \rho^2 \|x_j\|_2^2 \right) \\ &= e^{-2t} \Phi^*(X | V_\rho),\end{aligned}$$

as claimed.  $\square$

From these two lemmas we obtain the non-commutative analogue of the logarithmic-Sobolev inequality from classical probability. In particular, this yields an alternate proof for an inequality from [3, Section 7.2] stated here in the multi-variable case.

**Corollary 2.3.** *For  $X$  an  $n$ -tuple of non-commutative random variables,  $S$  a free  $(0, 1)$ -semicircular  $n$ -tuple that is free from  $X$ , and any  $\rho > 0$*

$$\chi^*(X | V_\rho) - \chi^*(\sqrt{\rho}^{-1}S | V_\rho) \leq \frac{1}{2\rho} \Phi^*(X | V_\rho).$$

In particular, if  $\text{Var}(X) = \frac{n}{\rho}$  then

$$\frac{n}{2} + \frac{n}{2} \log \left( \frac{2\pi e}{\rho} \right) - \chi^*(X) \leq \frac{1}{2\rho} \Phi^*(X).$$

*Proof.* Using Lemmas 2.1 and 2.2 we have

$$\begin{aligned}\chi^*(X | V_\rho) - \chi^*(\sqrt{\rho}^{-1}S | V_\rho) &= \chi^*(X(0) | V_\rho) - \lim_{t \rightarrow \infty} \chi^*(X(t) | V_\rho) \\ &= \int_0^\infty \rho^{-1} \Phi^*(X(t) | V_\rho) dt \\ &\leq \rho^{-1} \Phi^*(X | V_\rho) \int_0^\infty e^{-2t} dt = \frac{1}{2\rho} \Phi^*(X | V_\rho).\end{aligned}$$

The second inequality follows by expanding the definitions of the relative free entropy and Fisher information (and using  $\chi^*(\sqrt{\rho}^{-1}S) = \frac{n}{2} \log \left( \frac{2\pi e}{\rho} \right)$ ).  $\square$

## 2.2. Improving the free log-Sobolev Inequality via Stein's Method.

**Lemma 2.4.** *Suppose  $\Phi^*(X | V_\rho) < +\infty$ . For  $\rho > 0$  and  $X(t) = (x_1(t), \dots, x_n(t))$  as above*

$$\frac{1}{\rho} \Phi^*(X(t) | V_\rho) \leq \frac{e^{-4t}}{1 - e^{-2t}} \Sigma^*(X | V_\rho)^2$$

*Proof.* If  $\Sigma^*(X | V_\rho) = +\infty$ , there is nothing to prove. So assume  $\Sigma^*(X | V_\rho) < +\infty$ , and let  $A \in L^2(M_n(W^*(X) \bar{\otimes} W^*(X)^{op}))$  be a free Stein kernel for  $X$  relative to  $V_\rho$  such that

$$\|A - 1\|_{HS} = \Sigma^*(X | V_\rho).$$

Recall that since  $\Phi^*(X) < +\infty$ , this implies  $\mathcal{J}_X^*(A) = \rho X$ . Let  $\xi_1, \dots, \xi_n \in L^2(W^*(X))$  be the conjugate variables of  $X$ , collected in a vector as  $\Xi = (\xi_1, \dots, \xi_n)$  so that  $\mathcal{J}_X^*(1) = \Xi$ . Now, we compute:

$$\begin{aligned}\Phi^*(X(t) | V_\rho) &= \sum_{j=1}^n \|\xi_j(t) - \rho x_j(t)\|_2^2 \\ &= \sum_{j=1}^n \left\langle e^{-t} \xi_j + \sqrt{\rho(1 - e^{-2t})} s_j - \rho x_j(t), \xi_j(t) - \rho x_j(t) \right\rangle_2 \\ &= e^{-t} \sum_{j=1}^n \langle \xi_j - \rho x_j, \xi_j(t) - \rho x_j(t) \rangle_2 \\ &= e^{-t} \langle \Xi - \rho X, \Xi(t) - \rho X(t) \rangle_2 \\ &= e^{-t} \langle \mathcal{J}_X^*[1 - A], \Xi(t) - \rho X(t) \rangle_2.\end{aligned}\tag{6}$$

Recall that on  $\mathbb{C}\langle x_1(t), \dots, x_n(t) \rangle$

$$e^t \mathcal{J}_X = \mathcal{J}_{X(t)} = \frac{\sqrt{\rho}}{\sqrt{1 - e^{-2t}}} \mathcal{J}_S,$$

so that for any  $Z \in \mathcal{E}_t(M)^n$  we have

$$\langle \mathcal{J}_X^*[1 - A], Z \rangle_2 = \frac{e^{-t} \sqrt{\rho}}{\sqrt{1 - e^{-2t}}} \langle \mathcal{J}_S^*[1 - A], Z \rangle_2.$$

So continuing our computation we have

$$\begin{aligned} \Phi^*(X(t) | V_\rho) &= \frac{e^{-2t} \sqrt{\rho}}{\sqrt{1 - e^{-2t}}} \langle \mathcal{J}_S^*[1 - A], \Xi(t) - \rho X(t) \rangle_2 \\ &\leq \frac{e^{-2t} \sqrt{\rho}}{\sqrt{1 - e^{-2t}}} \|\mathcal{J}_S^*[1 - A]\|_2 \Phi^*(X(t) | V_\rho)^{1/2}. \end{aligned}$$

Now, since  $\partial_{s_j} |_{W^*(X)} \equiv 0$  for each  $j = 1, \dots, n$  we have by [39, Proposition 4.1]

$$\mathcal{J}_S^*[1 - A] = \left( \sum_{j=1}^n \partial_{s_j}^* [\delta_{i=j} 1 \otimes 1 - A_{ij}] \right)_{i=1}^n = \left( s_i - \sum_{j=1}^n A_{ij} \# s_j \right)_{i=1}^n = [1 - A] \# S.$$

Since  $s_1, \dots, s_n$  are free from  $W^*(X)$ , it is easy to see that

$$\|[1 - A] \# S\|_2 = \|1 - A\|_{HS} = \Sigma^*(X | V_\rho)$$

Thus we have

$$\Phi^*(X(t) | V_\rho) \leq \frac{e^{-2t} \sqrt{\rho}}{\sqrt{1 - e^{-2t}}} \Sigma^*(X | V_\rho) \Phi^*(X(t) | V_\rho)^{1/2},$$

which is equivalent to the desired inequality.  $\square$

**Remark 2.5.** The equality on line (6) is precisely where we use the stronger version of the definition of a free Stein kernel. It is a conjecture of Voiculescu that conjugate variables always lie in the  $L^2$ -closure of cyclic gradients. If this even weakly true for  $\Xi(t)$ , then the equality on line (6) will hold for the weaker version of the definition of a free Stein kernel.

With the previous lemmas in hand, the following proof is, aside from notation, identical to that of [23, Theorem 2.2]. We still present it here, for the sake of convenience.

**Theorem 2.6.** *Let  $X = (x_1, \dots, x_n)$  be an  $n$ -tuple of non-commutative random variables, and  $S = (s_1, \dots, s_n)$  a free  $(0, 1)$ -semicircular  $n$ -tuple, free from  $X$ . Then for any  $\rho > 0$  we have*

$$\chi^*(X | V_\rho) - \chi^*(\sqrt{\rho}^{-1} S | V_\rho) \leq \frac{1}{2} \Sigma^*(X | V_\rho)^2 \log \left( 1 + \frac{\Phi^*(X | V_\rho)}{\rho \Sigma^*(X | V_\rho)^2} \right).$$

*Proof.* If  $\Phi^*(X | V_\rho) = +\infty$ , then there is nothing to prove, so we assume that  $\Phi^*(X | V_\rho) < +\infty$ . If  $\Sigma^*(X | V_\rho) = +\infty$ , then the right-hand side reduces to  $\frac{1}{2\rho} \Phi^*(X | V_\rho)$  and the inequality holds by Corollary 2.3. Thus we may assume  $\Sigma^*(X | V_\rho) < +\infty$ .

Using Lemmas 2.1, 2.2, and 2.4 we have

$$\begin{aligned} \chi^*(X | V_\rho) - \chi^*(\sqrt{\rho}^{-1} S | V_\rho) &= \int_0^\infty \rho^{-1} \Phi^*(X(t) | V_\rho) dt \\ &\leq \rho^{-1} \Phi^*(X | V_\rho) \int_0^u e^{-2t} dt + \Sigma^*(X | V_\rho)^2 \int_u^\infty \frac{e^{-4t}}{1 - e^{-2t}} dt \\ &\leq \frac{1}{2\rho} \Phi^*(X | V_\rho) (1 - e^{-2u}) + \frac{1}{2} \Sigma^*(X | V_\rho)^2 (-e^{-2u} - \log(1 - e^{-2u})). \end{aligned}$$

Optimizing in  $u$  yields  $1 - e^{-2u} = \frac{\rho \Sigma^*(X | V_\rho)^2}{\Phi^*(X | V_\rho) + \rho \Sigma^*(X | V_\rho)^2}$  and the desired inequality.  $\square$

## 3. EXAMPLES OF FREE STEIN KERNELS

In this section we present several examples  $n$ -tuples of self-adjoint operators with known free Stein kernels relative to the potential  $V_1$ , for which equation (5) reduces to

$$\langle X, P \rangle_\varphi = \langle A, \mathcal{J}_X P \rangle_{(\varphi \otimes \varphi^{op}) \circ \text{Tr}}, \quad \forall P \in \mathcal{P}(X)^n,$$

which, in turn, is equivalent to the system of equations

$$\langle x_j, p(X) \rangle_\varphi = \sum_{k=1}^n \langle [A]_{jk}, [\partial_k p](X) \rangle_{\varphi \otimes \varphi^{op}}, \quad p \in \mathcal{P},$$

for  $j = 1, \dots, n$ .

Notably, these examples represent all known applications of the free monotone transport theorem of Guionnet and Shlyakhtenko [18] (excepting the infinite variable case considered in [28]). That is, under certain hypothesis, the  $n$ -tuples of operators considered in these examples have been shown by free transport arguments to generate a von Neumann algebra isomorphic to the free group factor  $L(\mathbb{F}_n)$  (the von Neumann algebra generated by a free  $(0, 1)$ -semicircular  $n$ -tuple), or (in Example 3) a free Araki-Woods factor (the non-tracial analogue of  $L(\mathbb{F}_n)$ ). Moreover, the free Stein discrepancy being small is a necessary condition under the hypotheses used in these free transport arguments. Indeed, the free Stein discrepancy is dominated by a norm derived from the projective tensor norm on  $\mathcal{P}^{(R)} \otimes (\mathcal{P}^{(R)})^{op}$ . At the core of the proofs of each of these applications is the fact that this latter norm must be sufficiently small (so that, for example, the free Stein kernel is invertible in a particular Banach algebra). In light of this, we posit the following conjecture. See [18] for further details on free monotone transport.

**Conjecture.** Suppose  $X = (x_1, \dots, x_n)$  has joint law  $\mu_X$ , and  $A = A^\dagger$  is a free Stein kernel for  $X$  relative to  $V_1$ . Then there exists  $\epsilon > 0$  such that if  $\Sigma^*(X | V_1) < \epsilon$ , then there exists free transport from the free semicircle law to  $\mu_X$ . In particular, we can embed  $W^*(X) \hookrightarrow L(\mathbb{F}_n)$ .

Each of the following examples arises from a similar construction: Let  $\mathcal{H}_{\mathbb{R}}$  let be a real Hilbert space with orthonormal basis  $\{e_1, \dots, e_n\}$ , contained in some complex Hilbert space  $\mathcal{H}$ . Then the operators in our examples will be represented on some Hilbert space completion  $\mathcal{F}$  of the following space:

$$\mathcal{F}^{\text{finite}} := \mathbb{C}\Omega \oplus \bigoplus_{d \geq 1} \mathcal{H}^{\otimes d},$$

where  $\Omega$  is the vacuum vector (to be regarded as the generator of “ $\mathcal{H}^{\otimes 0}$ ”). We consider linear operators  $\ell(e_1), \dots, \ell(e_n)$  on  $\mathcal{F}^{\text{finite}}$  defined by

$$\begin{aligned} \ell(e_j)\Omega &= e_j \\ \ell(e_j)f_1 \otimes \dots \otimes f_d &= e_j \otimes f_1 \otimes \dots \otimes f_d, \quad f_1, \dots, f_d \in \mathcal{H}. \end{aligned}$$

These operators, which will be bounded operators on all the completions of  $\mathcal{F}^{\text{finite}}$ , are known as *left creation operators*, possibly with some additional adjective depending on the completion. Furthermore, the particular completion considered will determine the behavior of the adjoints  $\ell(e_1)^*, \dots, \ell(e_n)^*$ , which are known as *left annihilation operators* (possibly with some additional adjectives). In fact, for the sake of these examples this behavior is more enlightening than an explicit description of the inner product on  $\mathcal{F}^{\text{finite}}$  which determines the completion  $\mathcal{F}$ , and so we forgo the latter.

In the following examples, the self-adjoint operators of interest are  $x_j = \ell(e_j) + \ell(e_j)^*$ ,  $j = 1, \dots, n$ . We consider their distribution with respect to the vacuum state  $\varphi(\cdot) = \langle \cdot, \Omega \rangle_{\mathcal{F}}$ . In particular, we will describe free Stein kernels for the  $n$ -tuple  $X = (x_1, \dots, x_n)$  relative to  $V_1$ .

**Example 1:  $q$ -semicircular systems** [6]. Let  $\mathcal{H}_{\mathbb{R}} = \mathbb{R}^n$ , contained in  $\mathcal{H} = \mathbb{C}^n$ . For  $q \in [-1, 1]$ , one can define an inner product on  $\mathcal{F}^{\text{finite}}$  (see [6, Lemma 3] for an explicit formula) so that for each  $j = 1, \dots, n$

$$\begin{aligned} \ell(e_j)^*\Omega &= 0 \\ \ell(e_j)^*f_1 \otimes \dots \otimes f_d &= \sum_{k=1}^d q^{k-1} \langle f_k, e_j \rangle f_1 \otimes \dots \otimes \hat{f}_k \otimes \dots \otimes f_d, \quad f_1, \dots, f_d \in \mathbb{C}^n, \end{aligned}$$

where  $\hat{f}_k$  denotes omission from the tensor product. In this case,  $\{\ell(e_j)\}_{j=1}^n$ ,  $\{\ell(e_j)^*\}_{j=1}^n$ , and  $\{x_j\}_{j=1}^n$  are respectively known as *left  $q$ -creation operators*, *left  $q$ -annihilation operators*, and  *$q$ -semicircular operators*. Also, the vacuum state is a trace which we denote  $\tau$ . When  $q = 0$ ,  $X$  is a free  $(0, 1)$ -semicircular  $n$ -tuple.

Towards exhibiting a free Stein kernel, for each  $d \geq 1$  let  $\pi_d \in \mathcal{B}(\mathcal{F})$  be the projection onto  $(\mathbb{C}^n)^{\otimes d}$ , and let  $\pi_0 \in \mathcal{B}(\mathcal{F}_q(n))$  be the projection onto  $\mathbb{C}\Omega$ . Consider the operator

$$\Xi_q = \sum_{d \geq 0} q^d \pi_d.$$

When  $q^2 n < 1$ ,  $\Xi_q$  is a Hilbert–Schmidt operator on  $\mathcal{F}$ . Consequently, using the identification

$$\begin{aligned} L^2(W^*(X) \bar{\otimes} W^*(X)^{op}) &\cong \text{HS}(\mathcal{F}) \\ a \otimes b^\circ &\mapsto \langle \cdot, b^* \Omega \rangle_{\mathcal{F}} a \Omega, \end{aligned}$$

we can consider  $A = \Xi_q \otimes I_n \in L^2(M_n(W^*(X) \bar{\otimes} W^*(X)^{op}))$ . It was observed by Shlyakhtenko in [32] that  $\mathcal{J}_X^*(A) = X$  (see also [33, Lemma 10]). Hence  $A$  is a free Stein kernel for  $X$  with respect to  $V_1$ , and

$$\Sigma^*(X | V_1) \leq \sqrt{n} \|\Xi_q - 1 \otimes 1^\circ\|_{\tau \otimes \tau^{op}} = \frac{|q|n}{\sqrt{1 - q^2 n}}.$$

**Example 2: mixed  $q$ -semicircular systems** [7]. As before, we let  $\mathcal{H}_{\mathbb{R}} = \mathbb{R}^n$  and  $\mathcal{H} = \mathbb{C}^n$ . Given a symmetric matrix  $Q = (q_{ij}) \in M_n([-1, 1])$ , one can define an inner product on  $\mathcal{F}^{\text{finite}}$  (see [7, Section 3] for a formula) so that for each  $j = 1, \dots, n$

$$\begin{aligned} \ell(e_j)^* \Omega &= 0 \\ \ell(e_j)^* e_{i_1} \otimes \cdots \otimes e_{i_d} &= \sum_{k=1}^n q_{j i_1} \cdots q_{j i_{k-1}} \langle e_{i_k}, e_j \rangle e_{i_1} \otimes \cdots \otimes \hat{e}_{i_k} \otimes \cdots \otimes e_{i_d}. \end{aligned}$$

The operators  $x_j = \ell(e_j) + \ell(e_j)^*$  are called *mixed  $q$ -Gaussian variables*, and the vacuum state is again a trace, which we denote by  $\tau$ .

As before, a free Stein kernel will arise from Hilbert–Schmidt operators. However, in order to properly describe them we require a finer orthogonal decomposition of  $\mathcal{F}$ . For  $d \geq 1$  and  $1 \leq i_1 \leq i_2 \leq \cdots \leq i_d \leq n$  define

$$\mathcal{F}_{(i_1, \dots, i_n)} = \text{span}\{e_{i_{\sigma(1)}} \otimes \cdots \otimes e_{i_{\sigma(d)}} : \sigma \in S_d\},$$

where  $S_d$  is the permutation group on  $d$  elements. Then these subspaces along with  $\mathbb{C}\Omega$  offer an orthogonal decomposition of  $\mathcal{F}$ . Let  $\pi_{(i_1, \dots, i_d)}$  be the orthogonal projection onto  $\mathcal{F}_{(i_1, \dots, i_d)}$  and let  $\pi_0$  be the orthogonal projection onto  $\mathbb{C}\Omega$ . For each  $j = 1, \dots, n$ , define

$$\Xi_j = \pi_0 + \sum_{d \geq 1} \sum_{1 \leq i_1 \leq \cdots \leq i_d \leq n} q_{j i_1} \cdots q_{j i_d} \pi_{(i_1, \dots, i_d)}.$$

Then  $\Xi_j$  is a Hilbert–Schmidt operator if and only if  $Q_j(2) := \sum_{i=1}^n |q_{ij}|^2 < 1$ . Assume  $Q_j(2) < 1$  for each  $j = 1, \dots, n$ . Using the same identification as in Example 1, we can consider

$$A = \sum_{j=1}^n \Xi_j \otimes E_{jj} \in L^2(M_n(W^*(X) \bar{\otimes} W^*(X)^{op})),$$

where  $E_{jj}$  is the diagonal matrix with  $jj$ th entry 1 and all others zero. By [27, Proposition 2]  $\mathcal{J}_X^*(A) = X$ , and hence using [28, Lemma 5.1] we see

$$\Sigma^*(X | V_1) \leq \left( \sum_{j=1}^n \|\Xi_j - 1 \otimes 1^{op}\|_{\tau \otimes \tau^{op}}^2 \right)^{1/2} = \left( \sum_{j=1}^n \frac{Q_j(2)}{1 - Q_j(2)} \right)^{1/2}.$$

**Example 3:  $q$ -deformed Araki-Woods algebras** [20]. Once again we have  $\mathcal{H}_{\mathbb{R}} = \mathbb{R}^n$ , however we modify  $\mathcal{H}$ . Let  $\{U_t : t \in \mathbb{R}\}$  be an orthogonal representation of  $\mathbb{R}$  onto  $\mathcal{H}_{\mathbb{R}}$ , which extends to a one parameter group of unitary transformations on  $\mathbb{C}^n$ . By Stone’s theorem, there exists a closed positive operator  $B$  such that  $U_t = B^{it}$  for every  $t \in \mathbb{R}$ . We then consider the following inner product

$$\langle f, g \rangle_U := \left\langle f, \frac{2}{1 + B^{-1}} g \right\rangle, \quad f, g \in \mathbb{C}^n,$$

and let  $\mathcal{H}$  be the completion of  $\mathbb{C}^n$  with respect to  $\|\cdot\|_U$ . Recall from [31] that this yields an isometric embedding of  $\mathcal{H}_{\mathbb{R}}$  into  $\mathcal{H}$  where the restriction of  $\text{Re} \langle \cdot, \cdot \rangle_U$  to  $\mathcal{H}_{\mathbb{R}}$  recovers the original inner product. In particular, our orthonormal basis  $\{e_1, \dots, e_n\}$  for  $\mathcal{H}_{\mathbb{R}}$ , are unit vectors inside in  $\mathcal{H}$  with completely imaginary covariance.

Given  $q \in [-1, 1]$ , one defines an inner product on  $\mathcal{F}^{\text{finite}}$  (see [20, Section 1]) so that for each  $j = 1, \dots, n$

$$\begin{aligned} \ell(e_j)^* \Omega &= 0 \\ \ell(e_j)^* f_1 \otimes \cdots \otimes f_d &= \sum_{k=1}^d q^{k-1} \langle f_k, e_j \rangle_U f_1 \otimes \cdots \otimes \hat{f}_k \otimes \cdots \otimes f_d, \quad f_1, \dots, f_d \in \mathcal{H}. \end{aligned}$$

The operators  $x_j = \ell(e_j) + \ell(e_j)^*$  are called *q-quasi-free semicircular operators*, and the vacuum state  $\varphi$  (which is not a trace unless the orthogonal representation  $\{U_t : t \in \mathbb{R}\}$  is trivial) is called a *q-quasi free state*. Furthermore, the action of the modular operator  $\Delta_\varphi$  is known:

$$\Delta_\varphi x_j = \sum_{k=1}^n [B]_{jk} x_k,$$

where  $[B]_{jk}$  is the  $jk$ th entry of the matrix representation of  $B$  as a map on  $\mathcal{H}_{\mathbb{R}}$  with respect to the basis  $\{e_1, \dots, e_n\}$ . In the case  $q = 0$ , the  $x_j$  are Shlyakhtenko's *quasi-free semicircular operators* and  $\varphi$  is a *free quasi-free state* [31].

Letting  $\Xi_q$  be exactly as in Example 1 (with  $q^2 n < 1$  so it is a Hilbert–Schmidt operator), it was shown in [29, Proposition 2.2] that

$$\langle x_j, p(X) \rangle_\varphi = \sum_{k=1}^n \left\langle \left[ \frac{2}{1+B} \right]_{jk} (\Delta_\varphi \otimes \Delta_{\varphi^{op}})(\Xi_q^*), [\partial_k p](X) \right\rangle_{\varphi \otimes \varphi^{op}}, \quad p \in \mathcal{P},$$

where  $\left[ \frac{2}{1+B} \right]_{jk}$  is the  $jk$ th entry of the matrix representation of  $\frac{2}{1+B}$  as a map on  $\mathcal{H}_{\mathbb{R}}$  with respect to the basis  $\{e_1, \dots, e_n\}$ . Thus

$$A = (\Delta_\varphi \otimes \Delta_{\varphi^{op}})(\Xi_q^*) \otimes \frac{2}{1+B} \in L^2(M_n(W^*(X) \bar{\otimes} W^*(X)^{op}), (\varphi \otimes \varphi^{op}) \circ \text{Tr})$$

is a free Stein kernel for  $X$  relative to  $V_1$ . Using  $\left[ \frac{2}{1+B} \right]_{jj} = 1$  we have

$$\Sigma^*(X | V_1) \leq \left( n \|(\Delta_\varphi \otimes \Delta_{\varphi^{op}})(\Xi_q^*) - 1 \otimes 1^{op}\|_{\varphi \otimes \varphi^{op}}^2 + \sum_{j \neq k} \left| \left[ \frac{2}{1+B} \right]_{jk} \right|^2 \|(\Delta_\varphi \otimes \Delta_{\varphi^{op}})(\Xi_q^*)\|_{\varphi \otimes \varphi^{op}}^2 \right)^{1/2}.$$

It is not difficult to show that if  $\{\xi_\ell^{(d)}\}_{\ell \in I_d}$  is an orthonormal basis for  $\pi_d \mathcal{F}$ , then

$$\|(\Delta_\varphi \otimes \Delta_{\varphi^{op}})(\Xi_q^*) - 1 \otimes 1^{op}\|_{\varphi \otimes \varphi^{op}}^2 = \sum_{d \geq 1} \sum_{\ell \in I_d} q^{2d} \|\Delta_\varphi^{-1} \xi_\ell^{(d)}\|_\varphi^2$$

Thus, in the estimate for  $\Sigma^*(X | V_1)$  we see that the first term vanishes as  $q \rightarrow 0$ , whereas the sum vanishes as  $B \rightarrow 1$ . Since  $W^*(X)$  is known to be a type III factor so long as  $B \neq 1$ , this would offer a counterexample to our conjecture were it not for the fact that

$$A^\dagger = (\Delta_\varphi \otimes \Delta_{\varphi^{op}})(\Xi_q^*) \otimes \frac{2}{1+B^{-1}} \neq A.$$

#### 4. ONE-DIMENSIONAL CASE

In this section, we give the statements of our results in the one-dimensional case. The statements become simpler, in particular because we have an explicit expression for the entropy of a probability measure on  $\mathbb{R}$ :

$$\chi(\mu) = \chi^*(\mu) := \int \log |x - y| \mu(dx) \mu(dy) + \frac{3}{4} + \frac{1}{2} \log(2\pi)$$

while the relative entropy with respect to the semicircular law  $d\sigma = \frac{1}{2\pi} \sqrt{4 - x^2} dx$  (with unit variance) is given by

$$\chi(\mu|\sigma) := \frac{1}{2} \int x^2 \mu(dx) - \chi(\mu).$$

We note  $\chi(\sigma) = \frac{1}{2} \log(2\pi e)$  and so  $\chi(\sigma | \sigma) = -\frac{1}{2} \log(2\pi)$ . The one-dimensional free Fisher information with respect to  $\sigma$  is given by

$$\Phi^*(\mu|\sigma) = \int (H\mu(x) - x)^2 \mu(dx)$$

where  $H\mu(x) = \int (x - y)^{-1} \mu(dy)$  is the Hilbert transform of  $\mu$ . The LSI then takes the form

$$\chi(\mu|\sigma) + \frac{1}{2} \log(2\pi) \leq \frac{1}{2} \Phi^*(\mu|\sigma).$$

In addition to the general proof of Biane and Speicher [3], several other techniques have been successfully applied to obtain this one-dimensional inequality. Ledoux also proved it using a free analogue of the Brunn-Minkowski inequality [22] (which itself was also proven using a random matrix approximation). Finally, Ledoux and Popescu [25] gave a proof using mass transport, adapting arguments of Cordero-Erausquin [12] in the classical case.

The Stein factor  $A$  associated to a probability measure (and relative to the semicircular law) is a function of two variables, such that

$$\int xP(x)\mu(dx) = \int \int A(x, y) \tilde{P}(x, y) \mu(dx) \mu(dy)$$

for all polynomials  $P$ , where  $\tilde{P}$  is the free difference quotient, which takes value  $\tilde{P}(x, y) = P'(x)$  if  $x = y$ , and  $\frac{P(x) - P(y)}{x - y}$  otherwise. The measure  $\mu$  is semicircular with unit variance iff  $A$  is constantly equal to 1, and the Stein discrepancy is given by

$$\Sigma^*(\mu | \sigma) := \left( \int |A(x, y) - 1|^2 \mu(dx) \mu(dy) \right)^{1/2}.$$

In this context, the free HSI inequality takes the following form:

**Corollary 4.1.** *For any centered measure  $\mu$ , given its Stein kernel  $A$ , we have*

$$\begin{aligned} & \frac{1}{2} \int x^2 \mu(dx) - \int \log|x - y| \mu(dx) \mu(dy) - \frac{3}{4} \\ & \leq \frac{1}{2} \int |A(x, y) - 1|^2 \mu(dx) \mu(dy) \times \log \left( 1 + \frac{\int (H\mu(x) - x)^2 \mu(dx)}{\int |A(x, y) - 1|^2 \mu(dx) \mu(dy)} \right). \end{aligned}$$

## 5. RATE OF CONVERGENCE IN THE FREE CLT

As in the classical case, the free HSI inequality can be used to give bounds on how fast the entropy converges along the free Central Limit Theorem. Let us first recall the statement of the free CLT:

**Theorem 5.1** (Free CLT, [34, 36]). *For each  $j = 1, \dots, N$  let  $(x_j^{(N)})_{N \in \mathbb{N}}$  be a sequence of freely independent, identically distributed random variables. Assume each  $x_j^{(N)}$  is centered and that the covariance of the  $n$ -tuples  $X^{(N)} := (x_1^{(N)}, \dots, x_n^{(N)})$  is the identity. If*

$$Y^{(N)} := \frac{X^{(1)} + \dots + X^{(N)}}{\sqrt{N}},$$

*then  $Y^{(N)} = (y_1^{(N)}, \dots, y_n^{(N)})$  converges (in moments) to a free  $(0, 1)$ -semicircular  $n$ -tuple.*

In [10], Chystiakov and Gotze showed that for a sequence of freely independent, identically distributed 1-tuples of random variables  $(x_1^{(N)})_{N \in \mathbb{N}}$ , the entropy of the renormalized convolution  $y_1^{(N)}$  satisfies

$$\chi^*(y_1^{(N)} | \sigma) = \chi^*(\sigma | \sigma) + O(N^{-1})$$

under the assumptions that the random variables  $x_1^{(N)}$  are bounded (actually, they show a more precise expansion, and identify the prefactor in the leading term of order  $N^{-1}$  in the expansion). The order of magnitude of the reminder is sharp. The proof is based on an Edgeworth expansion of the distribution function of normalized sums.

As in the classical case, we can use the HSI inequality to obtain a rate of convergence of order  $\frac{\log N}{N}$  for random variables with finite free Fisher information and free Stein discrepancy. While the rate is slightly

suboptimal, the technique has the advantage of avoiding technical computations, is easily applied to the multivariate case, and can be adapted to freely independent but not identically distributed  $X_i$ .

A quantitative multi-dimensional free CLT was obtained by Speicher in [35]. The estimates obtained there involve distances between Cauchy transforms, and are obtained via a reduction to a one-dimensional operator-valued free central limit theorem.

**Theorem 5.2.** *For each  $j = 1, \dots, N$  let  $(x_j^{(N)})_{N \in \mathbb{N}}$  be a sequence of freely independent, identically distributed random variables. Assume each  $x_j^{(N)}$  is centered and that the covariance of the  $n$ -tuples  $X^{(N)} := (x_1^{(N)}, \dots, x_n^{(N)})$  is the identity. For  $(a_1^{(N)}, \dots, a_N^{(N)})_{N \in \mathbb{N}}$  an array of real numbers satisfying  $\sum_{\ell=1}^N (a_\ell^{(N)})^2 = 1$ , let*

$$Y^{(N)} := a_1^{(N)} X^{(1)} + \dots + a_N^{(N)} X^{(N)}.$$

*Assume that the free Fisher information and the free Stein discrepancy of  $X^{(1)}$  relative to  $V_1$  are both finite. Let  $\sigma_N := \sum_{\ell=1}^N (a_\ell^{(N)})^4$ . Then*

$$|\chi^*(Y^{(N)}|V_1) - \chi^*(S|V_1)| = O(\sigma_N \log(\sigma_N^{-1})).$$

In particular, when  $a_\ell^{(N)} = N^{-1/2}$  for each  $\ell = 1, \dots, N$  we get a rate of order  $\frac{\log N}{N}$  in the free CLT (instead of the sharp  $\frac{1}{N}$  of [10]). So in exchange for allowing non-identical weights in the sum and covering the multivariate case, we end up with a slightly worse rate. This situation is the same as for the classical entropic CLT obtained via the HSI inequality of [23].

*Proof.* We first compute the free Stein kernel of  $Y^{(N)}$  in terms of the kernels for  $X^{(1)}, \dots, X^{(N)}$ , which we denote  $A_1, \dots, A_N$ . We may assume  $\|A_\ell - 1\|_{HS} = \Sigma^*(X^{(\ell)} | V_1) = \Sigma^*(X^{(1)} | V_1)$  for each  $\ell = 1, \dots, N$  (the latter equality following by the identical distributions of  $X^{(1)}, \dots, X^{(N)}$ ). Since the  $X^{(\ell)}$  are free, it follows that for  $P \in \mathcal{P}^n$

$$\begin{aligned} \langle Y^{(N)}, P(Y^{(N)}) \rangle_2 &= \sum_{\ell=1}^N a_\ell^{(N)} \langle X^{(\ell)}, P(Y^{(N)}) \rangle_2 \\ &= \sum_{\ell=1}^N a_\ell^{(N)} \langle A_\ell, a_\ell^{(N)} [\mathcal{J}P](Y^{(N)}) \rangle_{HS}. \end{aligned}$$

Thus if  $\mathcal{E}_N$  is the projection onto  $L^2(M_N(W^*(Y^{(N)}) \bar{\otimes} W^*(Y^{(N)})^{op}))$ , then

$$A := \mathcal{E}_N \left( \sum_{\ell=1}^N (a_\ell^{(N)})^2 A_\ell \right)$$

is a free Stein kernel for  $Y^{(N)}$  relative to  $V_1$ . Therefore

$$\begin{aligned} \Sigma^*(Y^{(N)} | V_1)^2 &\leq \|A - 1\|_{HS}^2 \\ &= \left\| \mathcal{E}_N \left( \sum_{\ell=1}^N (a_\ell^{(N)})^2 (A_\ell - 1) \right) \right\|_{HS}^2 \\ &= \sum_{\ell}^N (a_\ell^{(N)})^4 \|A_\ell - 1\|_{HS}^2 \\ &= \sigma_N \Sigma^*(X^{(1)} | V_1)^2, \end{aligned}$$

where the third line follows by freeness. Moreover, the Fisher information of  $Y^{(N)}$  is smaller than that of  $X^{(1)}$ , as a direct consequence of the free Stam inequality (see [39, Proposition 6.5]). The result then follows from a direct application of the HSI inequality to  $Y^{(N)}$  and monotonicity of  $t \mapsto t \log(1 + ct^{-1})$  for  $c, t \geq 0$ .  $\square$

## 6. BOUND ON THE DEFICIT IN THE LOG-SOBOLEV INEQUALITY

We produce the free analogue of [4, Theorem 1.1]. Let  $S = (s_1, \dots, s_n)$  and  $S' = (s'_1, \dots, s'_n)$  each be free  $(0, 1)$ -semicircular  $n$ -tuples in  $M^n$ , free from each other and free from  $x_1, \dots, x_n$ . Fix  $\rho > 0$  and let  $\frac{1}{\sqrt{\rho}}S = Y = (y_1, \dots, y_n)$ , a free  $(0, \rho^{-1})$ -semicircular  $n$ -tuple. For each  $t \geq 0$  we denote

$$\begin{aligned} X(t) &= (x_1(t), \dots, x_n(t)) = (x_1 + \sqrt{t}s'_1, \dots, x_n + \sqrt{t}s'_n) \\ Y(t) &= (y_1(t), \dots, y_n(t)) = (y_1 + \sqrt{t}s'_1, \dots, y_n + \sqrt{t}s'_n). \end{aligned}$$

Note that  $Y(t)$  is a free  $(0, \rho^{-1} + t)$ -semicircular  $n$ -tuple.

If  $Z$  is a free  $(0, \rho^{-1})$ -semicircular  $n$ -tuple, then denote

$$\Delta^*(X, Z) = \chi^*(X | V_\rho) - \chi^*(Z | V_\rho).$$

Also define  $d(t) = t - \log(1 + t)$  for  $t > -1$ .

**Theorem 6.1.**

$$\Phi^*(X | V_\rho) - 2\rho\Delta^*(X, \sqrt{\rho^{-1}}S) \geq n\rho \cdot d\left(\frac{\Phi^*(X)}{n\rho} - 1\right).$$

*Proof.* We compute using [39, Proposition 7.5]

$$\begin{aligned} \Delta^*(X(t), Y(t)) &= \chi^*(Y(t)) - \chi^*(X(t)) + \frac{1}{2(\rho^{-1} + t)} \sum_{j=1}^n \tau(x_j(t)^2 - y_j(t)^2) \\ &= \chi^*(Y(t)) - \chi^*(X(t)) + \frac{1}{2(\rho^{-1} + t)} \sum_{j=1}^n \tau(x_j^2 - y_j^2) \\ &= \chi^*(Y) - \chi^*(X) + \frac{1}{2} \int_0^t \Phi^*(Y(r)) - \Phi^*(X(r)) \, dr + \frac{1}{2(\rho^{-1} + t)} \sum_{j=1}^n \tau(x_j^2 - y_j^2) \\ &= \Delta^*(X, Y) + \frac{1}{2} \int_0^t \Phi^*(Y(r)) - \Phi^*(X(r)) \, dr - \frac{\rho t}{2(\rho^{-1} + t)} \sum_{j=1}^n \tau(x_j^2 - y_j^2). \end{aligned}$$

Equivalently,

$$\Delta^*(X, Y) = \Delta^*(X(t), Y(t)) + \frac{1}{2} \int_0^t \Phi^*(X(r)) - \Phi^*(Y(r)) \, dr + \frac{\rho t}{2(\rho^{-1} + t)} \sum_{j=1}^n \tau(x_j^2 - y_j^2).$$

From the free Stam inequality [39, Proposition 6.5] we have

$$\Phi^*(X(r)) \leq \frac{1}{\Phi^*(X)^{-1} + \Phi^*(\sqrt{r}S')^{-1}} = \frac{n\Phi^*(X)}{n + r\Phi^*(X)}.$$

Thus

$$\begin{aligned} \int_0^t \Phi^*(X(r)) - \Phi^*(Y(r)) \, dr &\leq \int_0^t \frac{n\Phi^*(X)}{n + r\Phi^*(X)} - \frac{n}{\rho^{-1} + r} \, dr \\ &= n \log\left(\frac{n + t\Phi^*(X)}{n(1 + \rho t)}\right). \end{aligned}$$

Continuing our previous computation we have

$$\Delta^*(X, Y) \leq \Delta^*(X(t), Y(t)) + \frac{n}{2} \log\left(\frac{n + t\Phi^*(X)}{n(1 + \rho t)}\right) + \frac{\rho t}{2(\rho^{-1} + t)} \sum_{j=1}^n \tau(x_j^2 - y_j^2). \quad (7)$$

We claim  $\lim_{t \rightarrow \infty} \Delta^*(X(t), Y(t)) = 0$ . By our previous computation, it suffices to show  $\lim_{t \rightarrow \infty} \chi^*(Y(t)) - \chi^*(X(t)) = 0$ . Note that  $\chi^*(Y(t)) = \frac{n}{2} \log(2\pi e(\rho^{-1} + t))$  and so by [39, Proposition 7.2] and [39, Proposition 7.5.(a)] we have

$$\frac{n}{2} \log\left(\frac{\rho^{-1} + t}{n^{-1}\text{Var}(X) + t}\right) \leq \chi^*(Y(t)) - \chi^*(X(t)) \leq \frac{n}{2} \log\left(\frac{\rho^{-1} + t}{t}\right),$$

which establishes the claim.



Hence, taking the limit as  $t \rightarrow \infty$  in (7) yields

$$\Delta^*(X, Y) \leq \frac{n}{2} \log \left( \frac{\Phi^*(X)}{n\rho} \right) + \frac{\rho}{2} \left[ \text{Var}(X) - \frac{n}{\rho} \right].$$

Replacing  $\rho \text{Var}(X) - n$  with  $\frac{1}{\rho} (\Phi^*(X | V_\rho) - \Phi^*(X) + n\rho)$  yields

$$\begin{aligned} \Delta^*(X, Y) &\leq \frac{n}{2} \log \left( \frac{\Phi^*(X)}{n\rho} \right) + \frac{1}{2\rho} \Phi^*(X | V_\rho) - \frac{1}{2\rho} \Phi^*(X) + \frac{n}{2} \\ &= \frac{1}{2\rho} \Phi^*(X | V_\rho) - \frac{n}{2} d \left( \frac{\Phi^*(X)}{n\rho} - 1 \right), \end{aligned}$$

which is equivalent to the claimed inequality.  $\square$

## 7. A MICROSTATES FREE LOG-SOBOLEV INEQUALITY

In this section, we prove a log-Sobolev inequality for the microstates free entropy. This result is simply a multi-variable analogue of [25, Theorem 4]. The proof of [25, Theorem 4] relies on the existence of transport between the measure of interest and the minimizer of the free entropy relative to a potential. Recall that by Lemma 1.6 the minimizer of the microstates free entropy relative to a potential  $V \in \mathcal{P}^{(R)}$  has as a joint law a free Gibbs state with potential  $V$ . Not only will we have to assume the existence of free transport in the multi-variable case, which is already a very restrictive assumption, but moreover we must assume that the free transport map is monotone and implemented by an invertible  $n$ -tuple of power series.

**Definition 7.1.** For  $f \in \mathcal{P}^{(R)}$ ,  $R > 0$ , we say that  $f$  is convex if  $\mathcal{J} \mathcal{D}f \geq 0$  as an element of  $M_n(\mathcal{P} \hat{\otimes}_R \mathcal{P}^{op})$ .

We have the following lemma, which is the analogue of the fact that a classical convex function lies above any of its tangent lines.

**Lemma 7.2.** Let  $(M, \tau)$  be a tracial von Neumann algebra and let  $f \in \mathcal{P}^{(R)}$ ,  $R > 0$ , be convex. If  $A, B \in M^n$  (self-adjoint) satisfy  $\|A\|, \|B\| \leq R$ , then

$$\tau(f(A)) \geq \tau(f(B) + [\mathcal{D}f](B) \cdot (A - B)).$$

*Proof.* The convexity of  $f$  implies  $[\mathcal{J} \mathcal{D}f](A, B)$  is positive in  $M_n(M \hat{\otimes} M^{op})$ . Hence

$$\begin{aligned} 0 &\leq \langle [\mathcal{J} \mathcal{D}f](A, B) \# (A - B), (A - B) \rangle_2 = \sum_{j,k=1}^n \tau((A_j - B_j) [\partial_k \mathcal{D}_j f](A, B) \# (A_k - B_k)) \\ &= \sum_{j=1}^n \tau((A_j - B_j) \{ [\mathcal{D}_j f](A) - [\mathcal{D}_j f](B) \}) = \tau([\mathcal{D}f](A) \cdot (A - B) - [\mathcal{D}f](B) \cdot (A - B)). \end{aligned}$$

Now, for  $t \in [0, 1]$  define  $A_t := tA + (1-t)B$  and

$$h(t) := \tau(f(A_t) - f(B) - [\mathcal{D}f](B) \cdot (A_t - B)).$$

Note that  $A_t - B = t(A - B)$ . We clearly have  $h(0) = 0$  and we wish to show  $h(1) \geq 0$ , which we shall demonstrate by showing  $h'(t) \geq 0$ . We compute

$$\begin{aligned} \frac{d}{dt} h(t) &= \tau \left( \sum_{j=1}^n [\partial_j f](A_t) \# (A - B) - [\mathcal{D}f](B) \cdot (A - B) \right) \\ &= \frac{1}{t} \tau([\mathcal{D}f](A_t) \cdot (A_t - B) - [\mathcal{D}f](B) \cdot (A_t - B)), \end{aligned}$$

which is positive by our initial observation above.  $\square$

We also have a much weaker converse of Lemma 1.6:

**Proposition 7.3.** Let  $Y = (y_1, \dots, y_n) \in M^n$  have as a joint law a free Gibbs state with potential  $V$  for some convex, self-adjoint  $V \in \mathcal{P}^{(R)}$ ,  $R > \|Y\|$ . Let  $F \in (\mathcal{P}^{(R)})^n$  be such that it has a compositional inverse  $G \in (\mathcal{P}^{(\|F\|_R)})^n$  and  $\mathcal{J}F \geq 0$ . Then  $\chi(F(Y) | V) \geq \chi(Y | V)$ .

*Proof.* Note that  $Y = G(F(Y))$ . We compute using Lemma 7.2 and [37, Proposition 3.5]

$$\begin{aligned} \chi(F(Y) | V) - \chi(Y | V) &= \tau(V(F(Y)) - V(Y)) - \chi(F(Y)) + \chi(Y) \\ &\geq \tau([\mathcal{D}V](Y) \cdot (F(Y) - Y)) + (\tau \otimes \tau^{op}) \circ \text{Tr}(\log |[\mathcal{J}G](F(Y))|). \end{aligned}$$

Now, since the joint law of  $Y$  is a free Gibbs state with potential  $V$ ,  $[\mathcal{D}_1V](Y), \dots, [\mathcal{D}_nV](Y)$  are the conjugate variables to  $Y$ . So using (2) we have

$$\chi(F(Y) | V) - \chi(Y | V) \geq (\tau \otimes \tau^{op}) \circ \text{Tr}([\mathcal{J}F](Y) - 1 + \log |[\mathcal{J}G](F(Y))|).$$

Then, since  $G \circ F$  is the identity, we have  $[\mathcal{J}G](F(Y)) = [(\mathcal{J}F)^{-1}](Y)$ . So continuing the above computation we have

$$\chi(F(Y) | V) - \chi(Y | V) \geq (\tau \otimes \tau^{op}) \circ \text{Tr}([\mathcal{J}F](Y) - 1 - \log[\mathcal{J}F](Y)) \geq 0,$$

since  $t - 1 - \log(t) \geq 0$  for  $t \geq 0$ . □

**Proposition 7.4.** *Suppose  $Y = (y_1, \dots, y_n) \in M^n$  has a free Gibbs state with potential  $V$  as a joint law, where  $V = V^* \in \mathcal{P}^{(R)}$ ,  $R > \|Y\|$ , is such that  $V - \rho V_1$  is convex for some  $\rho > 0$ . Suppose  $F \in (\mathcal{P}^{(R)})^n$  satisfies:*

- (i) *there exists  $G \in (\mathcal{P}^{(\|F\|_R)})^n$  a compositional inverse of  $F$ ;*
- (ii)  *$\mathcal{J}G \geq 0$ ; and*
- (iii)  *$F \in (\mathcal{P}^{(\|G\|_R)})^n$ .*

Then

$$\chi(F(Y) | V) - \chi(Y | V) \leq \frac{1}{2\rho} \Phi^*(F(Y) | V)$$

*Proof.* Denote  $F(Y) = X = (x_1, \dots, x_n)$ . If  $\Phi^*(X) = \infty$ , then there is nothing to show. Assume  $\Phi^*(X) < \infty$  and that  $\xi_1, \dots, \xi_n$  are the conjugate variables to  $x_1, \dots, x_n$ . Denote  $\Xi = (\xi_1, \dots, \xi_n)$ . Observe that the claimed inequality is equivalent to:

$$\begin{aligned} \frac{1}{2\rho} \|\Xi - [\mathcal{D}V](X)\|_2^2 + \tau(V(Y) - V(X) - [\mathcal{D}V](X) \cdot (Y - X)) \\ - \langle \Xi - [\mathcal{D}V](X), Y - X \rangle_2 + \langle \Xi, Y - X \rangle_2 + \chi(X) - \chi(Y) \geq 0. \end{aligned}$$

Let  $W = V - \rho V_1$ , so that by Lemma 7.2 we have  $\tau(W(Y) - W(X) - [\mathcal{D}W](X) \cdot (Y - X)) \geq 0$  or

$$\begin{aligned} \tau(V(Y) - V(X) - [\mathcal{D}V](X) \cdot (Y - X)) &\geq \rho \tau(V_1(Y) - V_1(X) - X \cdot (Y - X)) \\ &= \frac{\rho}{2} \tau(Y \cdot Y - X \cdot X - 2X \cdot (Y - X)) \\ &= \frac{\rho}{2} \|Y - X\|_2^2 \end{aligned}$$

Noting that

$$\frac{1}{2\rho} \|\Xi - [\mathcal{D}V](X)\|_2^2 + \frac{\rho}{2} \|Y - X\|_2^2 - \langle \Xi - [\mathcal{D}V](X), Y - X \rangle_2 = \frac{1}{2\rho} \|\Xi - [\mathcal{D}V](X) - \rho(Y - X)\|_2^2,$$

we see that it suffices to show  $\langle \Xi, Y - X \rangle_2 + \chi(X) - \chi(Y) \geq 0$ . Conditions (i)-(iii) imply that the full hypotheses of [37, Proposition 3.5] are satisfied. Using this and (2) (on  $Y = G(X)$ ) we have

$$\langle \Xi, Y - X \rangle_2 + \chi^*(X) - \chi^*(Y) = (\tau \otimes \tau^{op}) \circ \text{Tr}([\mathcal{J}G](X) - 1 - \log[\mathcal{J}G](X)),$$

which is positive since  $t - 1 - \log(t) \geq 0$  for  $t \geq 0$ . □

## REFERENCES

- [1] Cécile Ané, Sébastien Blachère, Djali Chafaï, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, and Grégory Scheffer, *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses [Panoramas and Syntheses], vol. 10, Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux. MR1845806
- [2] D. Bakry and Michel Émery, *Diffusions hypercontractives*, Séminaire de probabilités, XIX, 1983/84, 1985, pp. 177–206. MR889476
- [3] Philippe Biane and Roland Speicher, *Free diffusions, free entropy and free Fisher information*, Ann. Inst. H. Poincaré Probab. Statist. **37** (2001), no. 5, 581–606. MR1851716 (2002k:46167)
- [4] S. G. Bobkov, N. Gozlan, C. Roberto, and P.-M. Samson, *Bounds on the deficit in the logarithmic Sobolev inequality*, J. Funct. Anal. **267** (2014), no. 11, 4110–4138. MR3269872

- [5] François Bolley, Ivan Gentil, and Arnaud Guillin, *Dimensional improvements of the logarithmic Sobolev, Talagrand and Brascamp-Lieb inequalities*, ArXiv e-prints (July 2015), available at [1507.01086](#).
- [6] Marek Bożejko and Roland Speicher, *An example of a generalized Brownian motion*, *Comm. Math. Phys.* **137** (1991), no. 3, 519–531. MR1105428 (92m:46096)
- [7] ———, *Completely positive maps on Coxeter groups, deformed commutation relations, and operator spaces*, *Math. Ann.* **300** (1994), no. 1, 97–120. MR1289833
- [8] E. Carlen, *Some integral identities and inequalities for entire functions and their application to the coherent state transform*, *J. Funct. Anal.* **97** (1991), 231–249.
- [9] ———, *Superadditivity of Fisher’s information and logarithmic sobolev inequalities*, *J. Funct. Anal.* **101** (1991), 194–211.
- [10] G. P. Chistyakov and F. Gotze, *Asymptotic expansions in the clt in free probability*, *Probab. Theory Related Fields* **157** (2013), 107–156.
- [11] A. Cianchi, N. Fusco, F. Maggi, and A. Pratelli, *The sharp Sobolev inequality in quantitative form*, *J. Eur. Math. Soc.* **13** (2009), 1105–1139.
- [12] Dario Cordero-Erausquin, *Some applications of mass transport to Gaussian-type inequalities*, *Arch. Ration. Mech. Anal.* **161** (2002), no. 3, 257–269. MR1894593
- [13] Yoann Dabrowski, *A free stochastic partial differential equation*, *Ann. Inst. Henri Poincaré Probab. Stat.* **50** (2014), no. 4, 1404–1455. MR3270000
- [14] Ronen Eldan and Boaz Klartag, *Dimensionality and the stability of the Brunn-Minkowski inequality*, to appear in *Annali della Scuola Normale Superiore di Pisa* (2014).
- [15] Max Fathi, Emanuel Indrei, and Michel Ledoux, *Quantitative logarithmic Sobolev inequalities and stability estimates*, to appear in *Disc. Cont. Dynam. Syst. Series A* (2015).
- [16] A. Figalli, F. Maggi, and A. Pratelli, *A mass transportation approach to quantitative isoperimetric inequalities*, *Invent. Math.* **182** (2010), no. 1, 167–211. MR2672283
- [17] L. Gross, *Logarithmic sobolev inequalities*, *Amer. Math. J.* **97** (1975), 1061–1083.
- [18] A. Guionnet and D. Shlyakhtenko, *Free monotone transport*, *Invent. Math.* **197** (2014), no. 3, 613–661. MR3251831
- [19] F. Hiai, D. Petz, and Y. Ueda, *Free transportation cost inequalities via random matrix approximation*, *Probab. Theory Relat. Fields* **130** (2004), 199–221.
- [20] Fumio Hiai, *q-deformed Araki-Woods algebras*, *Operator algebras and mathematical physics (Constanța, 2001)*, 2003, pp. 169–202. MR2018229 (2004j:46086)
- [21] Emanuel Indrei and Diego Marcon, *A quantitative log-Sobolev inequality for a two parameter family of functions*, *Int. Math. Res. Not. IMRN* **20** (2014), 5563–5580. MR3271181
- [22] Michel Ledoux, *A (one-dimensional) free Brunn-Minkowski inequality*, *C. R. Math. Acad. Sci. Paris* **340** (2005), no. 4, 301–304. MR2121895
- [23] Michel Ledoux, Ivan Nourdin, and Giovanni Peccati, *Stein’s method, logarithmic Sobolev and transport inequalities*, *Geom. Funct. Anal.* **25** (2015), no. 1, 256–306. MR3320893
- [24] ———, *A Stein deficit for the logarithmic Sobolev inequality*, ArXiv e-prints (February 2016), available at [1602.08235](#).
- [25] Michel Ledoux and Ionel Popescu, *Mass transportation proofs of free functional inequalities, and free Poincaré inequalities*, *J. Funct. Anal.* **257** (2009), no. 4, 1175–1221. MR2535467
- [26] T. Mai, R. Speicher, and M. Weber, *Absence of algebraic relations and of zero divisors under the assumption of full non-microstates free entropy dimension*, ArXiv e-prints (February 2015), available at [1502.06357](#).
- [27] B. Nelson and Q. Zeng, *An application of free transport to mixed q-Gaussian algebras*, ArXiv e-prints (July 2015), available at [1507.04824](#).
- [28] ———, *Free monotone transport for infinite variables*, ArXiv e-prints (November 2015), available at [1511.06291](#).
- [29] Brent Nelson, *Free Monotone Transport Without a Trace*, *Comm. Math. Phys.* **334** (2015), no. 3, 1245–1298. MR3312436
- [30] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, *J. Funct. Anal.* **173** (2000), no. 2, 361–400. MR1760620
- [31] Dimitri Shlyakhtenko, *Free quasi-free states*, *Pacific J. Math.* **177** (1997), no. 2, 329–368. MR1444786 (98b:46086)
- [32] ———, *Some estimates for non-microstates free entropy dimension with applications to q-semicircular families*, *Int. Math. Res. Not.* **51** (2004), 2757–2772. MR2130608
- [33] ———, *Lower estimates on microstates free entropy dimension*, *Anal. PDE* **2** (2009), no. 2, 119–146. MR2547131 (2012a:46129)
- [34] Roland Speicher, *A new example of independence and white noise*, *Probab. Theory. Rlat. Fields* **84** (1990), 141–159.
- [35] ———, *On the rate of convergence and Berry-Esseen type theorems for a multivariate free central limit theorem*, ArXiv e-prints (December 2007), available at [0712.2974](#).
- [36] Dan Voiculescu, *Addition of certain non-commuting random variables*, *J. Funct. Anal.* **66** (1986), 323–346.
- [37] ———, *The analogues of entropy and of Fisher’s information measure in free probability theory. II*, *Invent. Math.* **118** (1994), no. 3, 411–440. MR1296352 (96a:46117)
- [38] ———, *The analogues of entropy and of Fisher’s information measure in free probability theory. IV. Maximum entropy and freeness*, *Free probability theory (Waterloo, ON, 1995)*, 1997, pp. 293–302. MR1426847 (97j:46071)
- [39] ———, *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. MR1618636 (99d:46087)
- [40] ———, *Free entropy*, *Bull. London Math. Soc.* **34** (2002), no. 3, 257–278. MR1887698 (2003c:46077)

•DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, BERKELEY  
*E-mail address:* `maxf@berkeley.edu`

°DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, BERKELEY  
*E-mail address:* `brent@math.berkeley.edu`