ENTROPY CONSIDERATIONS IN THE NONCOMMUTATIVE SETTING

PATRICIA GIURGESCU

ABSTRACT. An analogue of the classical link between the relative entropy and Fisher information entropy is presented in the context of free probability theory. Several generalizations of the relative entropy in terms of density matrices are also discussed.

1. Introduction. It is known that the equilibrium states of Hamiltonian models of interacting particle systems correspond to states with extremal entropic properties. The transition to equilibrium occurs through the metastable states, related to probabilistic limit results, such as the law of large numbers, the central limit theorem, large deviations. The principle of maximum entropy is well known and widely used in the construction of mathematical models. According to it, under the uncertainty of the choice of a specific type of distribution to be used as a model of a physical phenomenon, maximum entropy distributions (and/or their mixtures) should be tested first. Conversely, if a distribution fits a real phenomenon well, its relationship with maximum entropy distributions should be investigated.

Recall that for a discrete random variable X taking values $x_1,\ldots,x_n$ with probabilities $p_1,\ldots,p_n$, Shannon's entropy is $H(X) = -\sum_{k=1}^{n} p_k \log p_k$. For an absolutely continuous random variable X with Lebesgue density $p(x)$, Shannon's entropy is given by $H(X) = -\mathbf{E}p(X)\log p(X)$.

The probability distributions corresponding to the extremal entropy cases are the building blocks of infinitely divisible distributions, as proved by Goldman [4]. Namely, the normal and exponential/Poisson distributions maximize Shannon's entropy when certain moments are prescribed.

Exponential distributions have maximum entropy among all distributions concentrated on the positive half-line and having finite expectations (if Y_1 is a random variable with exponential distribution of parameter $\lambda > 0$ ($\mathbf{E}Y_1 = \lambda$), then for all X random variables such that $P(X \leq 0) = 1$ and $\mathbf{E}X = \lambda$, one has $H(X) \leq H(Y_1)$).

Normal distributions have maximum entropy among all distributions with finite variances (if Y_2 is a normal random variable with parameters 0, σ^2, then for all X random variables with $\mathbf{E}X = 0$ and $\mathbf{E}X^2 = \sigma^2$, one has $H(X) \leq H(Y_2)$).

This paper is organized as follows. In Section 2, we present the classical link between the relative entropy and Fisher information measure developed by Barron [1]. The concept of free random variable is introduced in Section 3, by analogy with the

classical notion of independence. In Section 4, we present the free relative entropy and free Fisher information and provide a step-by-step derivation of a relationship between these two measures, in analogy to the classical case presented in Section 2. We conclude with a discussion and comparison of various candidates for the quantum analogue of Fisher information which is not uniquely defined in the recent literature.

2. The relative entropy and Fisher information measure. In considering central limit theorem type results, it is more convenient to introduce the *relative entropy* (Kullback-Leibler information "distance"), as remarked by Barron in [1].

DEFINITION 2.1. If X is a random variable with finite variance and density $p(x)$ (with respect to the Lebesgue measure), then the relative entropy of X with respect to the normal density g, with the same mean and variance as p, is

$$S(X) = \int p(x) \log\left[\frac{p(x)}{g(x)}\right] dx. \tag{2.1}$$

Obviously, the relative entropy is not symmetric with respect to p and g, so it is not a proper metric (distance). Nonetheless, it measures the deviation of p from g in terms of entropy. If $\mathbf{E}X = a$ and $\operatorname{var}X = \sigma^2$, then

$$
\begin{aligned}
S(X) &= \int p(x) \log p(x) dx - \int p(x) \log g(x) dx \\
&= -H(X) - \int p(x)\left[\log\frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(x-a)^2}{2\sigma^2}\right] dx \\
&= \frac{1}{2}\left[\log\left(2\pi\sigma^2\right) + 1\right] - H(X).
\end{aligned}
\tag{2.2}
$$

Hence the relative entropy is the difference between Shannon's entropy of the normal law and that of X. The concavity of the logarithm implies that S is nonnegative and equals zero only if $p(x) = g(x)$ almost everywhere (i.e., the normal law has maximum Shannon's entropy for a given variance).

DEFINITION 2.2. For a random variable X, with continuously differentiable density $f(x)$ and $\mathbf{E}X = a$, $\operatorname{var}X = \sigma^2 < \infty$, the standardized Fisher information measure is defined as

$$\Phi(X) = \sigma^2 \mathbf{E}\left[\frac{f'(X)}{f(X)} - \frac{g'(X)}{g(X)}\right]^2, \tag{2.3}$$

where g is the normal density with the same mean and variance as f.

The normal density has minimum Fisher information measure for a given variance (since $\Phi \geq 0$, with equality only if $f = g$).

PROPOSITION 2.3 (see [1]). *The relative entropy $S(X)$ of any random variable X with finite variance is the integral of Fisher's standardized information, with respect to an independent normal random variable Y having the same mean and variance as X,*

namely

$$S(X) = \frac{1}{2} \int_0^1 \frac{1}{t} \Phi\left(\sqrt{t}X + \sqrt{1-t}Y\right) dt. \tag{2.4}$$

In what follows we derive an analogue of the above relationship in the noncommutative setting of free probability.

3. Freeness. We investigate results describing the entropy and the asymptotic behavior of sequences of observables (physical quantities), treated as elements of a von Neumann algebra $\mathcal{A}$, acting on a separable complex Hilbert space, and having a faithful normal state φ. The states are represented by the linear functionals φ on $\mathcal{A}$, satisfying the conditions of positivity ($\varphi(A^*A) \geq 0$, $A \in \mathcal{A}$) and normalization ($\varphi(I) = 1$); $\varphi(A)$ is then interpreted as the expected value of the observable A for the state φ. The states of a given system form a convex set, with the pure states being its extremal elements and the mixed states are represented by positive trace class operators (density matrices). The dynamical properties are reflected in the spectrum of the density evolution operators (density matrices). For instance, dynamical systems with continuous spectrum evolve in a mixing manner, as first noted by von Neumann and Koopman [6].

The classical notion of independence allows the calculation of mixed moments in terms of single moments. Consider two random variables X and Y, living on some probability space $(\Omega, \mathfrak{U}, P)$, assumed to be bounded (so that all their moments exist and determine their respective distributions). The expectation $\varphi(f(X,Y)) = \int f(X(\omega), Y(\omega)) dP(\omega)$ is well defined for any bounded function f of two variables. Then X is independent of Y if

$$\varphi\left(X^{n_1} Y^{m_1} \cdots X^{n_k} Y^{m_k}\right) = \varphi\left(X^{n_1 + \cdots + n_k}\right) \varphi\left(Y^{m_1 + \cdots + m_k}\right). \tag{3.1}$$

In particular, $\varphi(XY) = \varphi(X)\varphi(Y)$ and $\varphi(XXYY) = \varphi(XYXY) = \varphi(XX)\varphi(YY)$.

DEFINITION 3.1. A noncommutative probability space $(\mathcal{A}, \varphi)$ consists of a unital algebra $\mathcal{A}$ and a unital linear functional (analogue of the classical expectation map)

$$\varphi : \mathcal{A} \to \mathbb{C}, \qquad \varphi(I) = 1. \tag{3.2}$$

A random variable X is an element of the given algebra $\mathcal{A}$. The random variables $X_1, X_2, \ldots \in \mathcal{A}$ are free (with respect to φ) if, for all $m \in \mathbb{N}$ and for all polynomials $p_1(X), p_2(X), \ldots, p_m(X)$ in the variable X we have:

$$\varphi\left[p_1(X_{r(1)}) p_2(X_{r(2)}) \cdots p_m(X_{r(m)})\right] = 0 \tag{3.3}$$

whenever $\varphi[p_k(X_{r(k)})] = 0$, for all $k = 1, \ldots, m$ and $r(k) \neq r(k+1)$, for all $k = 1, \ldots, m-1$ (consecutive indices are distinct).

This definition provides a rule for calculating mixed moments in the free variables in terms of moments of the single variables.

Consider, for instance, the noncentered polynomial-like variables $p_1(X) = X - \varphi(X)1$ and $p_2(Y) = Y - \varphi(Y)1$. The definition of freeness yields

$$0 = \varphi[p_1(X)p_2(Y)] = \varphi[(X - \varphi(X)1) \cdot (Y - \varphi(Y)1)] = \varphi(XY) - \varphi(X)\varphi(Y), \tag{3.4}$$

as in the independent case. Similarly, $\varphi(XXYY) = \varphi(XX)\varphi(YY)$ follows from $\varphi[(X^2 - \varphi(X^2)1) \cdot (Y^2 - \varphi(Y^2)1)] = 0$. But considering the mixed moment $\varphi(XYXY)$, which can be resolved from $\varphi[(X - \varphi(X)1) \cdot (Y - \varphi(Y)1) \cdot (X - \varphi(X)1) \cdot (Y - \varphi(Y)1)] = 0$, we obtain $\varphi(XYXY) = \varphi(XX)\varphi(Y)\varphi(Y) + \varphi(X)\varphi(X)\varphi(YY) - \varphi(X)\varphi(Y)\varphi(X)\varphi(Y)$.

Hence $\varphi(XXYY) \neq \varphi(XYXY)$, illustrating the fact that freeness is the noncommutative analogue of independence.

Since for the free random variables X and Y the moments of $X + Y$ can be expressed in terms of the moments of X and the moments of Y, the distribution μ_{X+Y} depends on the distributions μ_X and μ_Y, through a special type of convolution introduced by Voiculescu in [12], called the *free additive convolution* and denoted $\boxplus$,

$$\mu_{X+Y} = \mu_X \boxplus \mu_Y. \tag{3.5}$$

The analogue of the logarithm of the Fourier transform used to linearize the convolution in the classical case $(\ln \mathcal{F}_{\mu * \nu}(t) = \ln \mathcal{F}_\mu(t) + \ln \mathcal{F}_\nu(t)$, with $\mathcal{F}_\mu(t) = \int_{\mathbb{R}} e^{itx} d\mu(x) = \varphi[e^{itX}])$ is the R-transform in the free setting. The R-transform $R_\mu(z)$, corresponding to each probability measure μ, is an analytic function on the upper half-plane, viewed as a formal power series in the indeterminate z. The desired feature of the R-transform is to linearize free additive convolution:

$$R_{\mu \boxplus \nu}(z) = R_\mu(z) + R_\nu(z). \tag{3.6}$$

The R-transform is uniquely determined by the Cauchy-transform G_μ of μ, defined by the analytic function

$$G_\mu(z) = \int \frac{1}{z - x} d\mu(x) = \varphi\left[\frac{1}{z - X}\right], \quad \text{on } \mathbb{C} \setminus \mathbb{R}, \tag{3.7}$$

such that $G_\mu(\bar{z}) = \overline{G_\mu(z)}$. Its Laurent series expansion, involving the moments of μ, is

$$G_\mu(z) = \sum_{n=0}^{\infty} z^{-n-1} \int_{\mathbb{R}} x^n d\mu(x). \tag{3.8}$$

This series can be formally inverted with respect to z:

$$K_\mu(z) = \frac{1}{z} + \sum_{n=1}^{\infty} C_n z^{n-1}. \tag{3.9}$$

The coefficients C_n (called free cumulants) are polynomial functions of the moments $m_k = \int_{\mathbb{R}} x^k d\mu(x)$ of μ. The link between the R-transform and the Cauchy-transform is that $G_\mu(z)$ and $R_\mu(z) + 1/z$ are inverses of each other with respect to composition: $G_\mu[R_\mu(z) + 1/z] = z$.

For example, if $\mu = (1/2)(\delta_0 + \delta_1)$, then $G_\mu(z) = (z - 1/2)/(z(z - 1))$ and $K_\mu(z) = (z + 1 + \sqrt{1 + z^2})/(2z)$. Hence, $K_{\mu \boxplus \mu}(z) = 1 + (1/z)\sqrt{1 + z^2}$ and $G_{\mu \boxplus \mu}(z) = 1/\sqrt{z(z - 1)}$. The inverted Cauchy transform, $d[\mu \boxplus \mu](x) = 1/(\pi\sqrt{x(2 - x)})dx$ on $[0, 2]$, is the arcsine distribution.

DEFINITION 3.2. An element $A = A^* \in \mathcal{A}$ is a *semicircular element* of radius $\sigma > 0$ if its distribution is absolutely continuous with respect to the Lebesgue measure and its density is $\rho(t) = 2(\pi\sigma^2)^{-1}\sqrt{\sigma^2 - t^2}$ on $[-\sigma, \sigma]$. (σ^2 is the free analogue of the variance parameter of a classical normal law.)

4. Free entropy. The free analogues of Shannon's entropy and Fisher information measure were first considered by Voiculescu in [13, 14], and further investigated in [8].

For a von Neumann algebra $\mathcal{A}$ and a faithful normal trace state $\varphi : \mathcal{A} \to \mathbb{C}$, consider the tracial W^*-probability space $(\mathcal{A}, \varphi)$. The distribution of any $X \in \mathcal{A}$, is defined by the unique probability measure μ, compactly supported on $\mathbb{R}$, having $\int_{-\infty}^{\infty} t^n d\mu(t) = \varphi(X^n)$, for all $n \geq 0$.

The Hilbert space $L^2(\mathcal{A}, \varphi)$ is obtained completing $\mathcal{A}$ with respect to the norm $\|X\|_2 = \sqrt{\varphi(X^*X)}$, $X \in \mathcal{A}$. For a positive integer N, denote by $\mathcal{M}_N(\mathcal{A})$ the W^*-algebra of the $N \times N$ matrices over $\mathcal{A}$, and $\varphi_N = \mathrm{Tr} \otimes \varphi : \mathcal{M}_N(\mathcal{A}) \to \mathbb{C}$ its faithful trace-state (i.e., $\varphi_N(X) = (1/N) \sum_{i=1}^{N} \varphi(a_{ii})$, for $X = (a_{ij})_{1 \leq i, j \leq N} \in \mathcal{M}_N(\mathcal{A})$). Since $\|X\|_{L^2(\varphi_N)}^2 = (1/N) \sum_{i, j=1}^{N} \|a_{ij}\|_{L^2(\varphi)}^2$, for any $X \in \mathcal{M}_N(\mathcal{A})$, one can extend the previous norm by continuity to the Hilbert space $L^2(\mathcal{M}_N(\mathcal{A}), \varphi_N)$.

For an n-tuple of selfadjoint elements $X_j \in \mathcal{A}^{\mathrm{sa}}$, the *free entropy*, denoted $\chi(X_1, \ldots, X_n)$, involves sets of points in $(\mathcal{M}_N(\mathcal{A}))^n$, that is, finite-dimensional matrices, so-called matrix microstates. The free entropy is then defined as the normalized limit of the logarithms of volumes of all such possible microstates. For one random variable $X \in \mathcal{A}^{\mathrm{sa}}$ with the distribution μ, the free entropy was defined in [13] as:

$$\chi(X) = \iint \log|s - t|\, d\mu(s) d\mu(t) + \frac{3}{4} + \frac{\log(2\pi)}{2}. \tag{4.1}$$

PROPOSITION 4.1. *Let $\Omega = \{\mathbf{X} = (X_1, \ldots, X_n) \in \mathcal{M}_N(\mathcal{A})^n \mid X_j = X_j^*, \|X_j\|_{L^2(\varphi_N)}^2 = 1\}$ and let $\mathbf{S} = (S_1, \ldots, S_n) \in \Omega$ be a semicircular system (i.e., $\{S_j\}$ are free and each S_j has a $(0,1)$ semicircular distribution). Then $\chi(\mathbf{X}) \leq \chi(\mathbf{S})$, for all $\mathbf{X} \in \Omega$.*

PROOF. Since $3/4 + \iint \log|s - t|\, d\mu(s) d\mu(t) = 1/2$, for μ with density $1/(2\pi)\sqrt{4 - t^2}$, concentrated on $[-2, 2]$, we have

$$\chi(\mathbf{S}) = n\chi(S_1) = n\left[\iint \log|s - t|\, d\mu(s) d\mu(t) + \frac{3}{4} + \frac{\log(2\pi)}{2} \right] = \frac{n}{2} \log(2\pi e). \tag{4.2}$$

Denote by $\varphi_N(X_j^2)$ the norm $\|X_j\|_{L^2(\varphi_N)}^2$. Then $\varphi_N(X_1^2 + \cdots + X_n^2) = n$ implies $\chi(\mathbf{X}) \leq (n/2) \log(2\pi e) = \chi(\mathbf{S})$. ∎

The extension of the above result to matrices is the following.

PROPOSITION 4.2. *Let $C = (c_{ij})_{1 \leq i, j \leq N}$ be a real positive definite invertible matrix, and $\Omega(C) = \{\mathbf{X} = (X_1, \ldots, X_N) \in \Omega \mid \varphi_N(X_i X_j) = c_{ij}, 1 \leq i, j \leq N\}$. Let $B = (b_{ij})_{1 \leq i, j \leq N} = C^{1/2}$. Then $\chi(\mathbf{X}) \leq \log(\det B) + \chi(\mathbf{S})$, for all $\mathbf{X} \in \Omega(C)$, where S is a free family of $(0,1)$ semicircular elements. Equality is attained if and only if $(X_1, \ldots, X_N)$ and $(\sum_j b_{1j} S_j, \ldots, \sum_j b_{nj} S_j)$ have the same distribution.*

PROOF. Apply the previous result to $\mathbf{Y} = (Y_1, \ldots, Y_N)$, with $Y_i = \sum_j a_{ij} X_j$ and $(a_{ij})_{1 \leq i, j \leq N} = A = B^{-1}$. Namely, $\chi(\mathbf{X}) = \log(\det B) + \chi(\mathbf{Y})$ and $\chi(\mathbf{Y}) \leq \chi(\mathbf{S})$, for $\mathbf{X} \in \Omega(C)$ and $\mathbf{Y} \in \Omega(I) = \Omega$, respectively. ∎

DEFINITION 4.3. $\{A_j\}_{j \in J} \subset \mathcal{A}$ is a *selfadjoint family* if there is an involutive bijection $\sigma : J \to J$ such that $A_j^* = A_{\sigma(j)}$, for all $j \in J$. The family of vectors $\{\xi_j\}_{j \in J} \subset$

$L^2(\mathcal{A}, \varphi)$ is said to fulfill the *conjugate relations* for $\{A_j\}_{j \in J}$, with respect to the W^*-subalgebra $\mathcal{B} \subseteq \mathcal{A}$, if for all $n \geq 0$, $B_0, \ldots, B_n \in \mathcal{B}$, $j, j_1, \ldots, j_n \in I$,

$$\varphi(\xi_j B_0 A_{j_1} B_1 \cdots A_{j_n} B_n)$$
$$= \sum_{m=1}^{n} \delta_{j,j_m} \varphi(B_0 A_{j_1} \cdots A_{j_{m-1}} B_{m-1}) \cdot \varphi(B_m A_{j_{m+1}} \cdots A_{j_n} B_n). \tag{4.3}$$

As usual, the conjugate relations allow the introduction of an inner product in $L^2(\mathcal{A}, \varphi)$.

DEFINITION 4.4. Let $(\mathcal{A}, \varphi)$ be a faithful trace W^*-probability space. If $\{A_j\}_{j \in J} \subset \mathcal{A}$ is a selfadjoint family, $\mathcal{B} \subseteq \mathcal{A}$ is a unital W^*-subalgebra such that $\{A_j\}_{j \in J}$ has a conjugate system $\{\xi_j\}_{j \in J}$ relative to $\mathcal{B}$, then the *free Fisher information* of $\{A_j\}_{j \in J}$ with respect to $\mathcal{B}$ is

$$\Phi^*(\{A_j\}_{j \in J} \mid \mathcal{B}) = \sum_{j \in J} \|\xi_j\|^2. \tag{4.4}$$

If $\{A_j\}_{j \in J}$ does not have a conjugate system relative to $\mathcal{B}$, then $\Phi^*(\{A_j\}_{j \in J} \mid \mathcal{B}) \overset{\text{def}}{=} \infty$ (see [8]).

An analogue of the classical *Cramér-Rao inequality* concerning the free Fisher information was formulated in [15].

If $(X_1, \ldots, X_n)$ *is a selfadjoint family of elements of* $L^2(\mathcal{M}_N(\mathcal{A}), \varphi_N)$, *such that the total variance* $\varphi_N(X_1^2 + \cdots + X_n^2)$ *is prescribed, then the free Fisher information* $\Phi^*(X_1, \ldots, X_n)$ *is minimized when the* X_j's *are free semicircular elements of equal radii.*

The link between the free entropy and the free Fisher information (similar to relation (2.4) for the classical case) is presented in the stages below.

(1) For the one-dimensional case, let $c \in \mathcal{A}$ be a *circular* variable (i.e., the real and imaginary parts of c are free and have semicircular distributions of equal radii). Assume further that the variance $\varphi_N(c) = 1$ and $\{c, c^*\}$ is free from $\{a, a^*\}$. Then,

$$\chi(\{a, a^*\}) = \frac{1}{2} \int_0^\infty \left[\frac{2}{1+t} - \Phi^*\left(a + c\sqrt{t}, (a + c\sqrt{t})^*\right) \right] dt + \log(2\pi e). \tag{4.5}$$

(2) Consider the matrix $A + S\sqrt{t} = \begin{bmatrix} 0 & a+c\sqrt{t} \\ (a+c\sqrt{t})^* & 0 \end{bmatrix}$, with $A, S \in \mathcal{M}_2(\mathcal{A})^{\text{sa}}$, S semicircular of variance 1 ($\varphi_2(S^{2n}) = \varphi[(c^*c)^n]$, $\varphi_2(S^{2n+1}) = 0$, for all $n \geq 0$) and free from A. Then,

$$\chi(A) = \frac{1}{2} \int_0^\infty \left[\frac{1}{1+t} - \Phi^*(A + S\sqrt{t}) \right] dt + \frac{1}{2} \log(2\pi e). \tag{4.6}$$

(3) The multidimensional version of (4.5) is

$$\chi\left(\{a_i, a_i^*\}_{i=1}^m \cup \{b_j\}_{j=1}^n\right)$$
$$= \frac{2m+n}{2} \log(2\pi e)$$
$$+ \frac{1}{2} \int_0^\infty \left[\frac{2m+n}{1+t} - \Phi^*\left(\{a_i + c_i\sqrt{t}, a_i^* + c_i^*\sqrt{t}\}_{i=1}^m \cup \{b_j + s_j\sqrt{t}\}_{j=1}^n\right) \right] dt, \tag{4.7}$$

where $s_j \in \mathscr{A}$ are semicircular and $c_i \in \mathscr{A}$ are circular elements, normalized by their variances ($\varphi_N(c_i^* c_i) = 1 = \varphi_N(s_j^2)$, for all $1 \le i \le m$, $1 \le j \le n$) such that $\{\{c_i, c_i^*\}_{i=1}^m$, $\{s_j\}_{j=1}^n\}$ and $\{\{a_i, a_i^*\}_{i=1}^m, \{b_j\}_{j=1}^n\}$ are free.

(4) Taking $\{\{b_{ij}, b_{ij}^*\}_{i,j=1}^N \cup \{b_{ii}\}_{i=1}^N\}$ instead of $\{\{a_i, a_i^*\}_{i=1}^m \cup \{b_j\}_{j=1}^n\}$ (i.e., when $n = N$ and $m = N(N-1)/2$), using relations (4.7) and (4.6), yields the matrix version:

$$\chi\Big[B = (b_{ij})_{i,j=1}^N \Big] = \frac{1}{2} \int_0^\infty \left[\frac{N^2}{1+t} - \Phi^*\Big((b_{ij} + s_{ij}\sqrt{t})_{i,j=1}^N \Big) \right] dt + \frac{N^2}{2} \log(2\pi e), \quad (4.8)$$

where the selfadjoint family $(b_{ij})_{i,j=1}^N$ is viewed as $(\{b_{ij}, b_{ij}^*\}_{i,j=1}^N \cup \{b_{ii}\}_{i=1}^N)$, and the family $(s_{ij})_{i,j=1}^N \subset \mathscr{A}$ has the properties:

- The elements s_{ii} are semicircular of variance 1, for all $1 \le i \le N$.
- $s_{ij}^* = s_{ji}$, for all $1 \le i, j \le N$, and the elements s_{ij} are circular of variance equal to 1, for all $1 \le i < j \le N$.
- The sets $\{s_{11}\}, \ldots, \{s_{NN}\}, \{s_{12}, s_{12}^*\}, \ldots, \{s_{N-1,N}, s_{N-1,N}^*\}, \{b_{ij}\}_{i,j=1}^N$ are free.

(5) Denoting $S = (s_{ij})_{i,j=1}^N \in \mathscr{M}_N(\mathscr{A})$, the matrix $(1/\sqrt{N})S$ is semicircular of variance 1 and free from B in $(\mathscr{M}_N(\mathscr{A}), \varphi_N)$ and relation (4.8) becomes

$$\chi\left(\frac{1}{\sqrt{N}} B \right) = \frac{1}{2} \int_0^\infty \left[\frac{1}{1+t} - \Phi^*\left(\frac{1}{\sqrt{N}} B + \sqrt{\frac{t}{N}} S \right) \right] dt + \frac{1}{2} \log(2\pi e). \quad (4.9)$$

(6) Using the scaling formulas: $\Phi^*(\kappa X) = \kappa^{-2}\Phi^*(X)$ and $\chi(\kappa X) = \chi(X) + \log \kappa$, for $\kappa > 0$ and X a selfadjoint random variable, then (4.9) leads to

$$\chi(B) - \frac{\log N}{2} = \frac{1}{2} \int_0^\infty \left[\frac{1}{1+t} - N\Phi^*(B + \sqrt{t}S) \right] dt + \frac{1}{2} \log(2\pi e), \quad (4.10)$$

which is the free setting analogue of relations (2.1) and (2.4).

Recently, Nica et al. [8] provided free entropy minimization criteria not only when the expectation or variance were prescribed, but also when the moments of all orders, that is, the whole distribution of an element A^*A, is specified.

5. Entropy considerations in quantum setting.

Since the quantum analogue of the classical Fisher information is not uniquely defined in the recent literature, it is useful to study and compare various candidates. The relative entropy has several generalizations in the context of density matrices describing the mixed quantum state space of a system. For two density matrices P and Q, with positive eigenvalues, the relative entropy would be $S(P, Q) = \operatorname{Tr} P \log(P/Q)$. To give meaning to the quotient P/Q, one can consider Umegaki's relative entropy [11],

$$S_U(P, Q) = \operatorname{Tr} P(\log P - \log Q) \quad (5.1)$$

or the Belavkin and Staszewski relative entropy [2],

$$S_{BS}(P, Q) = \operatorname{Tr} P \log(P^{1/2} Q^{-1} P^{1/2}) \quad (5.2)$$

(note that S_{BS} and S_U coincide in the case of commuting density matrices).

If $\mathscr{A}$ represents a finite quantum system, then the family of states φ corresponding to the density matrices P_φ in the Gelfand Naimark Segal (GNS) space $\mathscr{H}$ of φ, consists of essentially localized modifications of φ and

$$\varphi(A) = \operatorname{Tr}(P_\varphi A) \quad \forall A \in \mathscr{A}. \tag{5.3}$$

When φ is faithful, the representation is unique and the normalization $\varphi(I) = 1$ implies $\operatorname{Tr} P_\varphi = 1$.

The *relative entropy of the state φ with respect to ψ* is naturally defined by

$$S(\varphi,\psi) = \begin{cases} \operatorname{Tr} P_\varphi(\log P_\varphi - \log Q_\psi), & \text{if } \operatorname{supp} P_\varphi \geq \operatorname{supp} Q_\psi, \\ +\infty, & \text{otherwise}, \end{cases} \tag{5.4}$$

where $\operatorname{supp} P_\varphi$ designates the smallest projection V such that $\varphi(V) = \varphi(I)$. In particular, $S(\varphi,\psi)$ is always finite if the density of ψ has strictly positive eigenvalues (i.e., ψ is faithful). The functional $(\varphi,\psi) \mapsto S(\varphi,\psi)$ is lower semicontinuous with respect to the weak* topology on the state space of the C^*-algebra $\mathscr{A}$. If $\mathscr{A}$ is finitely dimensional, the functional $(\varphi,\psi) \mapsto S(\varphi,\psi)$ is measurable [9].

PROPOSITION 5.1. *The relative entropy functional $S(\varphi,\psi)$ has the following properties:*

(a) *For any φ,ψ positive functionals on the finite quantum system $\mathscr{A}$,*

$$S(\varphi,\psi) \geq \frac{1}{2}\operatorname{Tr}(P_\varphi - Q_\psi)^2 + \psi(I) - \varphi(I) \geq 0, \tag{5.5}$$

$S(\varphi,\psi) = 0$ *occurs only when $\varphi = \psi$.*

(b) $S(\varphi,\psi)$ *is monotonous under unital Schwarz mappings $\alpha: \mathscr{A}_1 \to \mathscr{A}_2$, for any states φ,ψ of $\mathscr{A}_2$,*

$$S(\varphi \circ \alpha, \psi \circ \alpha) \leq S(\varphi,\psi). \tag{5.6}$$

(Note: α is a Schwarz mapping if it is linear and satisfies the inequality $\alpha(A^*A) \geq \alpha(A)^*\alpha(A)$, for all $A \in \mathscr{A}_1$.) In particular, if α is an automorphism of $\mathscr{A}$, we obtain the invariance property: $S(\varphi \circ \alpha, \psi \circ \alpha) = S(\varphi,\psi)$.

(c) *If $\mathscr{A} = \mathscr{A}_1 \oplus \mathscr{A}_2$ and $\varphi_{12}(a \oplus b) = \lambda\varphi_1(a) + (1-\lambda)\varphi_2(b)$, $\psi_{12}(a \oplus b) = \lambda\psi_1(a) + (1-\lambda)\psi_2(b)$, for all $a \in \mathscr{A}_1, b \in \mathscr{A}_2, \lambda \in (0,1)$, then the following joint convexity holds:*

$$S(\varphi,\psi) \leq S(\varphi_{12},\psi_{12}) = \lambda S(\varphi_1,\psi_1) + (1-\lambda)S(\varphi_2,\psi_2). \tag{5.7}$$

PROOF. (a) Let $P_\varphi = \sum_i \lambda_i p_i$ and $Q_\psi = \sum_j \kappa_j q_j$ be the spectral decompositions of the density matrices P_φ and Q_ψ. The Taylor expansion of the function $\eta(x) = -x\log x$ yields the inequality

$$-\eta(x) + \eta(y) + (x-y)\eta'(y) = -\frac{1}{2}(x-y)^2\eta''(\theta) \geq \frac{1}{2}(x-y)^2 \tag{5.8}$$

for some $\theta \in (x,y)$. Since $\eta(x) = -x\log x$ is convex on real intervals $[x,y]$, the functional $F(P) = \operatorname{Tr}[\eta(P)]$ is convex on the set of selfadjoint matrices with spectrum

in $[x,y]$, $\{P \in \mathcal{A}^{\mathrm{sa}} : \mathrm{Sp}(P) \subset [x,y]\}$. Thus

$$
\begin{aligned}
\operatorname{Tr} p_i q_j &\left[-\eta(P_\varphi) + \eta(Q_\psi) + (P_\varphi - Q_\psi)\eta'(Q_\psi) - \frac{1}{2}(P_\varphi - Q_\psi)^2 \right] \\
&= \left[-\eta(\lambda_i) + \eta(\kappa_j) + (\lambda_i - \kappa_j)\eta'(\kappa_j) - \frac{1}{2}(\lambda_i - \kappa_j)^2 \right] \operatorname{Tr} p_i q_j \geq 0.
\end{aligned}
\tag{5.9}
$$

Summing over i and j gives the desired inequality

$$
\operatorname{Tr}\left[-\eta(P_\varphi) + \eta(Q_\psi) + (P_\varphi - Q_\psi)(I - \log Q_\psi) - \frac{1}{2}(P_\varphi - Q_\psi)^2 \right] \geq 0.
\tag{5.10}
$$

(b) Consider the linear map $\beta : \mathcal{A}_1 \to \mathcal{A}_2$ defined by $\beta(AT_1^{1/2}) = \alpha(A)T_2^{1/2}$, for $A \in \mathcal{A}_1$ and the invertible positive operators $T_i \in \mathcal{A}_i$, $i = 1,2$, satisfying $\operatorname{Tr} T_2 \alpha(A) \leq \operatorname{Tr} T_1 A$, for $A \in \mathcal{A}_1^+$. Note that β is a contraction (with respect to Hilbert-Schmidt norm):

$$
\left\| \alpha(A)T_2^{1/2} \right\|^2 = \operatorname{Tr} T_2 \alpha(A)^* \alpha(A) \leq \operatorname{Tr} T_2 \alpha(A^* A) \leq \operatorname{Tr} T_1 A^* A = \left\| AT_1^{1/2} \right\|^2.
\tag{5.11}
$$

Next, introduce the positive operators γ and δ by

$$
\gamma\left(AT_1^{1/2}\right) = S_1 A T_1^{-1/2}, \quad A \in \mathcal{A}_1,
\tag{5.12}
$$

$$
\delta\left(BT_2^{1/2}\right) = S_2 B T_2^{-1/2}, \quad B \in \mathcal{A}_2,
\tag{5.13}
$$

where $S_i \in \mathcal{A}_i$, $i = 1,2$, are positive operators satisfying $\operatorname{Tr} S_2 \alpha(A) \leq \operatorname{Tr} S_1 A$, for $A \in \mathcal{A}_1^+$. Then

$$
\begin{aligned}
\left\langle \beta^* \delta \beta\left(AT_1^{1/2}\right), AT_1^{1/2} \right\rangle &= \left\langle \delta \alpha(A)T_2^{1/2}, \alpha(A)T_2^{1/2} \right\rangle \\
&= \left\langle S_2 \alpha(A)T_2^{1/2}, \alpha(A)T_2^{1/2} \right\rangle \\
&= \operatorname{Tr} S_2 \alpha(A)\alpha(A)^* \\
&\leq \operatorname{Tr} S_2 \alpha(AA^*) \leq \operatorname{Tr} S_1 AA^* \\
&= \left\langle \gamma\left(AT_1^{1/2}\right), AT_1^{1/2} \right\rangle.
\end{aligned}
\tag{5.14}
$$

Hence, $\beta^* \delta \beta \leq \gamma$. In particular, for $0 < t < 1$, $\beta^* \delta^t \beta \leq (\beta^* \delta \beta)^t \leq \gamma^t$, yielding for any $A \in \mathcal{A}_1$,

$$
\operatorname{Tr} \alpha(A)^* S_2^t \alpha(A) T_2^{1-t} \leq \operatorname{Tr} A^* S_1^t A T_1^{1-t}.
\tag{5.15}
$$

The inequality (5.6) is obtained by differentiating in (5.15) at $t = 0$ and using the following representation of the relative entropy (for nonnegative density matrices $P_\varphi = T_1$ and $Q_\psi = S_1$):

$$
\operatorname{Tr} P_\varphi (\log P_\varphi - \log Q_\psi) = -\frac{d}{dt} \operatorname{Tr} Q_\psi^t P_\varphi^{1-t} \big|_{t=0}.
\tag{5.16}
$$

An integral formulation similar to (5.16) is

$$
S(P_\varphi, Q_\psi) = \operatorname{Tr} P_\varphi \int_0^\infty \left[(Q_\psi + tI)^{-1} (P_\varphi - Q_\psi)(P_\varphi + tI)^{-1} \right] dt.
\tag{5.17}
$$

(c) Let T_1 and S_1 be the densities of φ and ψ; likewise, let T_2 and S_2 be the densities of $\varphi_{12}(a \oplus b) = \lambda\varphi_1(a) + (1-\lambda)\varphi_2(b)$ and $\psi_{12}(a \oplus b) = \lambda\psi_1(a) + (1-\lambda)\psi_2(b)$, respectively. Then (5.17) holds and

$$\left[(1-t)^{-1}\left(1 - \operatorname{Tr} S_2^t T_2^{1-t}\right) \geq (1-t)^{-1}\left(1 - \operatorname{Tr} S_1^t T_1^{1-t}\right)\right]. \tag{5.18}$$

Letting $t \nearrow 1$ yields the joint convexity of the relative entropy: $S(\varphi, \psi) \leq \lambda S(\varphi_1, \psi_1) + (1-\lambda)S(\varphi_2, \psi_2)$. $\square$

Allowing other functions in place of the logarithm, broadens further the concept of relative entropy. For instance, considering the operator convex function $f : (0, \infty) \to \mathbb{R}$, with $f(1) = 0$, $f(x) = x\log x$, then $S_f(P, Q) = S_{BS}(P, Q)$. Another notable special case is $f(x) = (x-1)^2$, producing the so-called *quadratic relative entropy* (the χ^2-divergence)

$$S_{(x-1)^2}(P, Q) = \operatorname{Tr} P^{-1}(P - Q)^2. \tag{5.19}$$

PROPOSITION 5.2 (see [10]). *The quadratic relative entropy satisfies the properties:*
  (a) $S_{(x-1)^2}(P, Q) = \operatorname{Tr}(X^*X)$, *with* $X = P^{-1/2}(P - Q)$;
  (b) $S_{(x-1)^2}(P, Q) > 0$ *if* $P \neq Q$;
  (c) $S_{(x-1)^2}(\alpha(P), \alpha(Q)) \leq S_{(x-1)^2}(P, Q)$, *for every trace-preserving Schwarz mapping* α *of the matrix algebra, which sends density matrices into density matrices;*
  (d) *the quadratic entropy coefficient cannot exceed 1. If P and Q are invertible density matrices, then*

$$\eta_{(x-1)^2}(\alpha) = \sup_{P \neq Q}\left\{\frac{S_{(x-1)^2}(\alpha(P), \alpha(Q))}{S_{(x-1)^2}(P, Q)}\right\} \leq 1. \tag{5.20}$$

PROOF. (a) is obvious; (b) follows from (5.5); (c) is a restatement of (5.6); (d) follows from (c). $\square$

REMARK 5.3. If α is interpreted as a communication channel between two finite quantum systems, then the inequality (5.20) says that the output information corresponding to the states $\alpha(P)$, $\alpha(Q)$ is always less than or equal to the input information of the states P and Q; $S_f(\alpha(P), \alpha(Q))$ represents the amount of information correctly transmitted through the channel α, and $\eta_f(\alpha)$ can be viewed as the efficiency of the communication channel.

The quadratic relative entropy (5.19) can be extended for an arbitrary operator convex function $f : (0, \infty) \to \mathbb{R}$ with $f(1) = 0$ when the density matrix P is assumed to be invertible. The function f admits the integral representation

$$f(x) = a(x-1) + b(x-1)^2 + c\frac{(x-1)^2}{x} + \int_1^\infty \frac{(x-1)^2}{x+t}\,d\nu(t) \tag{5.21}$$

by means of a finite positive measure ν on $\mathbb{R}$. Based on the integral representation (5.21) it was recently shown by Petz and Ruskai [10] that $\eta_f(\alpha) = \eta_{(x-1)^2}(\alpha)$, for any operator convex function f. Since the relative entropy coefficient $\eta_f(\alpha)$ is independent of the operator convex function f, it is the object of current investigations to establish

whether the relative entropy coefficient depends on the Riemannian metric defined on the manifold of density matrices.

A generalized relative entropy induces a Riemannian metric on the state space. The set $\mathcal{M}$ of positive definite density matrices $P = (P_{ij})$ (with normalized trace, equal to 1) can be parametrized in an affine way, by the real numbers $\operatorname{Re} P_{ij}$, $\operatorname{Im} P_{ij}$ $(1 \le i < j \le n)$ and the positive numbers P_{ii}. Thus $\mathcal{M}$ can be embedded into the Euclidean k-space with $k = n^2 - 1$ and become a manifold. At each point $P \in \mathcal{M}$, the tangent space $T_P(\mathcal{M})$ is identified with the set of all traceless selfadjoint matrices. Then, an inner product can be defined on $T_P(\mathcal{M})$ by

$$g_P(A, A) = \operatorname{Tr} P^{-1} A^2. \tag{5.22}$$

In the commutative case, a positive and trace-preserving Riemannian metric corresponding to the (logarithmic) relative entropy

$$S_{\log}(P, Q) = \operatorname{Tr} P(\log P - \log Q) = \operatorname{Tr} P \int_0^\infty \left(\frac{1}{Q + tI}(P - Q)\frac{1}{P + tI} \right) dt \tag{5.23}$$

is

$$g_P^{\log}(A, B) = \operatorname{Tr} A^* P^{-1} B. \tag{5.24}$$

The quantum noncommutative case is richer, since the left and right multiplications by P^{-1} are not equivalent. In this setting, an analogue of $P^{-1}B$ can be defined by $\Omega_P^{\log}(B) = \int_0^\infty ((1/P + tI)B(1/P + tI)) dt$, hence

$$g_P^{\log}(A, B) = \operatorname{Tr} A^* \Omega_P^{\log}(B). \tag{5.25}$$

This Riemannian metric coincides with

$$\begin{aligned} g_P^{\log}(A, B) &= -\frac{\partial^2}{\partial u \, \partial v} S_{\log}(P + uA, P + vB)|_{u=v=0} \\ &= \operatorname{Tr} A \int_0^\infty \left(\frac{1}{P + tI} B \frac{1}{P + tI} \right) dt \end{aligned} \tag{5.26}$$

(for $A \in T_P(\mathcal{M}) = \{A = A^* : \operatorname{Tr} A = 0\}$). Obviously one may extend the metric (5.25) to other operator convex functions $f : (0, \infty) \to \mathbb{R}$, $f(1) = 0$ besides $f(x) = x \log x$. For $f(x) = (x - 1)^2$, corresponding to the symmetrized quadratic relative entropy

$$S_{(x-1)^2}^{\text{sym}}(P, Q) = \operatorname{Tr}(P - Q)P^{-1}(P - Q), \tag{5.27}$$

one has the Riemannian metric

$$g_P^{(x-1)^2}(A, B) = \operatorname{Tr} A^* \Omega_P(B) = \operatorname{Tr} A^* (BP^{-1} + P^{-1}B). \tag{5.28}$$

Further entropic properties of quantum systems can be discussed in terms of the geodesic distance defined as

$$\gamma_f(P, Q) = \inf_{\mathcal{S}} \int_0^1 \sqrt{\langle \dot{\omega}(t), \Omega_{\omega(t)}^f \dot{\omega}(t) \rangle} \, dt, \tag{5.29}$$

where $\mathcal{S} = \{\omega(t) \text{ smooth path} \mid \omega(0) = P, \ \omega(1) = Q\}$.

The squared geodesic distance $[\gamma_f(P, Q)]^2$ is a differentiable monotone relative entropy distance, that is, it satisfies the properties:

(a) $[\gamma_f(P,Q)]^2 \geq 0$, with equality if and only if $P = Q$.

(b) $\gamma_f(P,Q) = \gamma_f(Q,P)$, for all $P,Q \in \mathcal{M}$.

(c) $\gamma_f(P,Q) \leq \gamma_f(P,R) + \gamma_f(R,Q)$, for all $P,Q,R \in \mathcal{M}$.

(d) $\gamma_f(\kappa P, \kappa Q) = \kappa \gamma_f(P,Q)$, for scalars $\kappa > 0$.

(e) $\gamma_f(P,Q)$ is jointly convex in P and Q.

(f) $\gamma_f(\alpha(P),\alpha(Q)) \leq \gamma_f(P,Q)$, for any trace-preserving Schwarz map α.

(g) The function $h(u,v) = S(P + uA, Q + vB)$ is differentiable.

The above properties are easily verified in light of Propositions 5.1 and 5.2; (g) follows from Theorem 3.6(2) [5].

In [7] it was shown that $\gamma_f(P,Q)$ is majorized by $\gamma^{\text{sym}}_{(x-1)^2}(P,Q)$, for any operator convex function f. Moreover, for each fixed convex operator function f and Schwarz trace-preserving map α, the following entropy contraction coefficients:

$$\eta_f^{\text{RelEnt}}(\alpha) = \sup_{P \neq Q} \left\{ \frac{S_f(\alpha(P),\alpha(Q))}{S_f(P,Q)} \right\}, \tag{5.30}$$

$$\eta_f^{\text{Riem}}(\alpha) = \sup_{P} \sup_{A \in T_P(\mathcal{M})} \left\{ \frac{\langle \alpha(A), \Omega^f_{\alpha(P)}[\alpha(A)] \rangle}{\langle A, \Omega^f_P(A) \rangle} \right\}, \tag{5.31}$$

$$\eta_f^{\text{geod}}(\alpha) = \sup_{P \neq Q} \left\{ \frac{[\gamma_f(\alpha(P),\alpha(Q))]^2}{[\gamma_f(P,Q)]^2} \right\}, \tag{5.32}$$

satisfy the inequality: $\eta_f^{\text{geod}}(\alpha) \leq \eta_f^{\text{Riem}}(\alpha) \leq \eta_f^{\text{RelEnt}}(\alpha) \leq 1$ (see [7]).

6. Conclusion. The concept of entropy relies on having chosen certain measures (like phase space volumes, or Hilbert space norms) which are conserved under unitary time evolution. We highlighted the importance of the probability distributions corresponding to the extremal cases for the relative entropy and free entropy in various instances. These entropy inequalities may be viewed as manifestations of a universal principle of nondecrease of uncertainty, similar to the second law of thermodynamics, as remarked by Gnedenko and Korolev in [3].

The powerful blend of operatorial algebraic and probabilistic points of view afforded by the free probability theory provides the ideal framework for describing the limiting behavior of special classes of random matrices, as their size tends to infinity. Such large random matrices are extensively used in physical modeling (e.g., Ising models, phase transition phenomena, spin tunneling) motivating the interest in further investigating this topic in future work.

References

[1]   A. R. Barron, *Entropy and the central limit theorem*, Ann. Probab. **14** (1986), no. 1, 336–342. MR 87h:60048. Zbl 599.60024.

[2]   V. P. Belavkin and P. Staszewski, *C*-algebraic generalization of relative entropy and entropy*, Ann. Inst. H. Poincaré Sect. A (N.S.) **37** (1982), no. 1, 51–58. MR 84b:46083. Zbl 526.46060.

[3]   B. V. Gnedenko and V. Y. Korolev, *Random Summation*, Limit theorems and applications, CRC Press, Boca Raton, FL, 1996. MR 97e:60040. Zbl 857.60002.

[4]   S. Goldman, *Information Theory*, Prentice-Hall Inc., New York, 1953. MR 16,269c.

[5]   S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry. Vol I*, Interscience Publishers, a division of John Wiley & Sons, New York, London, 1963. MR 27#2945.

[6]   B. Koopman and J. von Neumann, *Dynamical systems of continuous spectra*, Proc. Nat. Acad. Sci. U.S.A. **18** (1932), 255–266.

[7]   A. Lesniewski and M. B. Ruskai, *Monotone Riemannian metrics and relative entropy on noncommutative probability spaces*, J. Math. Phys. **40** (1999), no. 11, 5702–5724. CMP 1 722 334. Zbl 991.40675.

[8]   A. Nica, D. Shlyakhtenko, and R. Speicher, *Some minimization problems for the free analogue of the Fisher information*, Adv. Math. **141** (1999), no. 2, 282–321. MR 2000b:46120. Zbl 929.46052.

[9]   M. Ohya and D. Petz, *Quantum Entropy and its Use*, Texts and Monographs in Physics, Springer-Verlag, Berlin, 1993. MR 94k:81002. Zbl 891.94008.

[10] D. Petz and M. B. Ruskai, *Contraction of generalized relative entropy under stochastic mappings on matrices*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **1** (1998), no. 1, 83–89. MR 99b:82015. Zbl 917.94006.

[11] H. Umegaki, *Conditional expectation in an operator algebra. IV. Entropy and information*, Kōdai Math. Sem. Rep. **14** (1962), 59–85. MR 25#5401. Zbl 199.19706.

[12] D. Voiculescu, *Addition of certain noncommuting random variables*, J. Funct. Anal. **66** (1986), no. 3, 323–346. MR 87j:46122. Zbl 651.46063.

[13] _____, *The analogues of entropy and of Fisher's information measure in free probability theory. I*, Comm. Math. Phys. **155** (1993), no. 1, 71–92. MR 94k:46137. Zbl 781.60006.

[14] _____, *Free Probability Theory*, Papers from the Workshop on Random Matrices and Operator Algebra Free Products held during the Special Year on Operator Algebra at the Fields Institute for Research in Mathematical Sciences, Waterloo, ON, March 1995. Edited by Dan-Virgil Voiculescu. Fields Institute Communications, vol. 12, American Mathematical Society, Providence, RI, 1997. MR 97g:46002. Zbl 859.00025.

[15] _____, *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. MR 99d:46087. Zbl 930.46053.

PATRICIA GIURGESCU: MATHEMATICS DEPARTMENT, PACE UNIVERSITY, 861 BEDFORD RD., PLEASANTVILLE, NY, 10507, USA

E-mail address: pgiurgescu@pace.edu