Large Deviations and Microstate
Free Relative Entropy

Masaru MIZUO

Graduate School of Information Sciences, Tohoku University,
Aoba-ku, Sendai 980-8577, Japan

Received April 14, 2000; final version accepted July 14, 2000

Voiculescu’s single variable free entropy is generalized to the free relative entropy in the microstate approach, and the integrat representation of the free relative entropy is obtained by large deviation results for random matrices. It is also shown that the free relative entropy shares common properties of the classical relative entropy.

1 Introduction

Since Voiculescu opened it in early 1980’s ([10]), a new noncommutative probability theory called free probability theory has been extensively developed, where the concept “freeness” plays a role of “independence” in classical probability theory. The use of freeness instead of independence makes free probability very different from classical one, however there is strong parallelism between their basic structures. Indeed, many “free” analogues of classical theory are known, for instance, the semicircle law versus the Gaussian law, the free central limit theorem ([10]), the free convolutions ([15]), the free infinite divisible distributions ([3]), the free entropy versus the classical Boltzmann-Gibbs entropy, etc. Random matrices play a vital role in free probability theory as asymptotic models of free random variables, and we have large deviations for random matrices where the main term of the rate function is the free entropy (or the logarithmic energy with different sign). The asymptotic freeness result for random matrices was first established by Voiculescu [11] and improved in [14]. On the other hand, the large deviation theorem was first proved for Gaussian random matrices by Ben Arous and Guionnet [2]. These themes are exposed in detail in [8].

Beyond the single variable case in [12], Voiculescu introduced in [13] the multi-variable free entropy for noncommutative random variables. Let $M_n(C)$ be the algebra of complex matrices (or microstates) of order $n$, and $tr_n$ be the normalized trace on $M_n(C)$. There is a natural linear bijection between the set $M_n(C)^{\otimes N}$ of selfadjoint matrices in $M_n(C)$ and the Euclidean space $\mathbb{R}^{n^2}$, which is an isometry with respect to the Hilbert-Schmidt and Euclidean norms. The Lebesgue measure $\mu_n$ on $M_n(C)^{\otimes N}$ is induced by the usual Lebesgue measure on $\mathbb{R}^{n^2}$ under this bijection, and $\mathcal{A}_n^\otimes N$ is the $N$-fold product measure on the product $(M_n(C)^{\otimes N})^\otimes N$ for $N \in \mathbb{N}$. Let $(\mathcal{M}, \tau)$ be a tracial $W^*$-probability space, that is, $\mathcal{M}$ is a von Neumann algebra and $\tau$ is a faithful normal tracial state on $\mathcal{M}$. Let $a_1, \ldots, a_N \in \mathcal{M}^{\otimes N}$, where $\mathcal{M}^{\otimes N}$ denotes the space of selfadjoint elements of $\mathcal{M}$. Throughout the paper we will write $a_1, \ldots, a_N \in (\mathcal{M}^{\otimes N}, \tau)$ for this case. The microstate free entropy $\chi(a_1, \ldots, a_N)$ of $a_1, \ldots, a_N \in (\mathcal{M}^{\otimes N}, \tau)$ is defined by the following procedure ([13]): For $n, r \in \mathbb{N}$ and $\varepsilon > 0$,

$$\Gamma_n(a_1, \ldots, a_N; n, r, \varepsilon) = \{ (a_1, \ldots, A_N) \in (M_n(C))^{\otimes N}; \|A\| \leq R, \|\text{tr}_n(A_{i_1} \cdots A_{i_r}) - \tau(a_{i_1} \cdots a_{i_r})\| \leq \varepsilon \}
$$

for all $1 \leq i_1, \ldots, i_k \leq N$, $1 \leq k \leq r$,

$$\chi_n(a_1, \ldots, a_n; r, \varepsilon) = \lim_{n \to \infty} \sup \left\{ \frac{1}{n^2} \log \mathcal{A}^{\otimes N}_n(\Gamma_n(a_1, \ldots, a_N; n, r, \varepsilon)) + \frac{N}{2} \log n \right\},$$

(1)

$$\chi(a_1, \ldots, a_N) = \lim_{r \to \infty} \lim_{\varepsilon \to 0} \chi_n(a_1, \ldots, a_N; r, \varepsilon),$$

$$\chi(a_1, \ldots, a_N) = \sup_{R \geq 0} \chi(R, a_1, \ldots, a_N).$$

In particular, in the case of single random variable $a \in (\mathcal{M}^{\otimes N}, \tau)$, if $\mu$ is the probability distribution of $a$ with respect to $\tau$, then we have

$$\Gamma_n(a; n, r, \varepsilon) = \{ A \in M_n(C)^{\otimes N}; \|A\| \leq R, \|\text{tr}_n(A^k) - m_k(\mu)\| \leq \varepsilon \text{ for all } 1 \leq k \leq r \},$$

where $m_k(\mu) = \int x^k d\mu(x)$ is the $k$th moment of $a$. Moreover, for any $R \geq \|a\|$, limsup becomes lim in (1) so that

$$\chi_n(a; r, \varepsilon) = \lim_{n \to \infty} \left[ \frac{1}{n^2} \log \mathcal{A}_n(\Gamma_n(a; n, r, \varepsilon)) + \frac{1}{2} \log n \right].$$
and
\[ \chi(a) = \chi_E(a) = \Sigma(\mu) + \frac{1}{2}\log(2\pi) + \frac{3}{4}, \]
where \( \Sigma(\mu) \equiv \int \log |x - y| \, d\mu(x) d\mu(y) \) which is the single variable free entropy introduced in [12]. In the following, identifying \( a \in (\mathcal{M}^*, \tau) \) with its distribution \( \mu \), we will use also \( \Sigma(a) \) for \( \Sigma(\mu) \) and \( I^E(\mu; n, r, e) \) for \( I^E_E(\mu; a, n, r, e) \). The above defined free entropy enjoys a number of properties considered natural as an entropic quantity measuring the degree of freeness. Some of important properties are listed as follows. Let \( a_1, \ldots, a_N \in (\mathcal{M}^*, \tau) \). Then:

1. **Subadditivity**: If \( C = \tau(a_1^2 + \cdots + a_N^2) \), then
   \[ \chi(a_1, \ldots, a_N) \leq \chi(a_1) + \cdots + \chi(a_N) \leq \frac{N}{2} \log \frac{2\pi e C}{N}. \]

2. **Upper semicontinuity**: Let \( a_{m,1}, \ldots, a_{m,N} \in (\mathcal{M}^*, \tau) (m \in \mathbb{N}) \). If \( (a_{m,1}, \ldots, a_{m,N}) \to (a_1, \ldots, a_N) \) in the distribution sense as \( m \to \infty \) and \( \sup_m \| a_{m,i} \| \to +\infty \) (\( 1 \leq i \leq N \)), then
   \[ \chi(a_1, \ldots, a_N) \geq \limsup_{m \to \infty} \chi(a_{m,1}, \ldots, a_{m,N}). \]

3. **Additivity**: If \( a_1, \ldots, a_N \) are in free relation, then
   \[ \chi(a_1, \ldots, a_N) = \chi(a_1) + \cdots + \chi(a_N). \] (2)

Conversely, if \( \chi(a_i) > -\infty \) (\( 1 \leq i \leq N \)) and the above equality (2) holds, then \( a_1, \ldots, a_N \) are in free relation.

In this paper we will generalize the single variable free entropy to its relative version and examine some properties of the free relative entropy. To justify our formulation, we start with the microstate representation of the classical relative entropy from the viewpoint of large deviations. (The relative version of the free entropy is also considered in [4] from another viewpoint.)

In the rest of the section we introduce some terminologies on random matrices for later use. Let \((\Omega, \mathcal{B}, P)\) be a usual probability space. An \( n \times n \) random matrix \( T = [T_{ij}] \) on \((\Omega, \mathcal{B}, P)\) is an \( M_n(C) \)-valued measurable function on \( \Omega \), or all entries \( T_{ij} (1 \leq i, j \leq n) \) are complex random variables on \( \Omega \). We always assume that each entry \( T_{ij} \) has all finite moments. The canonical tracial state \( \tau_n \) is defined for such random matrices by \( \tau_n(T) = E(\tr_n(T)) = (1/n) \sum_{j=1}^n E(T_{jj}) \), where \( E \) means the expectation with respect to \( P \). An \( n \times n \) random matrix \( T \) induces the probability distribution \( v_n \) on \( M_n(C) \), and an \( n \)-tuple \((T_1, \cdots, T_n)\) of \( n \times n \) random matrices does the distribution \( \mu_n \) on \( M_n(C)^n \), so we may take \((M_n(C), v_n)\) or \((M_n(C)^n, \mu_n)\) instead of \((\Omega, \mathcal{B}, P)\). In this way, we sometimes identify \( T \) with \((M_n(C), v_n)\), and for simplicity we write \( T = (M_n(C), v_n) \) and also \( T = (M_n(C)^n, v_n) \) if \( T \) is selfadjoint. Let \( X \in M_n(C)^n \) and \( \lambda_i(X) = \cdots \leq \lambda_n(X) \) be the eigenvalues of \( X \) ordered increasingly (with multiplicities). Then \( \lambda \) can be regarded as a mapping

\[ \lambda: X \in M_n(C)^n \mapsto (\lambda_1(X), \ldots, \lambda_n(X)) \in \mathbb{R}^n, \]

where \( \mathbb{R}^n_+ = \{(x_1, \ldots, x_n) \in \mathbb{R}^n: x_1 \leq \cdots \leq x_n\} \). The distribution of the eigenvalues of a random matrix \( T = (M_n(C)^n, v_n) \) is the image probability measure \( \lambda(v_n) \) on \( \mathbb{R}^n_+ \). But we treat the symmetrically extended probability measure on \( \mathbb{R}^n \) as the joint eigenvalue distribution of \( T \) so that it is invariant under the coordinate permutations on \( \mathbb{R}^n \). Furthermore, the empirical spectral density of \( T \) is defined to be a random discrete measure \((1/n)(\delta(\lambda_1(T)) + \cdots + \delta(\lambda_n(T)))\), where \( \delta(x) \) is the Dirac measure at \( x \).

## 2 Microstate representation of classical relative entropy

The next theorem due to Sanov (see [6]) is a typical example of level-2 large deviation theorems. When \( S \) is a closed subset of \( \mathbb{R} \), we denote by \( \mathcal{M}(S) \) the set of all probability Borel measures on \( S \), which is a complete separable metric space with respect to the weak topology.

**Theorem 2.1** Let \( \xi_1, \xi_2, \cdots \) be a sequence of independent identically distributed real random variables on a probability space \((\Omega, \mathcal{B}, P)\). Let \( v \) be the common distribution of \( \xi_n (n \in \mathbb{N}) \) whose support is included in the interval \( S = [-K, K] \). Then the sequence of random probability measures \((1/n)(\delta(\xi_1) + \cdots + \delta(\xi_n)) \in \mathcal{M}(S) \) (\( n \in \mathbb{N} \)) satisfies the large deviation principle in the scale \( 1/n \). Namely, for any closed set \( F \subset \mathcal{M}(S) \) and any open set \( G \subset \mathcal{M}(S) \) the following hold:

\[
\limsup_{n \to \infty} \frac{1}{n} \log P\left(\frac{\delta(\xi_1) + \cdots + \delta(\xi_n)}{n} \in F\right) \leq -\inf\{I(\mu): \mu \in F\},
\]

\[
\liminf_{n \to \infty} \frac{1}{n} \log P\left(\frac{\delta(\xi_1) + \cdots + \delta(\xi_n)}{n} \in G\right) \geq -\inf\{I(\mu): \mu \in G\},
\]
where the rate function $I(\mu) = S(\mu|v)$ for $\mu \in \mathcal{M}(S)$ is the classical relative entropy of $\mu$ with respect to $v$ defined by
\[
S(\mu|v) = \begin{cases} 
\frac{d\mu}{dv} \log \frac{d\mu}{dv} & \text{if } \mu \ll v, \\
\infty & \text{otherwise}.
\end{cases}
\]

It is well known that the classical relative entropy is weakly lower semicontinuous and strictly convex. Furthermore, it is a good rate function, that is, the level set $\{ \mu \in \mathcal{M}(S) : I(\mu) \leq c \}$ for any $c \geq 0$ is compact in $\mathcal{M}(S)$. Also, note that $v$ is a unique minimizer with $I(v) = 0$. The above large deviation theorem implies that the random measure $(1/n)(\delta(z_1) + \cdots + \delta(z_n))$ almost surely converges to $v$ in the weak topology. If the support of $v$ is compact, then the random measure almost surely converges in the distribution sense, that is, each moment of the random measure almost surely converges to that of $v$.

We can apply the Sanov theorem to obtain the microstate representation for the classical relative entropy. Let $\mathcal{M}_0(\mathbb{R})$ be the set of all probability Borel measures on $\mathbb{R}$ with compact support. We denote by $M_n(C)^{\text{diag}}$ the set of all selfadjoint diagonal matrices of order $n$. There is a natural linear bijection between $M_n(C)^{\text{diag}}$ and $\mathbb{R}^n$, which is an isometry with respect to the Hilbert-Schmidt and Euclidean norms. For $v \in \mathcal{M}_0(\mathbb{R})$ the probability measure induced by the $n$-fold product measure $v^n$ on $\mathbb{R}^n$ under this bijection will be denoted by $\Omega_n(v)$, which is also regarded as a measure on $M_n(C)^{\text{diag}}$ under the inclusion $M_n(C)^{\text{diag}} \subset M_n(C)^{\text{sa}}$.

**Proposition 2.2** Let $\mu, v \in \mathcal{M}_0(\mathbb{R})$. For $r \in \mathbb{N} \setminus 0$, $R > 0$ define
\[
S_R(\mu|v; r, \varepsilon) = \lim_{n \to \infty} \inf_{\mathcal{A}_n} \left\{ -\frac{1}{n} \log \Omega_n(v)(\Gamma_n(\mu; n, r, \varepsilon)) \right\},
\]
\[
S_R(\mu|v) = \lim_{r \to 0, \varepsilon \to 0} S_R(\mu|v; r, \varepsilon) = \sup_{R > 0} S_R(\mu|v; r, \varepsilon).
\]

Then
\[
\inf_{R > 0} S_R(\mu|v) = S(\mu|v).
\]
Moreover, if $\text{supp } \mu \subset [-R, R]$ for some $0 < R < +\infty$, then
\[
S_R(\mu|v; r, \varepsilon) = \lim_{n \to \infty} \left\{ -\frac{1}{n} \log \Omega_n(v)(\Gamma_n(\mu; n, r, \varepsilon)) \right\},
\]
\[
S_R(\mu|v) = S_R(\mu|v).
\]

See [8] for the proof of the proposition. Let $u_k$ be the uniform distribution on the interval $[-K, K]$ ($0 < K < +\infty$). If the support of $\mu \in \mathcal{M}_0(\mathbb{R})$ is included in $[-K, K]$, then
\[
\lim_{r \to 0, \varepsilon \to 0} \lim_{n \to \infty} \left\{ -\frac{1}{n} \log \Omega_n(u_k)(\Gamma_n(\mu; n, r, \varepsilon)) \right\} = S_R(\mu|u_k) = S(\mu|u_k) = -S(\mu) + \log 2K,
\]
where $S(\mu)$ is the Boltzmann-Gibbs entropy of $\mu$ defined by
\[
S(\mu) = \begin{cases} 
\frac{d\mu}{dx} \log \frac{d\mu}{dx} & \text{if } \mu \ll dx (dx \text{ is the Lebesgue measure}), \\
\infty & \text{otherwise}.
\end{cases}
\]

Hence for $\mu \in \mathcal{M}_0(\mathbb{R})$,
\[
S(\mu) = \sup_{R > 0} \lim_{r \to 0, \varepsilon \to 0} \lim_{n \to \infty} \left\{ -\frac{1}{n} \log \Omega_n(u_k)(\Gamma_n(\mu; n, r, \varepsilon)) + \log 2R \right\}
\]
\[
= \sup_{R > 0} \lim_{r \to 0, \varepsilon \to 0} \lim_{n \to \infty} \left\{ -\frac{1}{n} \log \Omega_n(\Gamma_n(\mu; n, r, \varepsilon)) \right\},
\]
where $\Omega_n$ is the measure on $M_n(C)^{\text{diag}}$ induced by the Lebesgue measure on $\mathbb{R}^n$ under the above bijection. (4) is the microstate representation for the Boltzmann-Gibbs entropy.

All $\Lambda_n$ in (1), $\Omega_n(v)$ in (3) and $\Omega_n$ in (4) are used as a measurement tool for the set $\Gamma_n(\mu; n, r, \varepsilon)$ of matrices (microstates) approximating $\mu$ in the distribution sense. So we will call $\Omega_n(v)$, $\Omega_n$ and $\Lambda_n$ reference measures. The reference measures of the Boltzmann-Gibbs entropy and the free entropy are the Lebesgue measures on $M_n(C)^{\text{diag}}$ and $M_n(C)^{\text{sa}}$, respectively. The reference measures of the classical relative entropy have the following properties which reflect classical probability theory.

**Proposition 2.3** (1) Let $v_i \in \mathcal{M}_0(\mathbb{R})$ ($i \in I$) and, for each $n \in \mathbb{N}$, $T_n$ ($i \in I$) be selfadjoint diagonal random matrices of order $n$ on $(\Omega, \mathcal{F}, P)$ having the distribution $(M_n(C)^{\text{diag}}, \Omega_n(v_i))$. Assume that, for each $n \in \mathbb{N}$,
$T_{n,i} \ (i \in I)$ are independent, that is, the distribution of the random matrices $T_{n,i} \ (i \in I)$ is $\otimes_{i \in I} \Omega_n(v_i)$ on $(M_n(C^{\text{diag}}))^I$. Then the random matrices $T_{n,i} \ (i \in I)$ almost surely converge as $n \to \infty$ to the product measure $\otimes_{i \in I} v_i$ in the distribution sense, that is, any joint moment of $T_{n,i} \ (i \in I)$ with respect to $\nu_n$ converges almost surely to that of $\otimes_{i \in I} v_i$ on $R^I$.

(2) Let $T$ be a selfadjoint diagonal random matrix of order $n$. If the mean $r_n(T)$ is 0 and the $p$th absolute moment $\tau_n(|T|^p)$ is fixed for some $p > 0$, then the random matrix $T = (M_n(C^{\text{diag}}, \Omega_n((1/Z) \exp(-c|x|^p))))$ has the maximal Boltzmann-Gibbs entropy, where $Z$ is a normalization constant and $c$ is a constant depending on $\tau_n(|T|^p)$. Here the Boltzmann-Gibbs entropy of a random matrices $(M_n(C^{\text{diag}}, v_n)$ is defined as that of an $n$-variable random vector under the identification between $M_n(C^{\text{diag}})$ and $R^n$.

In the above (1), the convergence of moments of random matrices $T_{n,i}$ for each fixed $i$ was already mentioned after Theorem 2.1. The almost sure convergence of joint moments of random matrices $T_{n,i} \ (i \in I)$ is due to the Sanov theorem for the multi-variable case (see [6]). The maximization problem of the Boltzmann-Gibbs entropy in (2) can be solved by the positivity of the relative entropy (see [5]); (2) says that the reference measures appearing in the above (2) are the most uniformly distributed on the selfadjoint diagonal matrices under a certain constraint.

In the next section we will look for the reference measures of the free relative entropy by replacing Proposition 2.3 in classical probability theory with its counterpart in free probability theory.

3 Definition of free relative entropy and its integral representation

The next theorem in [9] from potential theory will play a key role in the following discussions.

**Theorem 3.1** Let $S$ be a closed subset of $R$. Let $w: S \to [0, \infty)$ be a weight function, and assume the following conditions:

1. $w$ is continuous on $S$.
2. $S_0 = \{x \in S: w(x) > 0\}$ has the positive (inner logarithmic) capacity, that is, $E(\mu) < +\infty$ for some probability measure $\mu \in S(S)$ where $E(\nu) = -\int |x - y| d\nu(x) d\nu(y)$, the logarithmic energy.
2. $|x| w(x) \to 0$ as $x \in S$, $|x| \to \infty$, when $S$ is unbounded.

Let $Q(x) = -\log w(x)$ and define the weighted energy integral

$$E_Q(\mu) = \int (|x - y| + Q(x) + Q(y)) d\mu(x) d\mu(y).$$

(Note that $E_Q(\mu) > -\infty$ is well defined thanks to the condition 3.) Then, there exists a unique $\mu_Q \in S(S)$ such that $E_Q(\mu_Q) = \inf \{E_Q(\mu): \mu \in S(S)\}$; $E_Q(\mu)$ is finite and supp $\mu_Q$ is compact. Furthermore, this minimizer $\mu_Q$ is characterized as the measure with compact support such that for some real number $C_Q$ the following holds:

$$\int |x - y| d\mu_Q(y) \geq Q(x) - C_Q \quad \text{for all } x \in \text{supp } \mu_Q,$$

$$\int |x - y| d\mu_Q(y) \leq Q(x) - C_Q \quad \text{for } x \in S \text{ except in a set of capacity zero}.$$

In this case, $C_Q = E_Q(\mu_Q) - \int Q d\mu_Q$.

As in [8] with use of the above theorem, one can prove the next fundamental example of level-2 large deviation theorems in free probability, which is a counterpart of the Sanov theorem (Theorem 2.1) in classical probability.

**Theorem 3.2** Let $Q(x)$ be a continuous function on the interval $S = [-K, K] (0 < K < +\infty)$. For each $n \in N$ let $\xi_{1,n}, \xi_{2,n}, \ldots, \xi_{n,n}$ be real random variables on a probability space $(\Omega, \mathcal{F}, P)$ having the following joint probability distribution on $R^n$

$$\tilde{Z}_n(Q; K) = \frac{1}{Z_n(K)} \exp \left( -n \sum_{i=1}^n Q(\xi_i) \sum_{i,j=1}^n |x_i - x_j|^2 \sum_{i=1}^n \chi_{[-K,K]}(x_i) dx_i dx_j \cdots dx_n \right),$$

where $\chi_{[-K,K]}$ is the characteristic function of the interval $S$ and $Z_n(K)$ is a normalization constant

$$Z_n(K) = \int_{-K}^K \cdots \int_{-K}^K \exp \left( -n \sum_{i=1}^n Q(\xi_i) \sum_{i,j=1}^n |x_i - x_j|^2 \sum_{i=1}^n \chi_{[-K,K]}(x_i) dx_i dx_j \cdots dx_n \right).$$

Then the large deviation principle in the scale $1/n^2$ is satisfied for the sequence of random measures $(1/n) (\delta(\xi_{1,n}) + \ldots + \delta(\xi_{n,n}))$ as follows: for any closed $F$ and open $G$ in $S(S)$,

$$\limsup_{n \to \infty} \frac{1}{n^2} \log P \left( \frac{\delta(\xi_{1,n}) + \ldots + \delta(\xi_{n,n})}{n} \in F \right) \leq -\inf_{P(\mu) \in F} \{I(\mu) : \mu \in S(S)\},$$

$$\liminf_{n \to \infty} \frac{1}{n^2} \log P \left( \frac{\delta(\xi_{1,n}) + \ldots + \delta(\xi_{n,n})}{n} \in G \right) \geq -\inf_{P(\mu) \in G} \{I(\mu) : \mu \in S(S)\},$$

where the finite limit $B(K) = \lim_{n \to \infty} (1/n^2) \log Z_n(K)$ exists and the rate function $I(\mu)$ for $\mu \in S(S)$ is given
by
\[ I(\mu) = - \sum \mu(x) + \int Q(x) d\mu(x) + B(K). \]

The above \( I(\mu) \) is a convex function satisfying all requirements as a good rate function. Furthermore, there exists a unique \( \nu_0 \) such that \( I(\nu_0) = \inf \{ I(\mu) : \mu \in \mathcal{M}(S) \} = 0 \). The random probability measure \((1/n)\delta(\xi_1) + \delta(\xi_2) + \cdots + \delta(\xi_n))\) almost surely converges to \( \nu_0 \) in the distribution sense.

We are concerned with \( \nu \in \mathcal{M}_0(R) \) such that \( Q(\nu)(x) = \int |x - y| d\nu(y) \) is a continuous function on \( R \). For example, if \( \nu \) has the Radon-Nikodim derivative \( d\nu/dx \in L^\infty(R) \) with respect to the Lebesgue measure, then \( Q(\nu) \) is easily checked to be continuous. In [7] the next proposition was obtained.

**Proposition 3.3** For a probability measure \( \nu \) on \( R \) with supp \( \nu \subset [-K, K] \) for some \( 0 < K < +\infty \),
\[ \int \log |x - y| d\nu(y) \]
\[ = \begin{cases} \frac{K}{2} - \sum_{n=1}^\infty \frac{2}{n} T_n\left(\frac{x}{K}\right) T_n\left(\frac{y}{K}\right) d\nu(y) & \text{if } |x| \leq K, \\ \frac{x^2}{2} - \sum_{n=1}^\infty \frac{2}{n} \left(\frac{x - \sqrt{x^2 - K^2}}{K}\right)^n T_n\left(\frac{y}{K}\right) d\nu(y) & \text{if } x > K, \\ \frac{-x^2}{2} - \sum_{n=1}^\infty \frac{2}{n} \left(\frac{-x - \sqrt{x^2 - K^2}}{K}\right)^n T_n\left(-\frac{y}{K}\right) d\nu(y) & \text{if } x < -K, \end{cases} \]
and
\[ \int \log |x - y| d\nu(x) d\nu(y) = \int \frac{K}{2} - \sum_{n=1}^\infty \frac{2}{n} \left(\int \frac{y}{K} T_n\left(\frac{y}{K}\right) d\nu(y)\right)^2, \]
where \( T_n \) are the Chebyshev polynomials of the first kind.

So, if the sequence \( \{ T_n(y/K) d\nu(y) \} \) is in \( l^1(1/n) \), then \( \log |x - y| d\nu(y) \) is a continuous function on \( R \). Also, \( \{ T_n(y/K) d\nu(y) \} \) is in \( l^2(1/n) \) is a necessary and sufficient condition for \( \Sigma(\nu) > -\infty \). Here \( l^1(1/n) \) and \( l^2(1/n) \) are the \( l^1 \) and \( l^2 \) spaces with respect to the sequence \( (1/n) \). Obviously, the continuity of \( Q(\nu) \) implies the continuity of \( \Sigma(\nu) > -\infty \).

Let \( \nu \in \mathcal{M}_0(R) \) and assume that \( Q(\nu) \) is continuous on \( R \). We focus on the case \( Q(x) = 2Q(\nu)(x) \) in Theorem 3.2 so that the distribution \( (5) \) is
\[ \hat{A}_n(\nu; K) = \frac{1}{Z_n(K)} \exp \left( -2n \sum_{i=1}^n \log |x_i - y| d\nu(y) \right) \prod_{i < j} |x_i - x_j|^2 \prod_{i=1}^n \lambda_{1-K}(x_i) dx_1 dx_2 \cdots dx_n. \]

Also, we denote by the same symbol \( A_n(\nu; K) \) the induced measure on \( M_n(C)^{\text{diag}} \) under the bijection between \( M_n(C)^{\text{diag}} \) and \( R^n \). Now we define the unitary conjugation invariant measure \( A_n(\nu; K) \) on \( M_n(C)^w \) by
\[ A_n(\nu; K)(\cdot) = \int A_n(\nu; K)(U \cdot U^*) dU, \]
where \( U \) is in the \( n \)-dimensional unitary group and \( dU \) is the Haar probability measure. If supp \( \nu \subset [-K, K] \), then the random matrix \( \{ T_n(\nu(K)) \} \) almost surely converges to \( \nu \) in the distribution sense. In fact, the random matrix \( T_n(\nu) \) has the joint eigenvalue distribution \( \hat{A}_n(\nu; K) \), and thanks to Theorem 3.2 the almost sure limit distribution of the random matrix is the unique minimizer for the functional \( I(\mu) = -\Sigma(\mu) + 2 |Q(\nu) d\mu(x) + B(K) \). This minimizer is \( \nu \) itself due to the characterization for the minimizer in Theorem 3.1.

When \( a_i \in I \) are noncommutative random variables in free relation and \( v_i \) is the distribution of \( a_i \) for \( i \in I \), we denote by \( \bigstar_{i \in I} v_i \) the joint distribution of \( a_i \) for \( i \in I \). Note that \( \bigstar_{i \in I} v_i \) is the distribution in the sense of joint moments (there is no meaning as a measure) and it is determined by \( v_i \)'s only, independently of the choice of \( a_i \)'s. Similarly to Proposition 2.3 we state the following properties of the measures \( A_n(\nu; K) \).

**Proposition 3.4** (1) Let \( v_i \in \mathcal{M}_0(R) \) for \( i \in I \) be such that \( Q_i(\nu) \) is continuous for each \( i \in I \) and supp \( v_i \subset [-K, K] \) for some \( 0 < K < +\infty \). For each \( n \in \mathbb{N} \), let \( T_{n,i}(\nu; K) \) be \( n \times n \) selfadjoint random matrices on \( (\Omega, \mathcal{F}, \mathbb{P}) \) having the distribution \( (M_n(C)^{\text{diag}}, A_n(\nu; K)) \). Assume that, for each \( n \in \mathbb{N} \), \( T_{n,i}(\nu; K) \) are independent, that is, the distribution of the random matrices \( T_{n,i}(\nu; K) \) is \( \bigotimes_{i \in I} A_n(\nu; K) \) on \( (M_n(C)^{\text{diag}})^I \). Then the random matrices \( T_{n,i}(\nu; K) \) almost surely converge as \( n \to \infty \) to the distribution \( \bigstar_{i \in I} v_i \) in the distribution sense, that is, any joint moment of \( T_{n,i}(\nu; K) \) with respect to \( v_i \) converges to that of \( \bigstar_{i \in I} v_i \).

(2) Let \( T \) be an \( n \times n \) selfadjoint random matrix. If the mean \( T(\nu) = 0 \) and the pth absolute moment \( \tau_p(\nu)(T^p) \) is fixed for some \( 0 < p < +\infty \), then the random matrix \( T = (M_n(C)^{\text{diag}}, A_n(Q(\nu); K)) \) with \( Q(\nu) = c \lambda^p \) and \( K = +\infty \) has the maximal Boltzmann-Gibbs entropy. Here \( c \) is a constant depending on \( \tau_p(\nu)(T^p) \) and the Boltzmann-Gibbs entropy of selfadjoint random matrices \( (M_n(C)^{\text{diag}}, v_\lambda) \) is defined as an \( n^2 \)-variable random.
vector under the identification between $M_n(C)^w$ and $R^n$.

We notice that the reference measures appearing in the above (2) are the most uniformly distributed on the selfadjoint matrices under a certain constraint. The property (1) is a consequence of the almost sure asymptotic freeness for unitary conjugation invariant random matrices presented in [8]. Also see [8, Sect. 5.2] for (2).

Now it may be natural to define the free relative entropy in the microstate approach as follows.

**Definition 3.5** Let $\mu, \nu \in \mathcal{M}_n(R)$ and assume that $Q_{\nu}(x)$ is continuous. For $r \in N$ and $\varepsilon, R > 0$ define

$$
\Sigma_{R}(\mu|\nu) = \lim_{n \to \infty} \inf \left[ -\frac{1}{n^2} \log A_n(\nu; R) (F_R(\mu; n, r, \varepsilon)) \right],
$$

$$
\Sigma(\mu|\nu) = \inf_{R > 0} \Sigma_{R}(\mu|\nu).
$$

Then we call $\Sigma(\mu|\nu)$ the free relative entropy of $\mu$ with respect to $\nu$.

We need the logarithmic energy of a signed measure to show that the above defined microstate free relative entropy has a simple integral representation. Let $\nu$ be a finite signed measure with compact support. The logarithmic energy of $\nu$ familiar in potential theory ([9]) is defined by

**Definition (A)**

$$
E(\nu) = \left\{ \begin{array}{ll}
-\int \int \log|x-y|d\nu(x)d\nu(y) & \text{if } \int \int \log|x-y||d\nu(x)||d\nu(y) < +\infty, \\
+\infty & \text{otherwise.}
\end{array} \right.
$$

If the free entropies $\Sigma(\mu)$ and $\Sigma(\nu)$ are finite, then

$$
\int \int \log|x-y||d\mu - \nu(x)d\nu(y) < +\infty.
$$

Although the above definition of the logarithmic energy of a signed measure is natural, we will give another definition based on the next lemma. For a finite signed measure $\nu$ with compact support, we define

$$
E_\varepsilon(\nu) = -\int \int \log(\varepsilon + |x-y|)d\nu(x)d\nu(y).
$$

Since $\log(\varepsilon + |x-y|)$ is continuous on $R^2$, $E_\varepsilon(\nu)$ is well defined for any $\varepsilon > 0$.

**Lemma 3.6** Let $\nu$ be a finite signed measure with compact support. Then $E_\varepsilon(\nu)$ is monotone increasing as $\varepsilon \to +0$.

**Proof.** It suffices to prove the positive definiteness of $-\log(\varepsilon' + |x-y|) + \log(\varepsilon + |x-y|)$ for $\varepsilon > \varepsilon' > 0$. We have

$$
-\log(\varepsilon' + |x-y|) + \log(\varepsilon + |x-y|) = \log \left( 1 + \frac{\varepsilon - \varepsilon'}{\varepsilon' + |x-y|} \right) = \int_0^\infty \left( \frac{1}{1 + t} - \frac{\varepsilon - \varepsilon'}{\varepsilon' + |x-y|} \right) dt,
$$

and

$$
\frac{1}{1 + \varepsilon'} - \frac{1}{1 + \varepsilon' + |x-y|} = \frac{1}{(1 + \varepsilon)^2} \left( \frac{\varepsilon - \varepsilon'}{1 + \varepsilon' + |x-y|} \right).
$$

Since $l + |x-y|^{-1}$ is a positive definite kernel for any $l > 0$, so is $-\log(\varepsilon' + |x-y|) + \log(\varepsilon + |x-y|)$ due to $\varepsilon - \varepsilon' > 0$.

From this lemma we could also define logarithmic energy of a finite signed measure $\nu$ with compact support as

**Definition (B)**

$$
E(\nu) = \lim_{\varepsilon \to +0} E_\varepsilon(\nu).
$$

Note that if $\int \int \log|x-y||d\nu(x)||d\nu(y) < +\infty$, then the definitions (A) and (B) are equivalent. Although it is not known whether $\lim_{\varepsilon \to +0} E_\varepsilon(\nu)$ is infinite or not in the case when $\int \int \log|x-y||d\nu(x)||d\nu(y) = +\infty$, the next proposition is true in either case of the definitions (A) or (B). In fact, the same proof as in the case (A) in [8, Lemma 5.3.1] works also in the case (B).

**Proposition 3.7** Let $\nu$ be a compactly supported signed measure. If $\nu(R) = 0$, then $E(\nu) \geq 0$. Furthermore, $E(\nu) = 0$ if only if $\nu = 0$. 

Now we can give the integral representation for the microstate free relative entropy. D. Petz suggested to us that the free relative entropy would be the double logarithmic integral $E(\mu - v)$ in (7).

**Theorem 3.8** Let $\mu, v \in \mathcal{M}(\mathbb{R})$ and assume that $Q_\nu(x)$ is continuous. Then for any $R > 0$ such that supp $v$, supp $\mu \subset [-R, R]$,

$$\Sigma_R(\mu \mid v; r, \varepsilon) = \lim_{n \to \infty} \left[-\frac{1}{n^2} \log A_n(\nu; R)(I_n(\mu; n, r, \varepsilon))\right],$$

$$\Sigma(\mu \mid v) = \Sigma_R(\mu \mid v).$$

Furthermore,

$$\Sigma(\mu \mid v) = E(\mu - v) \tag{7}$$

in either case of the definitions (A) or (B) for the right-hand side.

**Proof.** We first prove the existence of lim in (6). Assume that supp $v$, supp $\mu \subset S = [-R, R]$ for $R > 0$. For $r \in \mathbb{N}$ and $\varepsilon > 0$ put

$$F(r, \varepsilon) = \{\kappa \in \mathcal{M}(S) : m_k(\kappa) - m_k(\mu) \leq e, k \leq r\},$$

$$G(r, \varepsilon) = \{\kappa \in \mathcal{M}(S) : m_k(\kappa) - m_k(\mu) < e, k \leq r\},$$

which are closed and open in $\mathcal{M}(S)$, respectively. Theorem 3.2 says that

$$\lim_{n \to \infty} \sup_{F(r, \varepsilon)} \left[\frac{1}{n^2} \log A_n(v; R)(I_n(\mu; n, r, \varepsilon))\right] = \inf \left\{ I(\mu) : \mu \in F(r, \varepsilon) \right\},$$

$$\lim_{n \to \infty} \inf_{G(r, \varepsilon)} \left[\frac{1}{n^2} \log A_n(v; R)(I_n(\mu; n, r, \varepsilon))\right] = \inf \left\{ I(\mu) : \mu \in G(r, \varepsilon) \right\},$$

where $\kappa = (1/n)(\delta(x_1) + \delta(x_2) + \cdots + \delta(x_n))$ for $x = (x_1, \ldots, x_n) \in S^*$ and $I(\mu)$ is the rate function given in Theorem 3.2. If $\inf \{ I(\mu) : \mu \in F(r, \varepsilon) \} = +\infty$ for some $r = r_1$ and $\varepsilon = \varepsilon_1$, then

$$\lim_{n \to \infty} \left[\frac{1}{n^2} \log A_n(v; R)(I_n(\mu; n, r, \varepsilon))\right] = -\inf \{ I(\mu) : \mu \in F(r, \varepsilon) \} = -\infty$$

for every $r \geq r_1$ and $\varepsilon \leq \varepsilon_1$. So we may assume that $\inf \{ I(\mu) : \mu \in F(r, \varepsilon) \} < +\infty$ for all $r \in \mathbb{N}$ and $\varepsilon > 0$. Then by the convexity of $I(\mu)$ we notice that

$$\inf \{ I(\mu) : \mu \in F(r, \varepsilon) \} = \inf \{ I(\mu) : \mu \in G(r, \varepsilon) \}.$$

Hence we get

$$\lim_{n \to \infty} \left[\frac{1}{n^2} \log A_n(v; R)(I_n(\mu; n, r, \varepsilon))\right] = -\inf \{ I(\mu) : \mu \in F(r, \varepsilon) \} = -\Sigma_R(\mu \mid v; r, \varepsilon).$$

So (6) is obtained. Note that the sets $F(r, \varepsilon)$ ($r \in \mathbb{N}, \varepsilon > 0$) form a neighborhood basis of $\mu$ in $\mathcal{M}(S)$. Thanks to the lower semicontinuity of $I(\mu)$ we have

$$\inf \{ I(\mu) : \mu \in F(r, \varepsilon) \} = -\Sigma_R(\mu \mid v; r, \varepsilon).$$

Since $I(\nu) = 0$, we get $B(R) = -\Sigma(\nu)$ for the constant term of $I(\mu) = -\Sigma(\mu) + 2 \int Q_\nu(x) d\mu(x) + B(R)$. Hence $\Sigma_R(\mu \mid v)$ is independent of $R > 0$ whenever supp $\mu$, supp $v \subset [-R, R]$, so $\Sigma(\mu \mid v) = \Sigma_R(\mu \mid v) = I(\mu)$.

Now it remains to show that

$$-\Sigma(\mu) + 2 \int Q_\nu(x) d\mu(x) = E(\mu - v). \tag{8}$$

Note that $\Sigma(\nu)$ and $\int Q_\nu(x) d\mu(x)$ are finite because of the continuity assumption of $Q_\nu(x)$. Moreover, for any $\varepsilon > 0$ it is clear that

$$-\Sigma(\mu) + 2 \int \log(\varepsilon + |x - y|) d\nu(y) d\mu(x) - \Sigma(\nu) = E(\mu - v). \tag{9}$$

In the case (B) we obtain (8) by taking the limits of both sides of (9) as $\varepsilon \to +0$. In the case (A), if $\Sigma(\mu) > -\infty$, then $\int \log(\varepsilon + |x - y|) d\mu - \varepsilon \int |x - y| d\mu - \varepsilon \int (y) < +\infty$ and we can let $\varepsilon \to +0$ again to get (8). If $\Sigma(\mu) = -\infty$, then the left-hand side of (8) is $+\infty$ and also $E(\mu - v) = +\infty$. \qed
4 Properties of free relative entropy

In Section 3 we defined the microstate free relative entropy with respect to \( \nu \) having the continuous potential \( Q_\nu(x) \) and its integral representation was also shown. In view of Theorem 3.8 we consider the integral form \( E(\mu - \nu) \) as the free relative entropy regardless of the continuity assumption of \( Q_\nu(x) \). Thus, in this section, we will adopt the notation \( \Sigma(\mu \mid \nu) = E(\mu - \nu) \) for \( \mu, \nu \in \mathcal{M}_\nu(\mathbb{R}) \) and examine properties of \( E(\mu - \nu) \) in either case of the definitions (A) or (B). They are summarized in the following. The free relative entropy differs from the classical one in the first property. But the other important properties are common; thus our use of the term "relative entropy" would be justified.

**Proposition 4.1** Let \( \mu, \nu \in \mathcal{M}_\nu(\mathbb{R}) \).

1. **Symmetry:** \( \Sigma(\mu \mid \nu) = \Sigma(\nu \mid \mu) \).
2. **Strict positivity:** \( \Sigma(\mu \mid \nu) \geq 0 \), and \( \Sigma(\mu \mid \nu) = 0 \) if and only if \( \mu = \nu \).
3. **Joint convexity:** If \( \Sigma(\mu_i) > -\infty \) and \( \Sigma(\nu_i) > -\infty \) \( (i = 1, 2) \), then
   \[
   \Sigma(\alpha \mu_1 + (1 - \alpha) \mu_2 \mid \alpha \nu_1 + (1 - \alpha) \nu_2) \leq \alpha \Sigma(\mu_1 \mid \nu_1) + (1 - \alpha) \Sigma(\mu_2 \mid \nu_2)
   \]
   (10)
   for \( 0 < \alpha < 1 \). Furthermore, in the case (B), (10) holds without the conditions \( \Sigma(\mu_i), \Sigma(\nu_i) > -\infty \) \( (i = 1, 2) \).
4. **Single strict convexity:** If \( Q_\nu(x) \) is continuous, then
   \[
   \Sigma(\alpha \mu_1 + (1 - \alpha) \mu_2 \mid \alpha \nu_1 + (1 - \alpha) \nu_2) \leq \alpha \Sigma(\mu_1 \mid \nu_1) + (1 - \alpha) \Sigma(\mu_2 \mid \nu_2)
   \]
   (11)
   for \( 0 < \alpha < 1 \). If \( \Sigma(\mu_1 \mid \nu_1) = +\infty \) \( (i = 1, 2) \) and \( \mu_1 \neq \mu_2 \), (11) can be replaced by strict inequality. Furthermore, in the case (B), (11) holds without the continuity of \( Q_\nu(x) \).
5. **Joint lower semicontinuity:** Let \( S \) be any compact subset of \( \mathbb{R} \) and \( \mathcal{M}_\nu(\mathbb{R}) \equiv \{ \mu \in \mathcal{M}_\nu(\mathbb{R}) : \Sigma(\mu) > -\infty \} \). Then \( \Sigma(\mu \mid \nu) \) is weakly jointly lower semicontinuous on \( \mathcal{M}(S) \cap \mathcal{M}_\nu(\mathbb{R}) \). Furthermore, in the case (B), it is weakly jointly lower semicontinuous on \( \mathcal{M}(S) \).
6. **Single lower semicontinuity:** Let \( S \) be any compact subset of \( \mathbb{R} \). If \( Q_\nu(x) \) is continuous, then \( \Sigma(\mu \mid \nu) \) is weakly lower semicontinuous in \( \mu \) on \( \mathcal{M}(S) \).

**Proof.** The property 1. is trivial and 2. is due to Proposition 3.7.

3. **Case (A):** For \( 0 < \alpha < 1 \) and \( \varepsilon > 0 \) we get
   \[
   \Sigma(\alpha \mu_1 + (1 - \alpha) \mu_2 \mid \alpha \nu_1 + (1 - \alpha) \nu_2) = \alpha^2 \Sigma(\mu_1 \mid \nu_1) + 2\alpha(1 - \alpha)\Sigma(\mu_1 - \nu_1, \mu_2 - \nu_2) + (1 - \alpha)^2 \Sigma(\mu_2 \mid \nu_2),
   \]
   (12)
   where
   \[
   \Sigma(\mu_1 \mid \nu_1) = -\iint \log(e + |x - y|)d(\mu_1 - \nu_1)(x)d(\mu_1 - \nu_1)(y),
   \]
   \[
   \Sigma(\mu_1 - \nu_1, \mu_2 - \nu_2) = -\iint \log(e + |x - y|)d(\mu_1 - \nu_1)(x)d(\mu_2 - \nu_2)(y),
   \]
   and similarly for \( \Sigma(\alpha \mu_1 + (1 - \alpha) \mu_2 \mid \alpha \nu_1 + (1 - \alpha) \nu_2) \) and \( \Sigma(\mu_2 \mid \nu_2) \). As mentioned after the definition (A), \( \Sigma(\mu_1 \mid \nu_1) < +\infty \) \( (i = 1, 2) \) and similarly
   \[
   \Sigma(\mu_1 - \nu_1, \mu_2 - \nu_2) = -\iint \log|x - y|d(\mu_1 - \nu_1)(x)d(\mu_2 - \nu_2)(y)
   \]
   (13)
   is finite. Also, the concavity of free entropy yields
   \[
   \Sigma(\alpha \mu_1 + (1 - \alpha) \mu_2) \geq \alpha \Sigma(\mu_1) + (1 - \alpha) \Sigma(\mu_2) > -\infty,
   \]
   \[
   \Sigma(\alpha \nu_1 + (1 - \alpha) \nu_2) \geq \alpha \Sigma(\nu_1) + (1 - \alpha) \Sigma(\nu_2) > -\infty,
   \]
   and therefore \( \Sigma(\alpha \mu_1 + (1 - \alpha) \mu_2 \mid \alpha \nu_1 + (1 - \alpha) \nu_2) < +\infty \). Thanks to what was mentioned after the definition (B), letting \( \varepsilon \to 0 + 0 \) in (12) implies that
   \[
   \Sigma(\alpha \mu_1 + (1 - \alpha) \mu_2 \mid \alpha \nu_1 + (1 - \alpha) \nu_2) = \alpha^2 \Sigma(\mu_1 \mid \nu_1) + 2\alpha(1 - \alpha)\Sigma(\mu_1 - \nu_1, \mu_2 - \nu_2) + (1 - \alpha)^2 \Sigma(\mu_2 \mid \nu_2),
   \]
   (14)
   and so
   \[
   \frac{d}{d\alpha^2} \Sigma(\alpha \mu_1 + (1 - \alpha) \mu_2 \mid \alpha \nu_1 + (1 - \alpha) \nu_2) = 2\Sigma(\mu_1 \mid \nu_1) - 4\Sigma(\mu_1 - \nu_1, \mu_2 - \nu_2) + 2\Sigma(\mu_2 \mid \nu_2)
   \]
   \[
   = 2E((\mu_1 - \nu_1, (\mu_2 - \nu_2) \geq 0.
   \]

**Proof.**
The last equality can be easily shown as above.  

Case (B): We may assume that \( \Sigma(\mu_i|v_i) < +\infty \) (\( i = 1, 2 \)). From the fact that \( \log(e + |x - y|) \) is a negative definite kernel (see the proof of [8, Lemma 5.3.1]), we notice that \( \Sigma(\kappa_1, \kappa_2) = -\int \log(e + |x - y|)d\kappa_1(x)d\kappa_2(y) \) is a positive bilinear form on the real linear space of compact supported signed measures \( \kappa \) on \( \mathbb{R} \) with \( \kappa(\mathbb{R}) = 0 \). So the Cauchy-Schwarz inequality yields 

\[
|E_v(\mu_i - v_1, \mu_2 - v_2)| \leq \Sigma(\mu_i|v_1)\Sigma(\mu_2|v_2), 
\]

and therefore by Lemma 3.6 we get 

\[
|E_v(\mu_1 - v_1, \mu_2 - v_2)| \leq \Sigma(\mu_1|v_1)^2\Sigma(\mu_2|v_2)^2 < +\infty. 
\]

Combining this and (12) together with Lemma 3.6 implies that all terms in both sides of (12) converge as \( \varepsilon \to +0 \), so we have (14) in the case (B) making use of 

\[
\Sigma(\mu_1 - v_1, \mu_2 - v_2) = \lim_{\varepsilon \to +0} \Sigma(\mu_i - v_i, \mu_2 - v_2). 
\]

(15)

The remaining proof is the same as in the case (A).

4. Case (A): If \( Q_\varepsilon(x) \) is continuous, then \( \Sigma(\mu|v) \) is the rate function \( I(\mu) = -\Sigma(\mu) + 2 \int Q_\varepsilon(x)d\mu(x) - \Sigma(v) \) of the large deviation principle in Theorem 3.2. Hence the strict convexity of \( \mu \to \Sigma(\mu|v) \) follows from the strict concavity of \( \Sigma(\mu) \).

Case (B): We may just put \( v_1 = v_2 \) in the second part of the proof of 3. The strictness is easy to check from the proof of 3. and the strict positivity of free relative entropy.

5. Case (A): If \( \Sigma(\mu, \Sigma(v)) > -\infty \), then, by Proposition 3.3 for signed measures (17),

\[
\Sigma(\mu|v) = -\int |\log|x - y||d(\mu - v)(x)d(\mu - v)(y) \leq \Sigma(\mu_1|v_1)^2\Sigma(\mu_2|v_2)^2 < +\infty. 
\]

where \( \supp \mu, \supp v \subset [-K, K] \). Hence \( \Sigma(\mu|v) \) is weakly jointly semicontinuous on \( \mathcal{M}(\mathcal{S}) \cap \mathcal{M}(\mathcal{R}) \).

Case (B): By Lemma 3.6 we get \( \Sigma(\mu|v) = \lim_{\varepsilon \to +0} \Sigma(\mu_i|v) \), which is the monotone increasing limit of weakly continuous functions \( \Sigma(\mu_i|v) \).

6. This follows from \( \Sigma(\mu|v) = I(\mu) \) as stated in the proof of 4.

Finally in this section, we give a remark on the free Fisher information matrix from the viewpoint of information geometry. Let \( V = \text{conv}(\mu_1, \mu_2, \ldots, \mu_{n+1}) \) be the convex hull of \( \mu_i \) (\( i = 1, 2, \ldots, n + 1 \)) considered as a manifold with boundary. The bijection between the \( n \)-dimensional simplex \( \{(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n: \alpha_i \geq 0 (1 \leq i \leq n), \alpha_1 + \alpha_2 + \cdots + \alpha_n \leq 1 \} \) and \( V \) provides a standard coordinate system \( \alpha \) of \( V \). Assume \( \Sigma(\mu_i) < +\infty \) (1 \( \leq i \leq n + 1 \)) in the case (A) or \( \Sigma(\mu_i|\mu_{n+1}) < +\infty \) (1 \( \leq i \leq n \)) in the case (B). In both cases we have 

\[
\Sigma(\alpha_1\mu_1 + \cdots + \alpha_n\mu_n + \left(1 - \sum_{i=1}^n \alpha_i\right)\mu_{n+1}) = \sum_{i,j=1}^n (\alpha_i - \alpha_j)^2 \Sigma(\mu_i|\mu_{n+1}) + \sum_{i,j=1}^n 2(\alpha_i - \alpha_j)(\alpha_i - \alpha_j) \Sigma(\mu_i - \mu_{n+1}, \mu_j - \mu_{n+1}), 
\]

with the notation (13) or (15). Hence we get 

\[
\frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \Sigma(\alpha_1\mu_1 + \cdots + \alpha_n\mu_n + \left(1 - \sum_{i=1}^n \alpha_i\right)\mu_{n+1}) = 2\Sigma(\mu_i|\mu_{n+1}) \quad \text{if} \quad i = j, 
\]

\[
2E(\mu_i - \mu_{n+1}, \mu_j - \mu_{n+1}) \quad \text{otherwise}. 
\]

Note that the Hessian of a convex function \( F(\alpha) \) at the point of its minimizer \( \alpha_0 \) becomes a tensor. In fact, if \( \alpha_i = \alpha_i(\alpha_j) \) is a coordinate change, then 

\[
\frac{\partial^2 F}{\partial \alpha_i \partial \alpha_j} \bigg|_{\alpha = \alpha_0} = \frac{\partial}{\partial \alpha_i} \left( \frac{\partial F}{\partial \alpha_j} \bigg|_{\alpha = \alpha_0} \right) = \frac{\partial}{\partial \alpha_i} \left( \frac{\partial F}{\partial \alpha_k} \bigg|_{\alpha = \alpha_0} \right) + \frac{\partial}{\partial \alpha_k} \frac{\partial F}{\partial \alpha_i} \bigg|_{\alpha = \alpha_0} 
\]

So the following matrix \( I_F = [I_F]_{ij} \) may be called the free Fisher information matrix with respect to the coor-
dinate $\alpha$:

$$I_{P_i} = \begin{cases} \frac{2\Sigma(\mu_i \mu_{i + 1})}{2\Sigma(\mu_i - \mu_{i + 1})} & \text{if } i = j, \\
\frac{2\Sigma(\mu_i - \mu_{i + 1}, \mu_j - \mu_{j + 1})}{2\Sigma(\mu_i - \mu_{i + 1})} & \text{otherwise.} \end{cases}$$

When we regard $I_P$ as a Riemannian metric tensor on the manifold $V$, this manifold is just the flat Euclidean space. This is very different from information geometry of classical probability theory (see [1]).

5 Multiple free relative entropy

For $\mu, \nu \in \mathcal{M}(\mathbb{R}^n)$, as in the single variable case, the $N$-variable classical relative entropy is defined as

$$S(\mu \mid \nu) = \left\{ \begin{array}{ll} \int \frac{d\mu}{d\nu} \log \frac{d\mu}{d\nu} & \text{if } \mu \ll \nu, \\
+\infty & \text{otherwise.} \end{array} \right.$$ When $\mu \in \mathcal{M}_0(\mathbb{R}^n)$, the microstate representation of the multi-variable classical relative entropy is obtained as in the single variable case by the multi-variable Sanov theorem or more general Sanov theorem for complete separable metric space (see [6]). For $n, r \in \mathbb{N}$ and $\varepsilon, R > 0$,

$$I_\kappa(\mu; n, r, \varepsilon) = \left\{ (A_1, \ldots, A_N \in (M_\kappa(C)^{\otimes n})^N \mid A_i \ll \mathcal{H} \leq R, \right.$$ 

$$\left. \|m_{i_1 \cdots i_k} \| \leq \varepsilon \text{ for all } 1 \leq i_1, \ldots, i_k \leq N, 1 \leq k \leq r \right\},$$

where $m_{i_1 \cdots i_k} = \sum x_{i_1} \cdots x_{i_k} d\mu(x_{i_1} \cdots x_{i_k})$. We denote by $\Omega_n(\nu)$ the measure on $(M_\kappa(C)^{\otimes n})$ induced by the $n$-fold product measure $\nu^n$ on $(\mathbb{R}^n)^n$ under the isometric bijection $(\mathbb{R}^n)^n \equiv (\mathbb{R}^n)^n \equiv (M_\kappa(C)^{\otimes n})$. Furthermore, this $\Omega_n(\nu)$ is also regarded as the measure on $(M_\kappa(C)^{\otimes n})$ under the natural inclusion $(M_\kappa(C)^{\otimes n}) \subseteq (M_\kappa(C)^{\otimes n})$ in the expression

$$S_\kappa(\mu \mid \nu; r, \varepsilon) = \lim_{n \to \infty} \inf \left\{ -\frac{1}{n} \log \Omega_n(\nu)(I_\kappa(\mu; n, r, \varepsilon)) \right\}.$$ If

$$S_\kappa(\mu \mid \nu) = \lim_{r \to \infty, \varepsilon \to 0} S_\kappa(\mu \mid \nu; r, \varepsilon) = \sup_{r \in \mathbb{N}, \varepsilon > 0} S_\kappa(\mu \mid \nu; r, \varepsilon),$$

then we have

$$S(\mu \mid \nu) = \inf_{R > 0} S_\kappa(\mu \mid \nu).$$

We have assumed that $\mu \in \mathcal{M}_0(\mathbb{R}^n)$ or $\mu$ is the distribution of commutative $N$ random variables, however the above expressions are meaningful even if $\mu$ is the distribution (as joint moments) of noncommutative random variables. For $a_1, \ldots, a_n \in (\mathcal{M}_\kappa, \tau)$ we define $S_\kappa(a_1, \ldots, a_n \mid \nu)$ by replacing $I_\kappa(\mu; n, r, \varepsilon)$ by $I_\kappa(a_1, \ldots, a_n; n, r, \varepsilon)$ (see Section 2) in (16). Then the classical relative entropy $S_\kappa(a_1, \ldots, a_n \mid \nu)$ is defined by

$$S(a_1, \ldots, a_n \mid \nu) = \inf_{R > 0} S_\kappa(a_1, \ldots, a_n \mid \nu),$$

even when $a_1, \ldots, a_n$ are noncommutative. For instance, if $a_1, \ldots, a_n$ are noncommutative with $\| \tau(a_1 a_2 a_3) - \tau(a_1 a_2 a_3) \| = \varepsilon > 0$, then since $\Omega_n(\nu)(I_\kappa(a_1, \ldots, a_n; n, 3, \varepsilon/3)) = 0$ for all $n \in \mathbb{N}$ we have $S(a_1, \ldots, a_n \mid \nu) = +\infty$.

Now let $\nu_1, \ldots, \nu_n \in \mathcal{M}_\kappa(\mathbb{R})$ such that $P_a(x) (1 \leq i \leq N)$ are continuous. Then the multiple free relative entropy with respect to $\nu_1 \otimes \nu_2 \otimes \cdots \otimes \nu_N$ is defined as follows.

**Definition 5.1** Let $a_1, \ldots, a_n \in (\mathcal{M}_\kappa, \tau)$. For $r \in \mathbb{N}$ and $\varepsilon, R > 0$ define

$$\Sigma_\kappa(a_1, \ldots, a_n; \nu_1 \otimes \cdots \otimes \nu_N; R, \varepsilon) = \lim_{n \to \infty} \inf \left\{ -\frac{1}{n} \log \otimes_{i=1}^n A_{a_i}(\nu_i; R)(I_\kappa(a_1, \ldots, a_n; n, r, \varepsilon)) \right\},$$

$$\Sigma_\kappa(a_1, \ldots, a_n; \nu_1 \otimes \cdots \otimes \nu_N) = \lim_{\varepsilon \to 0} \lim_{r \to \infty, \varepsilon \to 0} \Sigma_\kappa(a_1, \ldots, a_n; \nu_1 \otimes \cdots \otimes \nu_N; r, \varepsilon)$$

$$= \sup_{r \in \mathbb{N}, \varepsilon > 0} \inf_{\nu_1, \nu_2, \ldots, \nu_N} \Sigma_\kappa(a_1, \ldots, a_n; \nu_1 \otimes \cdots \otimes \nu_N; r, \varepsilon).$$

The multiple free relative entropy $\Sigma_\kappa(a_1, \ldots, a_n; \nu_1 \otimes \cdots \otimes \nu_N)$ with respect to $\nu_1 \otimes \nu_2 \otimes \cdots \otimes \nu_N$ is

$$\Sigma(a_1, \ldots, a_n; \nu_1 \otimes \cdots \otimes \nu_N) = \inf_{R > 0} \Sigma_\kappa(a_1, \ldots, a_n; \nu_1 \otimes \cdots \otimes \nu_N).$$
The properties of $\Sigma(a_1, \ldots, a_N | v_1 \cdots \cdots v_N)$ in the next proposition are similar to those of the free entropy $\chi(a_1, \ldots, a_N)$ mentioned in Section 1.

**Proposition 5.2** Let $a_1, \ldots, a_N \in (\mathcal{M}, \tau)$.

1. **Superadditivity:** 
   \[
   \Sigma(a_1, \ldots, a_N | v_1 \cdots \cdots v_N) \geq \Sigma(a_1 | v_1) + \cdots + \Sigma(a_N | v_N) \geq 0.
   \]

2. **Lower semicontinuity:** Let $a_{m,1}, \ldots, a_{m,N} \in (\mathcal{M}, \tau) (m \in \mathbb{N})$. If $(a_{m,1}, \ldots, a_{m,N}) \to (a_1, \ldots, a_N)$ in the distribution sense as $m \to \infty$ and $\sup m \|a_m\| < +\infty (1 \leq i \leq N)$, then 
   \[
   \Sigma(a_1, \ldots, a_N | v_1 \cdots \cdots v_N) \leq \lim\inf m \to \infty \Sigma(a_{m,1}, \ldots, a_{m,N} | v_1 \cdots \cdots v_N).
   \]

3. **Strong additivity:** If $a_1$ and $\{a_2, \ldots, a_N\}$ are in free relation, then 
   \[
   \Sigma(a_1, a_2, \ldots, a_N | v_1 \cdots \cdots v_N) = \Sigma(a_1 | v_1) + \Sigma(a_2, \ldots, a_N | v_2 \cdots v_N).
   \]

4. **Additivity:** If $a_1, \ldots, a_N$ are in free relation, then 
   \[
   \Sigma(a_1, \ldots, a_N | v_1 \cdots \cdots v_N) = \Sigma(a_1 | v_1) + \cdots + \Sigma(a_N | v_N).
   \]

**Proof.** The proofs for the free entropy in [8] equally work for the free relative one.

1. The proof of the superadditivity is the same as the subadditivity of free entropy, which is a direct consequence of the inclusion 
   \[
   \Gamma_R(a_1, \ldots, a_N; n; r, e) \subseteq \Gamma_R(a_1; n, r, e) \otimes \cdots \otimes \Gamma_R(a_N; n, r, e).
   \]

2. The proof of the lower semicontinuity is the same when we observe the fact that if $R > \|a_i\| (1 \leq i \leq N)$ then 
   \[
   \Sigma_R(a_1, \ldots, a_N | v_1 \cdots \cdots v_N) = \Sigma(a_1, \ldots, a_N | v_1 \cdots v_N).
   \]

But the proof of this fact is a slight modification of that of [8, Proposition 6.1.4].

3. The strong additivity is proved as that of free entropy by the method of approximate freeness (see the proof of [8, Theorem 6.4.1]). We essentially need here are the unitary invariance of the reference measure \(\otimes_{i=1}^{N} A_a(v_i; R)\) and the fact that \(\limsup\) could be replaced by \(\lim\) in the definition of the single variable free relative entropy (given in (6)).

4. Immediate from 3. \(\Box\)

**Acknowledgments**

The author is grateful to Professor Fumio Hiai for his advice and encouragement.

**REFERENCES**