J. Math. Soc. Japan
Vol. 54, No. 2, 2002

Free relative entropy for measures and a corresponding perturbation theory

By Fumio HIAI,[1] Masaru MIZUO and Dénes PETZ[2]

(Received Jan. 24, 2001)

Abstract. Voiculescu's single variable free entropy is generalized in two different ways to the free relative entropy for compactly supported probability measures on the real line. The one is introduced by the integral expression and the other is based on matricial (or microstates) approximation; their equivalence is shown based on a large deviation result for the empirical eigenvalue distribution of a relevant random matrix. Next, the perturbation theory for compactly supported probability measures via free relative entropy is developed on the analogy of the perturbation theory via relative entropy. When the perturbed measure via relative entropy is suitably arranged on the space of selfadjoint matrices and the matrix size goes to infinity, it is proven that the perturbation via relative entropy on the matrix space approaches asymptotically to that via free relative entropy. The whole theory can be adapted to probability measures on the unit circle.

Introduction.

One of the key points in free probability theory is that the important (noncommutative) distributions admit convenient matrix models. Let a be a noncommutative random variable in a noncommutative probability space $(\mathscr{A}, \varphi)$, and let an $n \times n$ random matrix X_n be given for every $n \in \boldsymbol{N}$. Then X_n is called an (almost sure) random matrix model for a if

$$\frac{1}{n}\sum_{i=1}^{n} P(X_n, X_n^*)_{ii} \to \varphi(P(a, a^*)) \text{ almost surely}$$

as $n \to \infty$ for any polynomial P of two noncommutative indeterminates. For example, selfadjoint random matrices with independent Gaussian entries form a model for the semicircular distribution when the variances of entries are suitably arranged; non-selfadjoint random matrices with independent Gaussian entries form a model for the circular distribution.

Matrix models play a crucial role in the definition of entropy which goes on the following lines. Suppose that the entropy of a variable a should be defined

2000 *Mathematics Subject Classification.* Primary 46L54; Secondary 94A17.
Key Words and Phrases. free entropy, free relative entropy, random matrix, large deviation.
[1] Supported by Grant-in-Aid for Exploratory Research 12874019.
[2] Supported by OTKA T032662.

with respect to another variable b which has a convenient random matrix model T_n. If λ_n is the distribution measure of T_n (on the space M_n of $n \times n$ matrices), then the asymptotics of the quantity

$$\frac{1}{n^2} \log \lambda_n(\{X \in M_n : |\tau(X^k) - \varphi(a^k)| \le \varepsilon, k \le r\}) \tag{0.1}$$

gives the entropy $\Sigma(a\|T_n)$; one takes

$$\lim_{\varepsilon \to +0} \lim_{r \to \infty} \limsup_{n \to \infty}$$

of the above expression. A general theory is not available for an arbitrary random matrix model, but Voiculescu's definition [**19**], [**20**] for the free entropy can be put in this setting and later in [**11**] Hiai and Petz considered the case when T_n is a Haar distributed unitary. Restricting ourselves to the case when the distribution of a is given by a compactly supported measure μ on $\boldsymbol{R}$, we have

$$\Sigma(\mu\|H_n) = \frac{1}{2} \iint_{\boldsymbol{R}^2} \log|x-y| \, d\mu(x) \, d\mu(y) - \frac{1}{4} \int x^2 \, d\mu(x) + \frac{3}{4}, \tag{0.2}$$

where H_n is the selfadjoint Gaussian random matrix model of the semicircular distribution. For a measure μ on the unit circle we also have

$$\Sigma(\mu\|U_n) = \iint_{\boldsymbol{T}^2} \log|x-y| \, d\mu(x) \, d\mu(y), \tag{0.3}$$

where U_n is a Haar distributed random unitary. In fact, the above formulas (0.2) and (0.3) are derived from the large deviation results for random matrices due to Ben Arous and Guionnet [**2**] and also [**12**].

The first aim of the present paper is to make a definition of the free relative entropy $\Sigma(\mu, \nu)$ of two measures. If we want to reach such a quantity in the above manner, then a random matrix model Q_n^ν of the measure ν should be find in order to proceed with (0.1). We make a proposal for Q_n^ν and find that

$$\Sigma(\mu\|Q_n^\nu) = -\iint \log|x-y| \, d(\mu-\nu)(x) \, d(\mu-\nu)(y).$$

This will be our definition of $\Sigma(\mu, \nu)$, the free relative entropy of μ with respect to ν. In fact, $\Sigma(\mu, \nu)$ is symmetric in the two variables and was investigated already in [**14**]. Note that Biane and Speicher [**3**] introduced the notion of free relative entropy (also free Fisher information) with respect to a function F while ours are defined with respect to a measure ν. After the discussion of the properties of the free relative entropy $\Sigma(\mu, \nu)$, we move to state perturbation.

Let us recall [**15**] that in the setting of operator algebras the relative entropy $S(\varphi, \omega)$ is defined for states φ and ω, and for a selfadjoint operator h the minimizer of the functional

$$S(\varphi, \omega) + \varphi(h)$$

is called the inner perturbation of the state ω; the notation ω^h is used. Following this pattern the minimizer of

$$\Sigma(\mu, \nu) + \int f(x)\, d\mu(x)$$

will be called the perturbed measure. In Sections 2 and 3 of the paper, this perturbation procedure is studied following the pattern of the state perturbation procedure in operator algebras. Roughly speaking several results are analogous, however in the perturbation theory via the free relative entropy there are slight differences.

Voiculescu originally introduced the free entropy $\Sigma(\mu)$ based on his speculation (so-called Voiculescu's heuristics) that $\Sigma(\mu)$ appears as a normalized limit of the relative entropies of the distributions of certain random matrices with respect to the Lebesgue measure on the matrix space. A more rigorous derivation of Voiculescu's heuristics was later given in [**2**]. In Sections 4 and 5, we show a similar but different result asserting that the free relative entropy $\Sigma(\nu^h, \nu)$ for the perturbed measure ν^h via free relative entropy is a normalized limit of the relative entropies of the distributions of random matrices perturbed according to h.

As is briefly explained in the last section, free relative entropy and the corresponding perturbation theory for probability measures on $\boldsymbol{R}$ can be fully adapted to the case of probability measures on the unit circle.

Throughout the paper our main reference is [**13**] concerning random matrix models, related large deviations and entropy in free probability theory.

1. Free entropy and free relative entropy.

For a probability Borel measure μ on $\boldsymbol{R}$ the *free entropy* $\Sigma(\mu)$ was introduced by Voiculescu [**19**] as

$$\Sigma(\mu) := \iint \log|x - y|\, d\mu(x)\, d\mu(y), \tag{1.1}$$

and it is indeed the minus sign of the so-called *logarithmic energy* of μ familiar in potential theory ([**18**]). In this paper the *support* of μ, denoted by $\operatorname{supp}\mu$, is that in the topological sense, i.e. the smallest closed subset $K \subset \boldsymbol{R}$ such that $\mu(K) = 1$. Note that the double integral (1.1) always exists with a value in

$[-\infty, +\infty)$ whenever μ is compactly supported. The free entropy functional $\Sigma(\mu)$ is upper semi-continuous in weak topology when the support of μ is restricted in a compact set, and it is strictly concave (see [**13**, 5.3.2]).

The microstates (or matricial) approach for free entropy was developed in [**20**]. For each $n \in \boldsymbol{N}$ let M_n denote the space of all $n \times n$ complex matrices and tr_n the normalized trace functional on M_n. The set of all selfadjoint matrices in M_n is denoted by M_n^{sa}. There is a natural linear bijection between M_n^{sa} and $\boldsymbol{R}^{n^2}$ which is an isometry for the Hilbert-Schmidt and Euclidean norms, so the "Lebesgue" measure Λ_n on M_n^{sa} is induced by the Lebesgue measure on $\boldsymbol{R}^{n^2}$ via this isometry. Let μ be a probability Borel measure supported in $[-R, R]$, $R > 0$. For $n, r \in \boldsymbol{N}$ and $\varepsilon > 0$ define

$$\Gamma_R(\mu; n, r, \varepsilon) := \{A \in M_n^{sa} : \|A\| \le R, |\mathrm{tr}_n(A^k) - m_k(\mu)| \le \varepsilon, k \le r\}, \tag{1.2}$$

where $\|A\|$ is the operator norm and $m_k(\mu) := \int x^k \, d\mu(x)$, the kth moment of μ. Then the limit

$$\chi_R(\mu; r, \varepsilon) := \lim_{n\to\infty} \left[\frac{1}{n^2} \log \Lambda_n(\Gamma_R(\mu; n, r, \varepsilon)) + \frac{1}{2} \log n\right] \tag{1.3}$$

exists for every $r \in \boldsymbol{N}$ and $\varepsilon > 0$, and

$$\lim_{r\to\infty, \varepsilon\to+0} \chi_R(\mu; r, \varepsilon) = \Sigma(\mu) + \frac{1}{2}\log(2\pi) + \frac{3}{4}. \tag{1.4}$$

(See [**13**, 5.6.2] for the existence of the limit in (1.3) while lim was originally lim sup in [**20**].)

The *Boltzmann-Gibbs entropy* $S(\mu)$ of a probability measure μ on $\boldsymbol{R}$ is given as

$$S(\mu) := -\int \frac{d\mu}{dx} \log \frac{d\mu}{dx} dx$$

if μ is absolutely continuous with respect to the Lebesgue measure dx and $d\mu/dx$ is the Radon-Nikodym derivative; otherwise $S(\mu) := -\infty$. The *relative entropy* (or the *Kullback-Leibler divergence*) $S(\mu, \nu)$ of μ with respect to another probability measure ν is defined as

$$S(\mu, \nu) := \int \frac{d\mu}{d\nu} \log \frac{d\mu}{d\nu} d\nu = \int \log \frac{d\mu}{d\nu} d\mu$$

if μ is absolutely continuous with respect to ν; otherwise $S(\mu, \nu) := +\infty$. If μ and ν are supported in $[-R, R]$, then these entropies have the asymptotic expressions as follows:

$$S(\mu) = \lim_{r\to\infty,\varepsilon\to+0} \lim_{n\to\infty} \frac{1}{n} \log L^n\left(\left\{(x_1,\dots,x_n) \in [-R,R]^n : \left|\frac{x_1^k+\cdots+x_n^k}{n} - m_k(\mu)\right| \le \varepsilon, k \le r\right\}\right), \tag{1.5}$$

$$-S(\mu,\nu) = \lim_{r\to\infty,\varepsilon\to+0} \lim_{n\to\infty} \frac{1}{n} \log \nu^n\left(\left\{(x_1,\dots,x_n) \in [-R,R]^n : \left|\frac{x_1^k+\cdots+x_n^k}{n} - m_k(\mu)\right| \le \varepsilon, k \le r\right\}\right), \tag{1.6}$$

where L^n is the n-dimensional Lebesgue measure and ν^n is the n-fold product of ν. These expressions can be derived from Sanov's large deviation theorem for the empirical distribution of i.i.d. random variables (see [**13**, 5.1.1] for details).

The free entropy $\Sigma(\mu)$ is the free analogue of the Boltzmann-Gibbs entropy $S(\mu)$, and the asymptotic expression given in (1.2)–(1.4) (with scale n^{-2}) is the "free" counterpart of the expression (1.5) (with scale n^{-1}). Now, naturally arises the following question: What is the free analogue of the relative entropy $S(\mu,\nu)$? The problem was recently investigated in [**14**], and it turned out that the *free relative entropy* $\Sigma(\mu,\nu)$ of μ with respect to ν can be defined as

$$\Sigma(\mu,\nu) = -\iint \log|x-y|\, d(\mu-\nu)(x)\, d(\mu-\nu)(y), \tag{1.7}$$

which is the logarithmic energy of a signed measure $\mu-\nu$. But the above (1.7) may not be well-defined, and to be more precise we adopt the following definition:

$$\Sigma(\mu,\nu) := \lim_{\varepsilon\to+0}\left[-\iint \log(|x-y|+\varepsilon)\, d(\mu-\nu)(x)\, d(\mu-\nu)(y)\right]. \tag{1.8}$$

In fact, this is well-defined because $\varepsilon > 0 \mapsto -\iint \log(|x-y|+\varepsilon)\, d(\mu-\nu)(x) \cdot d(\mu-\nu)(y)$ is increasing as $\varepsilon \searrow 0$ ([**14**, Lemma 3.6]). Of course, the integral (1.7) exists and coincides with (1.8) as long as $\log|x-y|$ is integrable with respect to $d|\mu-\nu|(x)\, d|\mu-\nu|(y)$; in particular, this is the case if $\Sigma(\mu) > -\infty$ and $\Sigma(\nu) > -\infty$ (see the proof of [**13**, 5.3.2]).

In [**14**] the asymptotic expression of the free relative entropy $\Sigma(\mu,\nu)$ was obtained in the microstates approach. Before stating it we here give a brief exposition on some large deviation result related to random matrices, which is a basis of deriving the asymptotic expression of $\Sigma(\mu,\nu)$. This large deviation will indeed play a crucial role in Section 4 as well.

Let $R > 0$ and Q be a real continuous function on $[-R,R]$. For each $n \in \mathbf{N}$ define the probability distribution $\tilde{\lambda}_n(Q;R)$ on $\mathbf{R}^n$ by

$$\tilde{\lambda}_n(Q;R) := \frac{1}{Z_n(Q;R)} \exp\left(-n\sum_{i=1}^{n} Q(x_i)\right) \prod_{i<j} |x_i - x_j|^2$$
$$\times \prod_{i=1}^{n} \chi_{[-R,R]}(x_i)\, dx_1 dx_2 \cdots dx_n, \tag{1.9}$$

where $Z_n(Q;R)$ is the normalizing constant:

$$Z_n(Q;R) := \int_{-R}^{R} \cdots \int_{-R}^{R} \exp\left(-n\sum_{i=1}^{n} Q(x_i)\right) \prod_{i<j} |x_i - x_j|^2\, dx_1 \cdots dx_n. \tag{1.10}$$

Moreover, let $\lambda_n(Q;R)$ be the probability distribution on M_n^{sa} which is invariant under unitary conjugation and whose joint eigenvalue distribution on $\boldsymbol{R}^n$ is $\tilde{\lambda}_n(Q;R)$; more explicitly,

$$\lambda_n(Q;R) := (dU \otimes \tilde{\lambda}_n(Q;R)) \circ \Phi_n^{-1}, \tag{1.11}$$

where dU is the Haar probability measure on the n-dimensional unitary group $\mathscr{U}_n$ and $\Phi_n : \mathscr{U}_n \times \boldsymbol{R}^n \to M_n^{sa}$ is defined as

$$\Phi_n(U,(x_1,\dots,x_n)) := U \operatorname{diag}(x_1,\dots,x_n) U^*.$$

One can consider $\lambda_n(Q;R)$ as the distribution of an $n \times n$ random selfadjoint matrix, or more explicitly $\lambda_n(Q;R)$ itself as a random matrix. The support of $\lambda_n(Q;R)$ is

$$(M_n^{sa})_R := \{A \in M_n^{sa} : \|A\| \le R\}. \tag{1.12}$$

The *empirical eigenvalue distribution* of this random matrix is the random discrete measure

$$\frac{\delta(x_1) + \delta(x_2) + \cdots + \delta(x_n)}{n},$$

where $\delta(x)$ is the point measure at x and the $\boldsymbol{R}^n$-vector $(x_1, x_2, \dots, x_n)$ is distributed subject to the distribution (1.9). Let $\mathscr{M}([-R,R])$ denote the set of all probability measures supported in $[-R,R]$ equipped with the weak topology. Then we have the following large deviation theorem.

THEOREM 1.1. *Let Q and Q_n $(n \in \boldsymbol{N})$ be real continuous functions on $[-R,R]$ such that $Q_n(x) \to Q(x)$ uniformly on $[-R,R]$. For each $n \in \boldsymbol{N}$ define the probability distribution $\tilde{\lambda}_n(Q_n;R)$ supported on $[-R,R]^n$ by* (1.9) *and the normalizing constant $Z_n(Q_n;R)$ by* (1.10) *with Q_n in place of Q. Then the finite limit*

$$B(Q;R) := \lim_{n\to\infty} \frac{1}{n^2} \log Z_n(Q_n;R) \tag{1.13}$$

exists, and if $(x_1, \dots, x_n) \in [-R, R]^n$ *is distributed with the joint distribution* $\tilde{\lambda}_n(Q_n; R)$, *then the empirical distribution* $(1/n)(\delta(x_1) + \dots + \delta(x_n))$ *satisfies the large deviation principle in the scale* n^{-2} *with the good rate function*:

$$I(\mu) := -\Sigma(\mu) + \mu(Q) + B(Q; R) \quad \textit{for } \mu \in \mathcal{M}([-R, R]).$$

There exists a unique minimizer μ_Q *of* I *with* $I(\mu_Q) = 0$ *and* $B(Q; R)$ *is determined by only* Q (*independently of the choice of* $\{Q_n\}$). *Furthermore, the above empirical distribution converges almost surely to* μ_Q *as* $n \to \infty$ *in weak topology.*

The above large deviation is a matricial counterpart of the famous Sanov large deviation theorem ([**5**], [**6**]). The probability distribution P_n on $\mathcal{M}([-R, R])$ of the random measure $(1/n)(\delta(x_1) + \dots + \delta(x_n))$ is given by

$$P_n(\Gamma) := \tilde{\lambda}_n(Q_n; R)\left(\left\{(x_1, \dots, x_n) \in [-R, R]^n : \frac{\delta(x_1) + \dots + \delta(x_n)}{n} \in \Gamma\right\}\right)$$

for Borel sets $\Gamma \subset \mathcal{M}([-R, R])$. Since $\mathcal{M}([-R, R])$ is weakly compact (hence the exponential tightness of (P_n) is automatic), it suffices (see [**5**, 4.1.11]) to prove that

$$-I(\mu) = \inf_G \left[\limsup_{n \to \infty} \frac{1}{n^2} \log P_n(G)\right] = \inf_G \left[\liminf_{n \to \infty} \frac{1}{n^2} \log P_n(G)\right]$$

for every $\mu \in \mathcal{M}([-R, R])$, where G runs over the neighborhoods of μ. The proof is more or less similar to that of [**13**, 5.4.3 and 5.5.1], so we omit the details. The assertion on the minimizer is a consequence of the fundamental result on weighted potentials (see [**18**, I.1.3 and I.3.1] or Lemma 2.1 below). The proof of the last statement is found in [**13**, p. 211].

Now let us return to the free relative entropy. Let ν be a compactly supported probability measure on $\boldsymbol{R}$, and assume that the function

$$Q_\nu(x) := 2 \int \log|x - y| \, d\nu(y) \tag{1.14}$$

is finite and continuous (as a function on $\boldsymbol{R}$) at every $x \in \operatorname{supp} \nu$. Then Q_ν is a continuous function on the whole $\boldsymbol{R}$, because $Q_\nu(x)$ is always continuous on $\boldsymbol{R} \setminus \operatorname{supp} \nu$. For instance, this is the case when ν is absolutely continuous with respect to dx and $d\nu/dx$ is bounded. For $R > 0$ define the probability distribution $\lambda_n(\nu; R)$ on M_n^{sa} by putting $Q = Q_\nu$ in (1.9) and (1.11): $\lambda_n(\nu; R) := \lambda_n(Q_\nu; R)$. Then the next theorem was proved in [**14**, Theorem 3.8] by appealing to the above large deviation theorem in the case $Q_n = Q = Q_\nu$.

Theorem 1.2. *Let* μ, ν *be compactly supported probability measures, and assume that* $Q_\nu(x)$ *in* (1.14) *is continuous on* $\boldsymbol{R}$. *Then for any* $R > 0$ *with* $\operatorname{supp} \mu$, $\operatorname{supp} \nu \subset [-R, R]$,

$$-\Sigma(\mu, \nu) = \lim_{r\to\infty, \varepsilon\to+0} \lim_{n\to\infty} \frac{1}{n^2} \log \lambda_n(\nu; R)(\Gamma_R(\mu; n, r, \varepsilon)), \tag{1.15}$$

where $\Gamma_R(\mu; n, r, \varepsilon)$ *is given in* (1.2).

The above expression (1.15) is the free analogue of (1.6). The reference measure $\lambda_n(\nu; R)$ on M_n^{sa} is a bit more complicated than the product ν^n on $\boldsymbol{R}^n$ in (1.6), but it is the right one in free (or matricial) probability. In fact, Theorem 1.1 (together with Lemma 2.1) says that the empirical eigenvalue distribution of the $n \times n$ selfadjoint random matrix having the distribution $\lambda_n(\nu; R)$ converges almost surely to ν, the minimizer of the rate function, as $n \to \infty$ in weak topology (hence in the distribution sense). In this way, Theorem 1.2 gives a justification for our free relative entropy $\Sigma(\mu, \nu)$. Another (more decisive) justification will be presented in Section 4.

We end the section by listing basic properties of $\Sigma(\mu, \nu)$ given in [**14**]. The properties except the first one are analogous to those of the relative entropy $S(\mu, \nu)$. Let μ, ν, μ_i, ν_i be compactly supported probability measures on $\boldsymbol{R}$.

(i) *Symmetry*: $\Sigma(\mu, \nu) = \Sigma(\nu, \mu)$.

(ii) *Strict positivity*: $\Sigma(\mu, \nu) \geq 0$, and $\Sigma(\mu, \nu) = 0$ if and only if $\mu = \nu$.

(iii) *Joint convexity*:

$$\Sigma(\lambda\mu_1 + (1-\lambda)\mu_2, \lambda\nu_1 + (1-\lambda)\nu_2) \leq \lambda\Sigma(\mu_1, \nu_1) + (1-\lambda)\Sigma(\mu_2, \nu_2)$$

for $0 \leq \lambda \leq 1$.

(iv) *Single strict convexity*:

$$\Sigma(\lambda\mu_1 + (1-\lambda)\mu_2, \nu) < \lambda\Sigma(\mu_1, \nu) + (1-\lambda)\Sigma(\mu_2, \nu)$$

if $0 < \lambda < 1$, $\mu_1 \neq \mu_2$, $\Sigma(\mu_1, \nu) < +\infty$ and $\Sigma(\mu_2, \nu) < +\infty$.

(v) *Joint lower semi-continuity*: $\Sigma(\mu, \nu)$ is jointly lower semi-continuous in weak topology when the supports of μ, ν are restricted in a compact set.

2. Perturbation via free relative entropy.

Let K be a fixed compact subset of $\boldsymbol{R}$, and let $\mathscr{M}(K)$ denote the set of all probability Borel measures supported in K. Also, let $C_{\boldsymbol{R}}(K)$ denote the space of all real continuous functions on K. For $\mu \in \mathscr{M}(K)$ and $h \in C_{\boldsymbol{R}}(K)$ we write $\mu(h)$ for $\int_K h\, d\mu$.

The (logarithmic) *capacity* of a compact set $C \subset \boldsymbol{R}$ is defined as

$$\operatorname{cap}(C) := \exp(\sup\{\Sigma(\mu) : \mu \in \mathscr{M}(C)\})$$

with convention $\operatorname{cap}(C) = 0$ if $\Sigma(\mu) = -\infty$ for all $\mu \in \mathscr{M}(C)$. Then the capacity of a general Borel set $A \subset \boldsymbol{R}$ is defined as $\operatorname{cap}(A) := \sup\{\operatorname{cap}(C) : C \subset A \text{ compact}\}$.

A property is said to hold for *quasi-every* $x \in A$ if it holds for all $x \in A$ except in a set of capacity zero.

Throughout this section, let $\nu \in \mathcal{M}(K)$ be such that the function $Q = Q_\nu$ given in (1.14) is finite and continuous on K; hence $Q \in C_{\boldsymbol{R}}(K)$ and K has positive capacity. For given $h \in C_{\boldsymbol{R}}(K)$ define the *weighted energy integral*

$$E_h(\mu) := \iint \log \frac{1}{|x-y|} \, d\mu(x) \, d\mu(y) + \int h \, d\mu = -\Sigma(\mu) + \mu(h)$$

for $\mu \in \mathcal{M}(K)$. For later use, we state the fundamental result in the theory of weighted potentials ([**18**, I.1.3 and I.3.1]) in a reduced form of the next lemma.

Lemma 2.1. *For every $h \in C_{\boldsymbol{R}}(K)$ the following assertions hold*:

(i) *There exists a unique $\mu_h \in \mathcal{M}(K)$ such that*

$$E_h(\mu_h) = \inf\{E_h(\mu) : \mu \in \mathcal{M}(K)\}.$$

(ii) *$E_h(\mu_h)$ and $\Sigma(\mu_h)$ are finite.*

(iii) *The minimizer μ_h is characterized as $\mu_h \in \mathcal{M}(K)$ such that for some $B \in \boldsymbol{R}$*

$$2 \int \log|x-y| \, d\mu_h(y) \begin{cases} \geq h(x) + B & \text{for all } x \in \operatorname{supp} \mu_h, \\ \leq h(x) + B & \text{for quasi-every } x \in K. \end{cases}$$

In this case, $B = -2E_h(\mu_h) + \mu_h(h)$.

For $\nu \in \mathcal{M}(K)$ fixed as above, the *Legendre transform* of $\mu \in \mathcal{M}(K) \mapsto \Sigma(\mu, \nu)$ is defined as

$$c(h, \nu) := \sup\{-\mu(h) - \Sigma(\mu, \nu) : \mu \in \mathcal{M}(K)\} \tag{2.1}$$

for each $h \in C_{\boldsymbol{R}}(K)$.

Theorem 2.2. *With the above definitions, the following assertions hold*:

(i) *$c(\cdot, \nu)$ is a convex function on $C_{\boldsymbol{R}}(K)$ satisfying*

$$-\nu(h) \leq c(h, \nu) \leq \|h\| \tag{2.2}$$

(*in particular*, $c(0, \nu) = 0$) *where $\|h\|$ is the sup-norm, and it is decreasing, i.e. $c(h_1, \nu) \geq c(h_2, \nu)$ if $h_1 \leq h_2$. Moreover*,

$$|c(h_1, \nu) - c(h_2, \nu)| \leq \|h_1 - h_2\| \tag{2.3}$$

for all $h_1, h_2 \in C_{\boldsymbol{R}}(K)$.

(ii) *For every $\mu \in \mathcal{M}(K)$*,

$$\Sigma(\mu, \nu) = \sup\{-\mu(h) - c(h, \nu) : h \in C_{\boldsymbol{R}}(K)\}. \tag{2.4}$$

(iii) *For every $h \in C_{\boldsymbol{R}}(K)$ there exists a unique $\nu^h \in \mathcal{M}(K)$ such that*

$$-\nu^h(h) - \Sigma(\nu^h, \nu) = c(h, \nu). \tag{2.5}$$

Moreover, $\Sigma(\nu^h)$ *is finite and*

$$c(h, \nu) = \Sigma(\nu^h) + \Sigma(\nu) - \nu^h(Q + h). \tag{2.6}$$

(iv) *For every* $h \in C_{\boldsymbol{R}}(K)$ *and* $\mu \in \mathscr{M}(K)$, $\mu = \nu^h$ *if and only if*

$$c(h + k, \nu) \geq c(h, \nu) - \mu(k) \quad \textit{for all } k \in C_{\boldsymbol{R}}(K).$$

Proof. (i) The convexity and the decreasingness of $c(\cdot\,, \nu)$ are obvious by definition. (2.2) follows immediately from the positivity of $\Sigma(\mu, \nu)$. For every $h_1, h_2 \in C_{\boldsymbol{R}}(K)$ and $\mu \in \mathscr{M}(K)$,

$$-\mu(h_1) - \Sigma(\mu, \nu) = -\mu(h_2) - \Sigma(\mu, \nu) + \mu(h_2 - h_1) \leq c(h_2, \nu) + \|h_1 - h_2\|,$$

and hence

$$c(h_1, \nu) \leq c(h_2, \nu) + \|h_1 - h_2\|.$$

This implies (2.3) by symmetry.

(ii) Let $C_{\boldsymbol{R}}(K)^*$ denote the space of all signed Borel measures on K, which is a Banach space with respect to the total variation norm and is identified with the dual Banach space of $C_{\boldsymbol{R}}(K)$ with the sup-norm. Note also that the weak topology on $\mathscr{M}(K)$ coincides with the w*-topology. The definition (2.1) means that $c(\cdot\,, \nu) : C_{\boldsymbol{R}}(K) \to \boldsymbol{R}$ is the conjugate function (or the Legendre transform) of the function $\varphi : C_{\boldsymbol{R}}(K)^* \to [0, +\infty]$ given by

$$\varphi(\mu) := \begin{cases} \Sigma(\mu, \nu) & \text{if } \mu \in \mathscr{M}(K), \\ +\infty & \text{if } \mu \in C_{\boldsymbol{R}}(K)^* \backslash \mathscr{M}(K). \end{cases}$$

By the properties of $\Sigma(\mu, \nu)$ listed at the end of Section 1, it follows that φ is a w*-lower semi-continuous and convex function. Hence (2.4) is a consequence of the general duality theorem for conjugate functions.

(iii) We first show that

$$\Sigma(\mu, \nu) = -\Sigma(\mu) - \Sigma(\nu) + \mu(Q) \tag{2.7}$$

for all $\mu \in \mathscr{M}(K)$ permitting both sides being $+\infty$. Indeed, for $\varepsilon > 0$ we write

$$\begin{aligned} &-\iint \log(|x - y| + \varepsilon)\, d(\mu - \nu)(x)\, d(\mu - \nu)(y) \\ &\quad = -\iint \log(|x - y| + \varepsilon)\, d\mu(x)\, d\mu(y) - \iint \log(|x - y| + \varepsilon)\, d\nu(x)\, d\nu(y) \\ &\qquad + 2\int \left(\int \log(|x - y| + \varepsilon)\, d\nu(y) \right) d\mu(x), \end{aligned}$$

and the first two terms in the right-hand side converge as $\varepsilon \searrow 0$ to $-\Sigma(\mu)$ and $-\Sigma(\nu)$, respectively. When $\varepsilon \le 1$ and $R := \max\{|x-y| : x, y \in K\}$, the diameter of K, since

$$\int \log(|x-y|+\varepsilon)\,d\nu(y) \le \log(R+1)$$

and $2\int \log(|x-y|+\varepsilon)\,d\nu(y) \searrow Q(x)$ as $\varepsilon \searrow 0$, the last term in the above equality converges to $\mu(Q)$ by the monotone convergence theorem. Hence (2.7) is obtained.

By (2.7) we have

$$-\mu(h) - \Sigma(\mu,\nu) = \Sigma(\mu) - \mu(Q+h) + \Sigma(\nu) = -E_{Q+h}(\mu) + \Sigma(\nu). \tag{2.8}$$

Hence Lemma 2.1 implies that there exists a unique ν^h (or μ_{Q+h}) in $\mathcal{M}(K)$ for which (2.5) is satisfied, and $\Sigma(\nu^h) > -\infty$. The formula (2.6) is obvious from (2.8).

(iv) Rewrite

$$c(h+k,\nu) \ge c(h,\nu) - \nu(k) \quad \text{for all } k \in C_{\boldsymbol{R}}(K)$$

as

$$c(h+k,\nu) + \mu(h+k) \ge c(h,\nu) + \nu(h) \quad \text{for all } k \in C_{\boldsymbol{R}}(K).$$

By (2.4) this condition is equivalent to

$$-\Sigma(\mu,\nu) \ge c(h,\nu) + \mu(h),$$

which means $\mu = \nu^h$. □

We call ν^h in Theorem 2.2 the *perturbed probability measure* of ν by h (via free relative entropy). Note that the variational expression (2.4) of $\Sigma(\mu,\nu)$ is valid for any choice of a compact $K \subset \boldsymbol{R}$ such that $K \supset \operatorname{supp}\mu, \operatorname{supp}\nu$. Clearly, $\nu^{h+\alpha} = \nu^h$ and $c(h+\alpha,\nu) = c(h,\nu) - \alpha$ for $\alpha \in \boldsymbol{R}$.

It is instructive to consider the perturbed measure ν^h in comparison with the similar perturbation via relative entropy. For any $\nu \in \mathcal{M}(K)$ and $h \in C_{\boldsymbol{R}}(K)$, it is well-known that

$$\log \nu(e^{-h}) = \sup\{-\mu(h) - S(\mu,\nu) : \mu \in \mathcal{M}(K)\}$$

and the probability measure $\mu_0 := (e^{-h}/\nu(e^{-h}))\nu$ (i.e. $d\mu_0/d\nu = e^{-h}/\nu(e^{-h})$) is a unique maximizer of $-\mu(h) - S(\mu,\nu)$ for $\mu \in \mathcal{M}(K)$. In fact, this can be easily verified by using the strict positivity of $S(\mu,\mu_0)$. Moreover, for every $\mu \in \mathcal{M}(K)$,

$$S(\mu,\nu) = \sup\{-\mu(h) - \log \nu(e^{-h}) : h \in C_{\boldsymbol{R}}(K)\}.$$

The probability measure μ_0 perturbed from ν via the relative entropy $S(\mu, \nu)$ is the so-called *Gibbs ensemble*. The above $c(h, \nu)$ is considered as the "free" counterpart of $\log \nu(e^{-h})$, and the characterization of ν^h in the above (iv) is the "free" analogue of the so-called *variational principle* for Gibbs ensembles ([**17**]). It is worth noting, as mentioned in Introduction, that this type of perturbation theory via relative entropy was developed even in the quantum probabilistic setting on operator algebras ([**16**], [**7**], [**15**, Section 12]).

We shall write $\nu^{h,\Sigma}$ for ν^h in Theorem 2.2 and $\nu^{h,S}$ for the above μ_0, when both perturbed measures via $\Sigma(\mu, \nu)$ and $S(\mu, \nu)$ are simultaneously treated. A simple expression of $c(h, \nu)$ such as $\log \nu(e^{-h})$ is not available; nevertheless in Section 4 we shall give an asymptotic expression of $c(h, \nu)$.

Proposition 2.3. *For every* $\mu \in \mathcal{M}(K)$,

$$\Sigma(\mu, \nu^h) \le \Sigma(\mu, \nu) + \mu(h) + c(h, \nu). \tag{2.9}$$

Moreover, if $\operatorname{supp} \mu \subset \operatorname{supp} \nu^h$, *then*

$$\Sigma(\mu, \nu^h) = \Sigma(\mu, \nu) + \mu(h) + c(h, \nu). \tag{2.10}$$

Proof. Since ν^h is the minimizer of $E_{Q+h}(\mu)$ due to (2.8), we have by Lemma 2.1

$$2\int \log|x-y|\, d\nu^h(y) \begin{cases} \ge Q(x) + h(x) + B & \text{for all } x \in \operatorname{supp} \nu^h, \\ \le Q(x) + h(x) + B & \text{for quasi-every } x \in K, \end{cases} \tag{2.11}$$

where

$$\begin{aligned} B &= -2E_{Q+h}(\nu^h) + \nu^h(Q+h) \\ &= 2\Sigma(\nu^h) - \nu^h(Q+h) \\ &= \Sigma(\nu^h) - \Sigma(\nu) + c(h, \nu) \end{aligned}$$

from (2.6). For (2.9) there is nothing to prove when $\Sigma(\mu, \nu) = +\infty$, so assume that $\Sigma(\mu, \nu) < +\infty$ and hence $\Sigma(\mu) > -\infty$ by (2.7). For $\varepsilon > 0$ we write

$$\begin{aligned} &-\iint \log(|x-y| + \varepsilon)\, d(\mu - \nu^h)(x)\, d(\mu - \nu^h)(y) \\ &\quad = -\iint \log(|x-y| + \varepsilon)\, d\mu(x)\, d\mu(y) - \iint \log(|x-y| + \varepsilon)\, d\nu^h(x)\, d\nu^h(y) \\ &\qquad + 2\iint \log(|x-y| + \varepsilon)\, d\nu^h(y)\, d\mu(x). \end{aligned} \tag{2.12}$$

The first two terms of (2.12) have the finite limits $\Sigma(\mu)$ and $\Sigma(\nu^h)$ as $\varepsilon \searrow 0$, and the monotone convergence theorem can be applied to the last term of (2.12) as in the proof (iii) of Theorem 2.2. Hence we get

$$\Sigma(\mu, \nu^h) = -\Sigma(\mu) - \Sigma(\nu^h) + 2\int\left(\int \log|x-y|\,d\nu^h(y)\right)d\mu(x).$$

Since $\Sigma(\mu) > -\infty$ implies that $\mu(A) = 0$ for any Borel set A of capacity zero (see [**18**, I.1.7]), we can use the second inequality of (2.11) to obtain (2.9):

$$\begin{aligned}\Sigma(\mu, \nu^h) &\le -\Sigma(\mu) - \Sigma(\nu^h) + \mu(Q+h) + B \\ &= -\Sigma(\mu) - \Sigma(\nu) + \mu(Q+h) + c(h, \nu) \\ &= \Sigma(\mu, \nu) + \mu(h) + c(h, \nu).\end{aligned}$$

Next, assume that $\operatorname{supp}\mu \subset \operatorname{supp}\nu^h$. Then the first inequality of (2.11) implies that

$$2\int\left(\int \log|x-y|\,d\nu^h(y)\right)d\mu(x) \ge \mu(Q+h) + B > -\infty.$$

Hence we can take the limit $\varepsilon \searrow 0$ of (2.12) (regardless of $\Sigma(\mu) > -\infty$ or not) so that

$$\begin{aligned}\Sigma(\mu, \nu^h) &\ge -\Sigma(\mu) - \Sigma(\nu^h) + \mu(Q+h) + B \\ &= \Sigma(\mu, \nu) + \mu(h) + c(h, \nu).\end{aligned}$$

This together with (2.9) gives (2.10). □

Corollary 2.4. *For every* $h \in C_{\boldsymbol{R}}(K)$,

$$\Sigma(\nu^h, \nu) \le \frac{\nu(h) - \nu^h(h)}{2} \le \|h\|,$$

$$c(h, \nu) \ge -\nu(h) + \Sigma(\nu^h, \nu) \ge -\frac{\nu(h) + \nu^h(h)}{2}.$$

Furthermore, if $\operatorname{supp}\nu \subset \operatorname{supp}\nu^h$, *then*

$$\Sigma(\nu^h, \nu) = \frac{\nu(h) - \nu^h(h)}{2},$$

$$c(h, \nu) = -\nu(h) + \Sigma(\nu^h, \nu) = -\frac{\nu(h) + \nu^h(h)}{2}.$$

Proof. Putting $\mu = \nu$ in (2.9) gives

$$\begin{aligned}\Sigma(\nu^h,\nu) = \Sigma(\nu,\nu^h) &\le \nu(h) + c(h,\nu)\\ &= \nu(h) - \nu^h(h) - \Sigma(\nu^h,\nu)\end{aligned}$$

from (2.5), and this implies the first assertion. If $\operatorname{supp}\nu \subset \operatorname{supp}\nu^h$, then equality occurs in the above thanks to (2.10). □

The next proposition is the chain rule for the perturbation $\nu \mapsto \nu^h$.

Proposition 2.5. *Let* $h, k \in C_{\boldsymbol{R}}(K)$. *If* $Q_{\nu^h}(x) := 2\int \log|x-y|\,d\nu^h(y)$ *as well as* $Q = Q_\nu$ *is continuous on* K *and* $\operatorname{supp}(\nu^h)^k \subset \operatorname{supp}\nu^h$, *then*

$$(\nu^h)^k = \nu^{h+k},$$

$$c(h+k,\nu) = c(h,\nu) + c(k,\nu^h).$$

In particular, these hold if $\operatorname{supp}\nu^h = K$ *and* $Q_{\nu^h} = Q + h$ *on* K.

Proof. Since Q_{ν^h} is continuous on K by assumption, the perturbation $(\nu^h)^k$ of ν^h is well-defined and it is characterized as follows: for some $B' \in \boldsymbol{R}$,

$$2\int \log|x-y|\,d(\nu^h)^k(y) \begin{cases} \ge Q_{\nu^h}(x) + k(x) + B' & \text{for all } x \in \operatorname{supp}(\nu^h)^k, \\ \le Q_{\nu^h}(x) + k(x) + B' & \text{for quasi-every } x \in K. \end{cases}$$

Since $\operatorname{supp}(\nu^h)^k \subset \operatorname{supp}\nu^h$, combining this with (2.11) gives

$$2\int \log|x-y|\,d(\nu^h)^k(y) \begin{cases} \ge Q(x) + h(x) + k(x) + B + B' & \text{for all } x \in \operatorname{supp}(\nu^h)^k, \\ \le Q(x) + h(x) + k(x) + B + B' & \text{for quasi-every } x \in K, \end{cases}$$

which characterizes ν^{h+k} so that $(\nu^h)^k = \nu^{h+k}$.

By (2.5) we get

$$c(h+k,\nu) = -\nu^{h+k}(h+k) - \Sigma(\nu^{h+k},\nu),$$

$$c(k,\nu^h) = -(\nu^h)^k(k) - \Sigma((\nu^h)^k,\nu^h) = -\nu^{h+k}(k) - \Sigma(\nu^{h+k},\nu^h).$$

Since $\operatorname{supp}\nu^{h+k} \subset \operatorname{supp}\nu^h$, we also get by (2.10)

$$\Sigma(\nu^{h+k},\nu^h) = \Sigma(\nu^{h+k},\nu) + \nu^{h+k}(h) + c(h,\nu).$$

Combining the above three equalities yields

$$c(h+k,\nu) = c(h,\nu) + c(k,\nu^h),$$

as desired. □

Corollary 2.6. *Assume one of the following conditions*:

(a) *$\mu \in \mathcal{M}(K)$ is such that Q_μ as well as Q_ν is continuous on K, and $h := Q_\mu - Q_\nu$,*
(b) *$h \in C_{\boldsymbol{R}}(K)$ and $\mu := \nu^h$ satisfies* $\operatorname{supp} \nu \subset \operatorname{supp} \mu$.

Then for each $0 \le \lambda \le 1$,

$$\nu^{\lambda h} = (1-\lambda)\nu + \lambda\mu,$$

$$\Sigma(\nu^{\lambda h}, \nu) = \lambda^2 \Sigma(\mu, \nu),$$

$$c(\lambda h, \nu) = -\lambda\nu(h) + \lambda^2 \Sigma(\mu, \nu).$$

Proof. We just prove the case (b); the proof of the case (a) is similar. For $0 < \lambda < 1$ put $\mu_\lambda := (1-\lambda)\nu + \lambda\mu$, so we have $\operatorname{supp} \mu_\lambda = \operatorname{supp} \mu$ thanks to $\operatorname{supp} \nu \subset \operatorname{supp} \mu$. From the characterization of $\mu = \nu^h$ we have for some $B \in \boldsymbol{R}$

$$Q_{\mu_\lambda}(x) = (1-\lambda)Q_\nu(x) + \lambda Q_\mu(x) \begin{cases} \ge Q_\nu(x) + \lambda h(x) + \lambda B & \text{for all } x \in \operatorname{supp} \mu_\lambda, \\ \le Q_\nu(x) + \lambda h(x) + \lambda B & \text{for quasi-every } x \in K, \end{cases}$$

which implies that $\mu_\lambda = \nu^{\lambda h}$. By Corollary 2.4,

$$\begin{aligned} \Sigma(\nu^{\lambda h}, \nu) &= \frac{\nu(\lambda h) - ((1-\lambda)\nu + \lambda\mu)(\lambda h)}{2} \\ &= \lambda^2 \frac{\nu(h) - \mu(h)}{2} = \lambda^2 \Sigma(\mu, \nu), \end{aligned}$$

$$c(\lambda h, \nu) = -\nu(\lambda h) + \Sigma(\nu^{\lambda h}, \nu) = -\lambda\nu(h) + \lambda^2 \Sigma(\mu, \nu),$$

and the latter holds also for $\lambda = 0, 1$. □

As for the perturbation $\nu \mapsto \nu^{h,S}$ via relative entropy, $\operatorname{supp} \nu^{h,S} = \operatorname{supp} \nu$ is obvious and the formulas

$$S(\mu, \nu^{h,S}) = S(\mu, \nu) + \mu(h) + \log \nu(e^{-h}), \tag{2.13}$$

$$(\nu^{h,S})^{k,S} = \nu^{h+k,S},$$

$$\log \nu(e^{-(h+k)}) = \log \nu(e^{-h}) + \log \nu^{h,S}(e^{-k})$$

hold in general.

The relation between ν and $\nu^h = \nu^{h,\Sigma}$ is more complicated than that between ν and $\nu^{h,S}$. However, the formulas in Corollary 2.6 (though they do not generally hold) are quite simple compared with those for $\nu^{\lambda h,S}$; in fact, $\nu^{\lambda h,S}$ $(0 \le \lambda \le 1)$ is not a line segment, and $(d^2/d\lambda^2)S(\nu^{\lambda h,S}, \nu)$ and $(d^2/d\lambda^2)S(\nu, \nu^{\lambda h,S})$ are non-constant functions of λ. The simple formulas for $\nu^{\lambda h,\Sigma}$ in Corollary 2.6 correspond to the flatness of the Riemannian metric induced by the free entropy (see [**14**, Section 4]).

The following example shows that the support assumptions in Corollaries 2.4, 2.6 and Proposition 2.5 cannot be removed.

EXAMPLE 2.7. Let $K := [-R, R]$ and ν be the *arcsine law* on $[-R, R]$, i.e.

$$\nu := \frac{1}{\pi\sqrt{R^2 - x^2}}\chi_{(-R,R)}(x)\,dx. \tag{2.14}$$

We notice $Q_\nu(x) \equiv 2\log(R/2)$ and $\Sigma(\nu) = \log(R/2)$. For each $r > 0$ set $h_r \in C_{\boldsymbol{R}}([-R, R])$ by $h_r(x) := 2x^2/r^2$. For $0 < r \le R$ let w_r be the *semicircle law* of radius r, i.e.

$$w_r(x) := \frac{2}{\pi r^2}\sqrt{r^2 - x^2}\chi_{[-r,r]}(x)\,dx.$$

Since $Q_{w_r}(x) = 2\int_{-r}^{r} w_r(y)\log|x - y|\,dy$ satisfies (see [**18**, Section IV.5])

$$Q_{w_r}(x) = \frac{2x^2}{r^2} + 2\log\frac{r}{2} - 1 \quad \text{if } |x| \le r$$

and

$$\begin{aligned} Q_{w_r}(x) &= \frac{2x^2}{r^2} + 2\log\left|\frac{x + \sqrt{x^2 - r^2}}{2}\right| - \frac{2|x|}{r^2}\sqrt{x^2 - r^2} - 1 \\ &< \frac{2x^2}{r^2} + 2\log\frac{r}{2} - 1 \quad \text{if } |x| > r, \end{aligned}$$

we see that $w_r = \nu^{h_r}$ when $0 < r \le R$. Furthermore, note that Q_{w_r} is continuous on $[-R, R]$. On the other hand, for $r \ge R$ let μ_r be a convex combination of w_R and ν given by

$$\mu_r := \left(1 - \frac{R^2}{r^2}\right)\nu + \frac{R^2}{r^2}w_R.$$

Recall ([**10**, Proposition 3.3]) that μ_r is a unique maximizer of the free entropy $\Sigma(\mu)$ under the constraint that μ is supported in $[-R, R]$ and $\int x^2\,d\mu(x) \le R^2/2 - R^4/4r^2$. Since

$$\begin{aligned} Q_{\mu_r}(x) &= \left(1 - \frac{R^2}{r^2}\right)Q_\nu(x) + \frac{R^2}{r^2}Q_{w_R}(x) \\ &= \frac{2x^2}{r^2} + 2\log\frac{R}{2} - \frac{R^2}{r^2} \quad \text{for } |x| \le R, \end{aligned}$$

we have $\mu_r = \nu^{h_r}$ when $r \ge R$.

One can explicitly compute

$$\Sigma(\nu^{h_r}, \nu) = -\Sigma(\nu^{h_r}) - \Sigma(\nu) + 2\log\frac{R}{2} = \begin{cases} \log(R/r) + 1/4 & \text{if } 0 < r \le R, \\ R^4/4r^4 & \text{if } r \ge R, \end{cases}$$

$$\frac{\nu(h_r) - \nu^{h_r}(h_r)}{2} = \begin{cases} R^2/2r^2 - 1/4 & \text{if } 0 < r \le R, \\ R^4/4r^4 & \text{if } r \ge R, \end{cases}$$

$$c(h_r, \nu) = \begin{cases} \log(r/R) - 3/4 & \text{if } 0 < r \le R, \\ R^4/4r^4 - R^2/r^2 & \text{if } r \ge R. \end{cases}$$

Hence all equalities in Corollaries 2.4 and 2.6 are confirmed when $r \ge R$ (hence $\operatorname{supp} \nu^{h_r} = \operatorname{supp} \nu = [-R, R]$). Since $\log(R/r) + 1/4 < R^2/2r^2 - 1/4$ if $0 < r < R$, the assumption $\operatorname{supp} \nu \subset \operatorname{supp} \nu^h$ in Corollary 2.4 (also Corollary 2.6) is essential. For $r \ge R$ one has $\nu = (\mu_r)^{-h_r} = (\nu^{h_r})^{-h_r}$ because $Q_\nu(x) = Q_{\mu_r}(x) - h_r(x) + R^2/r^2$ for $|x| \le R$. However, for $0 < r < R$ one cannot choose a constant $B \in \boldsymbol{R}$ such that $Q_\nu(x) = Q_{w_r}(x) - h_r(x) + B$ for $|x| \le R$; hence $\nu \ne (w_r)^{-h_r} = (\nu^{h_r})^{-h_r}$. This says that the assumption $\operatorname{supp}(\nu^h)^k \subset \operatorname{supp} \nu^h$ in Proposition 2.5 is essential. But, of course, $\nu = (w_r)^h$ for $h := -Q_{w_r}$. So, both $\operatorname{supp} \nu \subset \operatorname{supp} \nu^h$ and $\operatorname{supp} \nu \supset \operatorname{supp} \nu^h$ can occur in general.

The next proposition gives a simple sufficient condition for $\mu \in \mathscr{M}(K)$ to be a perturbed probability measure of ν.

Proposition 2.8. *If $\mu \in \mathscr{M}(K)$ satisfies $\mu \le \alpha\nu$ for some constant $\alpha \ge 1$, then $Q_\mu(x) := 2\int \log|x - y|\, d\mu(y)$ is continuous on K and there exists an $h \in C_{\boldsymbol{R}}(K)$ such that $\mu = \nu^h$.*

Proof. It is enough to prove the first assertion. Indeed, we then have $h := Q_\mu - Q_\nu \in C_{\boldsymbol{R}}(K)$ and

$$2\int \log|x - y|\, d\mu(y) = Q_\mu(x) = Q_\nu(x) + h(x) \quad \text{for all } x \in K,$$

showing $\mu = \nu^h$.

Let $x_0 \in K$ and $\varepsilon > 0$. Since $y \mapsto \log|y - x_0|$ is integrable with respect to ν, one can choose $0 < \delta < 1$ such that

$$(0\ge) \int_{K \cap \{|y - x_0| \le \delta\}} \log|y - x_0|\, d\nu(y) \ge -\varepsilon. \tag{2.15}$$

Furthermore, one can choose $\delta_0 > 0$ such that if $x \in K$ and $|x - x_0| \le \delta_0$, then

$$\frac{1}{2}|Q_\nu(x) - Q_\nu(x_0)| \le \varepsilon, \tag{2.16}$$

$$\left| \int_{K \cap \{|y - x| > \delta\}} \log|y - x|\, d\nu(y) - \int_{K \cap \{|y - x_0| > \delta\}} \log|y - x_0|\, d\nu(y) \right| \le \varepsilon, \tag{2.17}$$

$$\left| \int_{K \cap \{|y - x| > \delta\}} \log|y - x|\, d\mu(y) - \int_{K \cap \{|y - x_0| > \delta\}} \log|y - x_0|\, d\mu(y) \right| \le \varepsilon. \tag{2.18}$$

In the above, the estimate (2.17) follows because the Lebesgue bounded convergence theorem yields

$$\lim_{x\to x_0}\int_{K\cap\{|y-x_0|>\delta\}}\log|y-x|\,d\nu(y)=\int_{K\cap\{|y-x_0|>\delta\}}\log|y-x_0|\,d\nu(y)$$

and

$$\lim_{x\to x_0}\left(\int_{K\cap\{|y-x|>\delta\}}-\int_{K\cap\{|y-x_0|>\delta\}}\right)\log|y-x|\,d\nu(y)=0.$$

The estimate (2.18) is similar. Assume $x\in K$ and $|x-x_0|\le\delta_0$; then by (2.16) and (2.17)

$$\begin{aligned}
&\left|\int_{K\cap\{|y-x|\le\delta\}}\log|y-x|\,d\nu(y)-\int_{K\cap\{|y-x_0|\le\delta\}}\log|y-x_0|\,d\nu(y)\right|\\
&\quad\le|Q_\nu(x)-Q_\nu(x_0)|\\
&\qquad+\left|\int_{K\cap\{|y-x|>\delta\}}\log|y-x|\,d\nu(y)-\int_{K\cap\{|y-x_0|>\delta\}}\log|y-x_0|\,d\nu(y)\right|\\
&\quad\le 2\varepsilon,
\end{aligned}$$

and hence

$$\int_{K\cap\{|y-x|\le\delta\}}\log|y-x|\,d\nu(y)\ge\int_{K\cap\{|y-x_0|\le\delta\}}\log|y-x_0|\,d\nu(y)-2\varepsilon\ge-3\varepsilon \tag{2.19}$$

thanks to (2.15). From $\delta<1$ and the assumption $\mu\le\alpha\nu$ we have

$$\begin{aligned}
&\frac{1}{2}|Q_\mu(x)-Q_\mu(x_0)|\\
&\quad\le\left|\int_{K\cap\{|y-x|>\delta\}}\log|y-x|\,d\mu(y)-\int_{K\cap\{|y-x_0|>\delta\}}\log|y-x_0|\,d\mu(y)\right|\\
&\qquad-\int_{K\cap\{|y-x|\le\delta\}}\log|y-x|\,d\mu(y)-\int_{K\cap\{|y-x_0|\le\delta\}}\log|y-x_0|\,d\mu(y)\\
&\quad\le\varepsilon-\alpha\int_{K\cap\{|y-x|\le\delta\}}\log|y-x|\,d\nu(y)-\alpha\int_{K\cap\{|y-x_0|\le\delta\}}\log|y-x_0|\,d\nu(y)\\
&\quad\le\varepsilon+3\alpha\varepsilon+\alpha\varepsilon=(1+4\alpha)\varepsilon
\end{aligned}$$

by (2.18), (2.19) and (2.15). Hence $Q_\mu(x)$ is continuous at each $x_0\in K$. $\square$

Under the assumption $\mu \le \alpha\nu$ in the above proposition, Q_μ and Q_ν can be compared as follows.

LEMMA 2.9. *If $\mu \in \mathscr{M}(K)$ satisfies $\mu \le \alpha\nu$ for some $\alpha \ge 1$, then*

$$Q_\mu(x) \ge \alpha Q_\nu(x) + 2(1-\alpha)\log R \quad (x \in K),$$

where R is the diameter of K.

PROOF. Choose $c \in \boldsymbol{R}$ such that $K \subset [c - R/2, c + R/2]$, and transform K and μ, ν to K' ($\subset [-1/2, 1/2]$) and $\mu', \nu' \in \mathscr{M}(K')$ via the affine transformation $x \mapsto (x-c)/R$, so $\mu \le \alpha\nu$ implies $\mu' \le \alpha\nu'$. Then for every $x \in K'$ we estimate

$$\begin{aligned}
Q_\mu(c+Rx) &= 2\int_K \log|(c+Rx) - y|\,d\mu(y) \\
&= 2\int_{K'} \log|(c+Rx) - (c+Ry)|\,d\mu'(y) \\
&= 2\left(\int_{K'} \log|x-y|\,d\mu'(y) + \log R\right) \\
&\ge 2\left(\alpha\int_{K'} \log|x-y|\,d\nu'(y) + \log R\right) \\
&= 2\left(\alpha\left(\int_K \log|(c+Rx)-y|\,d\nu(y) - \log R\right) + \log R\right) \\
&= \alpha Q_\nu(c+Rx) + 2(1-\alpha)\log R,
\end{aligned}$$

showing the desired conclusion. □

COROLLARY 2.10. *If $\mu \in \mathscr{M}(K)$ satisfies $\beta\nu \le \mu \le \alpha\nu$ for some constants $0 < \beta \le 1 \le \alpha$, then there exists an $h \in C_{\boldsymbol{R}}(K)$ such that $\mu = \nu^h$ and*

$$(1-\alpha)(2\log R - Q_\nu) \le h \le (1-\beta)(2\log R - Q_\nu),$$

where R is the diameter of K. (Note $Q_\nu \le 2\log R$.)

PROOF. By Proposition 2.8 and Lemma 2.9, $Q_\mu \in C_{\boldsymbol{R}}(K)$ and

$$Q_\mu(x) \ge \alpha Q_\nu(x) + 2(1-\alpha)\log R \quad (x \in K).$$

Since $\nu \le \beta^{-1}\mu$, the roles of μ, ν can be interchanged so that

$$Q_\nu(x) \ge \beta^{-1}Q_\mu(x) + 2(1-\beta^{-1})\log R,$$

that is,

$$Q_\mu(x) \le \beta Q_\nu(x) + 2(1-\beta)\log R \quad (x \in K).$$

Hence the required estimates are satisfied for $h := Q_\mu - Q_\nu$, and we have $\mu = \nu^h$ for this h. □

COROLLARY 2.11. *If $\mu \in \mathcal{M}(K)$ satisfies $\beta\nu \le \mu \le \alpha\nu$ for some $0 < \beta \le 1 \le \alpha$, then*

$$\Sigma(\mu, \nu) \le (\alpha(\alpha - 1) + (1 - \beta))(\log R - \Sigma(\nu)).$$

PROOF. Corollary 2.10 gives

$$\begin{aligned}\nu(h) &\le (1 - \beta)(2\log R - \nu(Q_\nu)) = 2(1 - \beta)(\log R - \Sigma(\nu)),\\ \mu(h) &\ge (1 - \alpha)\mu(2\log R - Q_\nu)\\ &\ge \alpha(1 - \alpha)\nu(2\log R - Q_\nu) = 2\alpha(1 - \alpha)(\log R - \Sigma(\nu)).\end{aligned}$$

Hence by Corollary 2.4 we have

$$\Sigma(\mu, \nu) \le \frac{\nu(h) - \mu(h)}{2} \le (\alpha(\alpha - 1) + (1 - \beta))(\log R - \Sigma(\nu)).$$

(Note $\Sigma(\nu) \le \log(R/4)$, see Example 2.7). □

3. Convergence of perturbed measures.

As in the previous section, let K be a compact subset of $\boldsymbol{R}$ and $\nu \in \mathcal{M}(K)$ be such that $Q = Q_\nu \in C_{\boldsymbol{R}}(K)$. The aim of this section is to show the continuity properties (with respect to h) of the perturbation ν^h introduced in the previous section.

Set

$$\mathcal{M}_\Sigma(K) := \{\mu \in \mathcal{M}(K) : \Sigma(\mu) > -\infty\},$$

and for $\mu_1, \mu_2 \in \mathcal{M}_\Sigma(K)$ define

$$d(\mu_1, \mu_2) := \Sigma(\mu_1, \mu_2)^{1/2} \in [0, +\infty).$$

The next lemma is an application of the series expansions of the function $x \mapsto \int \log|x - y|\, d\mu(y)$ and of the free entropy $\Sigma(\mu)$ due to Haagerup [**8**], and it will play a key role in the proof of the following theorem.

LEMMA 3.1. *The above defined $d(\mu_1, \mu_2)$ is a metric on $\mathcal{M}_\Sigma(K)$ and the d-topology is stronger than the weak topology (restricted on $\mathcal{M}_\Sigma(K)$).*

PROOF. The free relative entropy $\Sigma(\mu_1, \mu_2)$ is symmetric and strictly positive as stated at the end of Section 1. To prove the triangular inequality of $\Sigma(\mu_1, \mu_2)^{1/2}$, we may assume $K = [-R, R]$ without loss of generality. First, assume $K = [-2, 2]$ and set

$$L_r(x, y) := -\sum_{n=1}^{\infty} \frac{2r^n}{n} T_n\left(\frac{x}{2}\right) T_n\left(\frac{y}{2}\right)$$

for $0 < r < 1$ and $x, y \in [-2, 2]$, where T_n's are the Chebyshev polynomials of the first kind. In [**8**] Haagerup estimated

$$\begin{aligned} 2\log 2 \geq L_r(x, y) &= \frac{1}{2}\log((1-r^2)^2 + r^2(x^2+y^2) - r(1+r^2)xy) \\ &\geq \log|x-y| + 2\log\frac{1+r}{2}, \end{aligned}$$

and showed the series expansion

$$\Sigma(\mu) = -\sum_{n=1}^{\infty} \frac{2}{n}\left(\int_{-2}^{2} T_n\left(\frac{x}{2}\right) d\mu(x)\right)^2 \tag{3.1}$$

for every $\mu \in \mathscr{M}([-2, 2])$. When $\mu_1, \mu_2 \in \mathscr{M}_\Sigma([-2, 2])$, since $\log|x-y|$ is integrable with respect to $d|\mu_1 - \mu_2|(x)\, d|\mu_1 - \mu_2|(y)$ and $L_r(x, y) \to \log|x-y|$ as $r \nearrow 1$, the Lebesgue convergence theorem yields

$$\begin{aligned} \Sigma(\mu_1, \mu_2) &= -\iint \log|x-y|\, d(\mu_1-\mu_2)(x)\, d(\mu_1-\mu_2)(y) \\ &= \lim_{r\nearrow 1}\left[-\iint L_r(x, y)\, d(\mu_1-\mu_2)(x)\, d(\mu_1-\mu_2)(y)\right] \\ &= \sum_{n=1}^{\infty} \frac{2}{n}\left(\int_{-2}^{2} T_n\left(\frac{x}{2}\right) d(\mu_1-\mu_2)(x)\right)^2. \end{aligned}$$

For $\mu_1, \mu_2 \in \mathscr{M}_\Sigma([-R, R])$ let $\tilde{\mu}_i := \mu_i((R/2)\cdot) \in \mathscr{M}_\Sigma([-2, 2])$. Then, since $\Sigma(\mu_1, \mu_2) = \Sigma(\tilde{\mu}_1, \tilde{\mu}_2)$, the above formula is transformed to

$$\Sigma(\mu_1, \mu_2) = \sum_{n=1}^{\infty} \frac{2}{n}\left(\int_{-R}^{R} T_n\left(\frac{x}{R}\right) d(\mu_1-\mu_2)(x)\right)^2. \tag{3.2}$$

Now the triangular inequality of $\Sigma(\mu_1, \mu_2)^{1/2}$ is obvious. If $\mu, \mu_k \in \mathscr{M}_\Sigma([-R, R])$ and $\Sigma(\mu_k, \mu) \to 0$, then one can see from (3.2) that $\mu_k(p) \to \mu(p)$ for any polynomial p, which says that $\mu_k \to \mu$ in w*-topology. So the d-topology is stronger than the w*-topology (or the weak topology). □

Remark 3.2. Concerning the above metric $d(\mu_1, \mu_2)$ on $\mathscr{M}_\Sigma(K)$ it is worth noting that the d-topology is strictly stronger than the w*-topology and $(\mathscr{M}_\Sigma(K), d)$ is a non-compact Polish space. Indeed, we may assume $K = [-2, 2]$, and let ν

be the arcsine law (2.14) with $R=2$. For $0<\delta,\varepsilon<1$ let ν_1 be the uniform distribution on $[2-\delta,2]$ and $\mu_1 := (1-\varepsilon)\nu+\varepsilon\nu_1$. Then $\mu_1 \in \mathcal{M}_\Sigma([-2,2])$ and $\|\mu_1-\nu\| \le 2\varepsilon$. But, since $\int_{-2}^2 T_n(x/2)\,d\nu(x)=0$ for all $n\in\boldsymbol{N}$, we get

$$\Sigma(\mu_1,\nu) = \varepsilon^2\sum_{n=1}^\infty \frac{2}{n}\left(\int_{-2}^2 T_n\left(\frac{x}{2}\right)d(\nu_1-\nu)(x)\right)^2$$

$$= \varepsilon^2\sum_{n=1}^\infty \frac{2}{n}\left(\frac{1}{\delta}\int_{2-\delta}^2 T_n\left(\frac{x}{2}\right)dx\right)^2,$$

which can be arbitrarily large as $\delta\to+0$ (for any ε fixed). Therefore, one can choose a sequence $\{\mu_k\}$ in $\mathcal{M}_\Sigma([-2,2])$ such that $\|\mu_k-\nu\|\to 0$ (hence $\mu_k\to\nu$ in w*-topology) but $\Sigma(\mu_k,\nu)\to+\infty$. (Incidentally, one can get a sequence $\{\mu_k'\}$ in $\mathcal{M}_\Sigma([-2,2])$ such that $\Sigma(\mu_k',\nu)\to 0$ and $\|\mu_k'-\nu\|\not\to 0$, unlike the relative entropy case, see [**4**], [**9**].)

Next, let $\{\mu_k\}$ be a d-Cauchy sequence in $\mathcal{M}_\Sigma([-2,2])$. By (3.2) this means that

$$\left(\int_{-2}^2 T_n\left(\frac{x}{2}\right)d\mu_k(x)\right)_{n=1}^\infty \quad (k\in\boldsymbol{N})$$

form a Cauchy sequence in $\ell^2(1/n)$, the ℓ^2-space with respect to the sequence $(1/n)$. Hence the above sequence converges as $k\to\infty$ to some $(a_n)\in\ell^2(1/n)$ in the norm of $\ell^2(1/n)$, so the limit $\mu(p) := \lim_{k\to\infty}\mu_k(p)$ exists for every polynomial p. Since $|\mu(p)|\le\|p\|_\infty$, μ extends to a bounded linear functional on $C_{\boldsymbol{R}}([-2,2])$ and $\mu_k\to\mu$ in w*-topology. We get

$$a_n = \lim_{k\to\infty}\int_{-2}^2 T_n\left(\frac{x}{2}\right)d\mu_k(x) = \int_{-2}^2 T_n\left(\frac{x}{2}\right)d\mu(x) \quad (n\in\boldsymbol{N})$$

so that by (3.1)

$$\Sigma(\mu) = -\sum_{n=1}^\infty\frac{2}{n}\left(\int_{-2}^2 T_n\left(\frac{x}{2}\right)d\mu(x)\right)^2 = -\sum_{n=1}^\infty\frac{2}{n}a_n^2 > -\infty.$$

This implies that $\mu\in\mathcal{M}_\Sigma([-2,2])$ and $d(\mu_k,\mu)\to 0$. Since $(\mathcal{M}_\Sigma([-2,2]),d)$ is isometrically imbedded in $\ell^2(1/n)$, it is separable. Furthermore, since $\mathcal{M}_\Sigma([-2,2])$ is a w*-dense proper subset of $\mathcal{M}([-2,2])$ as easily seen, there is a sequence in $\mathcal{M}_\Sigma([-2,2])$ converging to an element in $\mathcal{M}([-2,2])\backslash\mathcal{M}_\Sigma([-2,2])$ in w*-topology. Then this cannot have a subsequence converging in d-topology, so $(\mathcal{M}_\Sigma([-2,2]),d)$ is not compact.

THEOREM 3.3. *If $h, h_n \in C_{\boldsymbol{R}}(K)$, $n \in \boldsymbol{N}$, satisfy $\|h_n - h\| \to 0$, then the following convergences hold*:

(i) $c(h_n, \nu) \to c(h, \nu)$.

(ii) $\Sigma(\nu^{h_n}, \mu) \to \Sigma(\nu^h, \mu)$ *for every* $\mu \in \mathscr{M}_\Sigma(K)$; *in particular*, $\Sigma(\nu^{h_n}, \nu^h) \to 0$.

(iii) $\nu^{h_n} \to \nu^h$ *weakly*.

(iv) $\nu^{h_n}(h_n) \to \nu^h(h)$.

(v) $\Sigma(\nu^{h_n}) \to \Sigma(\nu^h)$.

PROOF. (i) is obvious from (2.3).

(ii) By (2.9) we have

$$0 \le \Sigma(\nu^h, \nu^{h_n}) \le \Sigma(\nu^h, \nu) + \nu^h(h_n) + c(h_n, \nu),$$

and this right-hand side tends to

$$\Sigma(\nu^h, \nu) + \nu^h(h) + c(h, \nu) = 0$$

thanks to (i) and (2.5). Hence $\Sigma(\nu^{h_n}, \nu^h) \to 0$. For every $\mu \in \mathscr{M}_\Sigma(K)$, since Lemma 3.1 implies that

$$|\Sigma(\nu^{h_n}, \mu)^{1/2} - \Sigma(\nu^h, \mu)^{1/2}| \le \Sigma(\nu^{h_n}, \nu^h)^{1/2},$$

we get $\Sigma(\nu^{h_n}, \mu) \to \Sigma(\nu^h, \mu)$.

(iii) is a consequence of (ii) and Lemma 3.1.

(iv) follows from

$$\begin{aligned} |\nu^{h_n}(h_n) - \nu^h(h)| &\le |\nu^{h_n}(h_n) - \nu^{h_n}(h)| + |\nu^{h_n}(h) - \nu^h(h)| \\ &\le \|h_n - h_n\| + |\nu^{h_n}(h) - \nu^h(h)| \to 0 \end{aligned}$$

thanks to (iii).

(v) We apply (2.6) to have

$$\begin{aligned} \Sigma(\nu^{h_n}) &= c(h_n, \nu) - \Sigma(\nu) + \nu^{h_n}(Q + h_n) \\ &\to c(h, \nu) - \Sigma(\nu) + \nu^h(Q + h) = \Sigma(\nu^h) \end{aligned}$$

due to (i), (iii) and (iv). □

Concerning the perturbation $\nu^{h,S}$ via relative entropy, the continuity of $h \mapsto \nu^{h,S}$ can be straightforwardly seen from the explicit formula $\nu^{h,S} = (e^{-h}/\nu(e^{-h}))\nu$. In fact, when $h, h_n \in C_{\boldsymbol{R}}(K)$ and $h_n \to h$ boundedly pointwise, i.e. $\sup_n \|h_n\| < +\infty$ and $h_n(x) \to h(x)$ for every $x \in K$, one gets the w*-convergence $\nu^{h_n,S} \to \nu^{h,S}$ by the Lebesgue bounded convergence theorem.

The next proposition says that the weak convergence and the d-convergence are equivalent for a sequence $\{\mu_n\}$ in $\mathcal{M}(K)$ such that μ_n's are uniformly dominated by ν.

Proposition 3.4. *Let $\mu, \mu_n \in \mathcal{M}(K)$, $n \in \boldsymbol{N}$, and assume that there is an $\alpha \geq 1$ such that $\mu_n \leq \alpha\nu$ for all $n \in \boldsymbol{N}$. Then $\mu_n \to \mu$ weakly if and only if $\Sigma(\mu_n, \mu) \to 0$. In this case, $\Sigma(\mu_n) \to \Sigma(\mu)$ and $\Sigma(\mu_n, \mu') \to \Sigma(\mu, \mu')$ for every $\mu' \in \mathcal{M}_\Sigma(K)$.*

Proof. First, assume that $\Sigma(\mu_n, \mu) \to 0$. Proposition 2.8 implies that $Q_{\mu_n} \in C_{\boldsymbol{R}}(K)$ for $n \in \boldsymbol{N}$. Hence, by (2.7) with μ_n in place of ν, we get

$$\Sigma(\mu, \mu_n) = -\Sigma(\mu) - \Sigma(\mu_n) + \mu(Q_{\mu_n}). \tag{3.3}$$

This implies $\mu \in \mathcal{M}_\Sigma(K)$ as well as $\mu_n \in \mathcal{M}_\Sigma(K)$. Hence the weak convergence $\mu_n \to \mu$ follows from Lemma 3.1.

Conversely, assume that $\mu_n \to \mu$ weakly. Then $\mu \leq \alpha\nu$ holds as well, so $Q_\mu \in C_{\boldsymbol{R}}(K)$ by Proposition 2.8. Since as (3.3) we have

$$\Sigma(\mu_n, \mu) = -\Sigma(\mu_n) - \Sigma(\mu) + \mu_n(Q_\mu),$$

it suffices for $\Sigma(\mu_n, \mu) \to 0$ to prove $\Sigma(\mu_n) \to \Sigma(\mu)$. Indeed, we then obtain

$$\Sigma(\mu_n, \mu) \to -2\Sigma(\mu) + \mu(Q_\mu) = 0.$$

Now let us prove that $\Sigma(\mu_n) \to \Sigma(\mu)$. For any $\varepsilon > 0$ choose $0 < \delta < 1$ such that

$$\iint_{\{|x-y|<\delta\}} |\log|x-y|| \, d\nu(x) \, d\nu(y) \leq \varepsilon.$$

We estimate

$$\begin{aligned}
|\Sigma(\mu_n) - \Sigma(\mu)| &\leq \left| \iint \log|x-y| \, d(\mu_n - \mu)(x) \, d\mu_n(y) \right| \\
&\quad + \left| \iint \log|x-y| \, d\mu(x) \, d(\mu_n - \mu)(y) \right| \\
&\leq \int \left| \int \log|x-y| \, d(\mu_n - \mu)(x) \right| d\mu_n(y) \\
&\quad + \int \left| \int \log|x-y| \, d(\mu_n - \mu)(y) \right| d\mu(x) \\
&\leq 2\alpha \int \left| \int \log|x-y| \, d(\mu_n - \mu)(y) \right| d\nu(x)
\end{aligned}$$

and

$$\left|\int \log|x-y|\,d(\mu_n-\mu)(y)\right| \le \left|\int \log(|x-y| \vee \delta)\,d(\mu_n-\mu)(y)\right|$$
$$+\int(\log(|x-y| \vee \delta) - \log|x-y|)\,d(\mu_n+\mu)(y)$$
$$\le \left|\int \log(|x-y| \vee \delta)\,d(\mu_n-\mu)(y)\right|$$
$$+\,2\alpha\int_{\{|x-y|<\delta\}} |\log|x-y||\,dv(y).$$

Therefore,

$$|\Sigma(\mu_n)-\Sigma(\mu)| \le 2\alpha\int\left|\int \log(|x-y| \vee \delta)\,d(\mu_n-\mu)(y)\right| dv(x)$$
$$+\,4\alpha^2\iint_{\{|x-y|<\delta\}} |\log|x-y||\,dv(x)\,dv(y).$$

Since $\mu_n \to \mu$ weakly, we have

$$\lim_{n\to\infty}\int \log(|x-y| \vee \delta)\,d(\mu_n-\mu)(y) = 0 \quad (x \in K)$$

and

$$\sup_{x\in K}\left|\int \log(|x-y| \vee \delta)\,d(\mu_n-\mu)(y)\right| < +\infty,$$

so the Lebesgue bounded convergence theorem yields

$$\lim_{n\to\infty}\int\left|\int \log(|x-y| \vee \delta)\,d(\mu_n-\mu)(y)\right| dv(x) = 0.$$

Hence

$$\limsup_{n\to\infty}|\Sigma(\mu_n)-\Sigma(\mu)| \le 4\alpha^2\varepsilon,$$

implying $\Sigma(\mu_n) \to \Sigma(\mu)$.

It remains to show that $\Sigma(\mu_n,\mu') \to \Sigma(\mu,\mu')$ for every $\mu' \in \mathcal{M}_\Sigma(K)$ whenever $\mu_n,\mu \in \mathcal{M}_\Sigma(K)$ and $\Sigma(\mu_n,\mu) \to 0$. But this is immediate from Lemma 3.1 as in the proof (ii) of Theorem 3.3. $\square$

As for relative entropy, it is known that if μ_n, v_n are probability measures on $\boldsymbol{R}$ such that $\|\mu_n-\mu\| \to 0$, $\|v_n-v\| \to 0$ and there is an $\alpha > 0$ such that $\mu_n \le \alpha v_n$

for all $n \in \boldsymbol{N}$, then $S(\mu_n, \nu_n) \to S(\mu, \nu)$. (This is true in the operator algebra setting, see [**1**, Theorem 3.7].) However, this fails to hold for free relative entropy; one can easily provide an example of $\mu_n, \nu_n \in \mathscr{M}_\Sigma(K)$ such that $\|\mu_n - \nu\| \to 0$, $\|\nu_n - \nu\| \to 0$ and $\mu_n \le \alpha\nu_n$ for all $n \in \boldsymbol{N}$, but $\Sigma(\mu_n, \nu_n) \not\to 0$.

4. From relative entropy to free relative entropy.

In this section, given ν and h as before, for each n we consider an $n \times n$ selfadjoint random matrix naturally perturbed via relative entropy, and show that the perturbed measure ν^h via free relative entropy is the limit distribution of the empirical eigenvalue distributions of perturbed random matrices as the size n goes to ∞. In so doing, we can also express the free relative entropy $\Sigma(\nu^h, \nu)$ as the limit (with normalization) of the relative entropy defined on the matrix space M_n^{sa}.

Throughout this section, we assume for simplicity that K is a finite interval $[-R, R]$. Let $\nu \in \mathscr{M}([-R, R])$ be fixed so that $Q = Q_\nu$ in (1.14) is a continuous function on $[-R, R]$. For each $n \in \boldsymbol{N}$ we simply write $\lambda_n(\nu)$ for the probability measure $\lambda_n(\nu; R) = \lambda_n(Q; R)$ on $(M_n^{sa})_R$ given in (1.9)–(1.12). Here note that $(M_n^{sa})_R$ is a compact subset of M_n^{sa} being identified with a Euclidean space $\boldsymbol{R}^{n^2}$. For a given $h \in C_{\boldsymbol{R}}([-R, R])$ and $n \in \boldsymbol{N}$, let $\phi_n(h)$ denote a real continuous function on $(M_n^{sa})_R$ defined by

$$\phi_n(h)(A) := n^2 \operatorname{tr}_n(h(A)) \quad \text{for } A \in (M_n^{sa})_R, \tag{4.1}$$

where $h(A)$ is defined via functional calculus and tr_n is the normalized trace on M_n. Then one can get the probability measure $\lambda_n(\nu)^{\phi_n(h), S}$ on $(M_n^{sa})_R$ which is the perturbed measure of $\lambda_n(\nu)$ by $\phi_n(h)$ via relative entropy; namely, $\lambda_n(\nu)^{\phi_n(h), S}$ is a unique maximizer of the functional

$$-\eta(\phi_n(h)) - S(\eta, \lambda_n(\nu)) \quad \text{for } \eta \in \mathscr{M}((M_n^{sa})_R),$$

where $\mathscr{M}((M_n^{sa})_R)$ is the set of all probability Borel measures on $(M_n^{sa})_R$. In fact, as mentioned after Theorem 2.2, it is given by

$$\lambda_n(\nu)^{\phi_n(h), S} = \frac{e^{-\phi_n(h)}}{\lambda_n(\nu)(e^{-\phi_n(h)})} \lambda_n(\nu) \tag{4.2}$$

and

$$-\lambda_n(\nu)^{\phi_n(h), S}(\phi_n(h)) - S(\lambda_n(\nu)^{\phi_n(h), S}, \lambda_n(\nu)) = \log \lambda_n(\nu)(e^{-\phi_n(h)}). \tag{4.3}$$

In the sequel we use the following notations for short:

$$\Delta(x) := \prod_{i<j} (x_i - x_j)^2, \quad dx := dx_1 dx_2 \cdots dx_n.$$

Lemma 4.1. *With the above notations,*

$$\lambda_n(v)^{\phi_n(h), S} = \lambda_n(Q + h; R),$$

that is, $\lambda_n(v)^{\phi_n(h), S}$ *is invariant under unitary conjugation and its joint eigenvalue distribution is*

$$\tilde{\lambda}_n(Q + h; R) = \frac{1}{Z_n(Q + h; R)} \exp\left(-n \sum_{i=1}^{n} (Q(x_i) + h(x_i))\right) \Delta(x) \prod_{i=1}^{n} \chi_{[-R, R]}(x_i)\, dx, \tag{4.4}$$

where $Z_n(Q + h; R)$ *is defined by* (1.10) *with* $Q + h$ *in place of* Q. *Furthermore,*

$$\lambda_n(v)(e^{-\phi_n(h)}) = \frac{Z_n(Q + h; R)}{Z_n(Q; R)}. \tag{4.5}$$

Proof. Since it is obvious from the definition (4.1) that $\phi_n(h)$ is invariant under unitary conjugation, so is the measure $\lambda_n(v)^{\phi_n(h), S}$ due to the expression (4.2). Since $\lambda_n(v) = \lambda_n(Q; R)$ has the joint eigenvalue distribution (1.9) and by (4.1)

$$\phi_n(h)(A) = n \sum_{i=1}^{n} h(x_i)$$

for the eigenvalues $x_1, x_2, \ldots, x_n$ of $A \in (M_n^{sa})_R$, we get

$$\begin{aligned}
\lambda_n(v)(e^{-\phi_n(h)}) &= \frac{1}{Z_n(Q; R)} \int_{-R}^{R} \cdots \int_{-R}^{R} \exp\left(-n \sum_{i=1}^{n} (Q(x_i) + h(x_i))\right) \Delta(x)\, dx \\
&= \frac{Z_n(Q + h; R)}{Z_n(Q; R)}.
\end{aligned}$$

This and (4.2) imply that the joint eigenvalue distribution of $\lambda_n(v)^{\phi_n(h), S}$ is (4.4). □

The measure $\lambda_n(v)^{\phi_n(h), S}$ on $(M_n^{sa})_R$ may be considered as an $n \times n$ selfadjoint random matrix which is a perturbation of $\lambda_n(v)$ via relative entropy. The next theorem says that this perturbation of $\lambda_n(v)$ via relative entropy on the matrix space approaches asymptotically as $n \to \infty$ to v^h $(= v^{h, \Sigma})$, the perturbation of v via free relative entropy. In particular, it justifies our formulation of free relative entropy. In the theorem we actually treat a sequence of perturbed measures $\lambda_n(v)^{\phi_n(h_n), S}$ determined by separate $h_n \in C_{\mathbf{R}}([-R, R])$ for each n satisfying $\|h_n - h\| \to 0$. The proof is based on the large deviation result presented in Theorem 1.1.

Theorem 4.2. *Let* $\nu \in \mathcal{M}([-R, R])$ *be as above. If* $h, h_n \in C_{\boldsymbol{R}}([-R, R])$, $n \in \boldsymbol{N}$, *satisfy* $\|h_n - h\| \to 0$, *then the following hold*:

(i) *The empirical eigenvalue distribution of* $\lambda_n(\nu)^{\phi_n(h_n), S}$ *converges almost surely to* ν^h *as* $n \to \infty$ *in weak topology.*

(ii)

$$\nu^h(h) = \lim_{n\to\infty} \frac{1}{n^2} \lambda_n(\nu)^{\phi_n(h_n), S}(\phi_n(h_n)). \tag{4.6}$$

(iii)

$$\Sigma(\nu^h, \nu) = \lim_{n\to\infty} \frac{1}{n^2} S(\lambda_n(\nu)^{\phi_n(h_n), S}, \lambda_n(\nu)). \tag{4.7}$$

(iv) *With* $B(Q; R)$ *defined by* (1.13) *and* $B(Q + h; R)$ *similarly with* $Q + h$ *in place of* Q,

$$c(h, \nu) = \lim_{n\to\infty} \frac{1}{n^2} \log \lambda_n(\nu)(e^{-\phi_n(h_n)}) = B(Q + h; R) - B(Q; R). \tag{4.8}$$

(v)

$$\nu(h) - \nu^h(h) - \Sigma(\nu^h, \nu) = \lim_{n\to\infty} \frac{1}{n^2} S(\lambda_n(\nu), \lambda_n(\nu)^{\phi_n(h_n), S}).$$

Hence, if $\operatorname{supp} \nu \subset \operatorname{supp} \nu^h$, *then*

$$\Sigma(\nu^h, \nu) = \lim_{n\to\infty} \frac{1}{n^2} S(\lambda_n(\nu), \lambda_n(\nu)^{\phi_n(h_n), S}).$$

Proof. First, note that ν is the minimizer of the rate function in Theorem 1.1, and the definition of ν^h in Section 2 means that ν^h is the minimizer of the rate function when Q is replaced by $Q + h$ in Theorem 1.1. (With the notation in Lemma 2.1, $\nu^h = \mu_{Q+h}$ as well as $\nu = \mu_Q$.) Hence Theorem 1.1 implies the following:

$$-\Sigma(\nu) + \nu(Q) + B(Q; R) = 0, \tag{4.9}$$

$$B(Q; R) = \lim_{n\to\infty} \frac{1}{n^2} \log Z_n(Q; R), \tag{4.10}$$

$$-\Sigma(\nu^h) + \nu^h(Q + h) + B(Q + h; R) = 0, \tag{4.11}$$

$$B(Q + h; R) = \lim_{n\to\infty} \frac{1}{n^2} \log Z_n(Q + h_n; R). \tag{4.12}$$

(i) By Lemma 4.1 the measure (or a selfadjoint random matrix) $\lambda_n(\nu)^{\phi_n(h_n), S}$ has the joint eigenvalue distribution $\tilde{\lambda}_n(Q + h_n; R)$ given in (4.4) with h_n in place of h. Theorem 1.1 (applied to $Q+h_n$ instead of Q_n) says that the empirical distribution of $\tilde{\lambda}_n(Q+h_n;R)$ converges almost surely to ν^h as $n \to \infty$ in weak topology. This implies the assertion.

(ii) By Lemma 4.1 and (4.1),

$$\frac{1}{n^2}\lambda_n(\nu)^{\phi_n(h_n),S}(\phi_n(h_n)) = \int_{-R}^{R}\cdots\int_{-R}^{R}\left(\frac{1}{n}\sum_{i=1}^{n} h_n(x_i)\right) d\tilde{\lambda}_n(Q+h_n;R)(x).$$

Theorem 1.1 tells us that if $(x_1,\dots,x_n)$ is distributed subject to $\tilde{\lambda}_n(Q+h_n;R)$, then for every $f \in C_{\mathbf{R}}([-R,R])$ the random variable

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i) = \left(\frac{\delta(x_1)+\cdots+\delta(x_n)}{n}\right)(f)$$

converges almost surely to $\nu^h(f)$ as $n \to \infty$. Since $\|h_n - h\| \to 0$, this shows that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} h_n(x_i) = \nu^h(h) \text{ almost surely}$$

so that (4.6) follows by the Lebesgue dominated convergence theorem.

(iii) Since by (4.3) and (4.5)

$$\begin{aligned}-\lambda_n(\nu)^{\phi_n(h_n),S}(\phi_n(h_n)) &- S(\lambda_n(\nu)^{\phi_n(h_n),S},\lambda_n(\nu)) \\ &= \log Z_n(Q+h_n;R) - \log Z_n(Q;R),\end{aligned}$$

it follows from (4.6), (4.10) and (4.12) that $\lim_{n\to\infty}(1/n^2)S(\lambda_n(\nu)^{\phi_n(h_n),S},\lambda_n(\nu))$ exists, and we have

$$\begin{aligned}-\nu^h(h) &- \lim_{n\to\infty}\frac{1}{n^2}S(\lambda_n(\nu)^{\phi_n(h_n),S},\lambda_n(\nu)) \\ &= B(Q+h;R) - B(Q;R) \\ &= \Sigma(\nu^h) - \nu^h(Q+h) - \Sigma(\nu) + \nu(Q) \qquad (4.13)\end{aligned}$$

thanks to (4.9) and (4.11). Therefore,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n^2}S(\lambda_n(\nu)^{\phi_n(h_n),S},\lambda_n(\nu)) &= -\Sigma(\nu^h) + \Sigma(\nu) + \nu^h(Q) - \nu(Q) \\ &= -\Sigma(\nu^h) - \Sigma(\nu) + \nu^h(Q) \\ &= \Sigma(\nu^h,\nu)\end{aligned}$$

due to (2.7). Hence (iii) follows.

(iv) By (2.5), (4.13), (4.12) and (4.5) we furthermore have

$$\begin{aligned} c(h,\nu) &= -\nu^h(h) - \Sigma(\nu^h,\nu) \\ &= B(Q+h;R) - B(Q;R) \\ &= \lim_{n\to\infty}\frac{1}{n^2}\log\frac{Z_n(Q+h_n;R)}{Z_n(Q;R)} \\ &= \lim_{n\to\infty}\frac{1}{n^2}\log\lambda_n(\nu)(e^{-\phi_n(h_n)}), \end{aligned} \tag{4.14}$$

as desired.

(v) By (2.13),

$$S(\lambda_n(\nu),\lambda_n(\nu)^{\phi_n(h_n),S}) = \lambda_n(\nu)(\phi_n(h_n)) + \log\lambda_n(\nu)(e^{-\phi_n(h_n)}).$$

Theorem 1.1 implies as (4.6) that

$$\lim_{n\to\infty}\frac{1}{n^2}\lambda_n(\nu)(\phi_n(h_n)) = \nu(h).$$

These together with (4.14) give

$$\lim_{n\to\infty}\frac{1}{n^2}S(\lambda_n(\nu),\lambda_n(\nu)^{\phi_n(h_n),S}) = \nu(h) - \nu^h(h) - \Sigma(\nu^h,\nu).$$

If $\operatorname{supp}\nu \subset \operatorname{supp}\nu^h$, then by Corollary 2.4

$$\nu(h) - \nu^h(h) - \Sigma(\nu^h,\nu) = \Sigma(\nu^h,\nu).$$

Hence (v) is obtained. □

Besides its conceptual importance, Theorem 4.2 supplies the asymptotic formulas of $\nu^h(h)$ and $c(h,\nu)$ (when $h_n = h$ for all n); thus we obtain the asymptotic formula of $\Sigma(\nu^h,\nu) = -\nu^h(h) - c(h,\nu)$. In particular, we state the following:

Corollary 4.3. *Let μ,ν be compactly supported probability measures on $\mathbf{R}$ such that Q_μ and Q_ν are continuous. Then for any $R>0$ with* $\operatorname{supp}\mu$, $\operatorname{supp}\nu \subset [-R,R]$,

$$\begin{aligned} &\Sigma(\mu,\nu) \\ &= \lim_{n\to\infty}\frac{\int_{-R}^R\cdots\int_{-R}^R((1/n)\sum_{i=1}^n(Q_\nu(x_i)-Q_\mu(x_i)))\exp(-n\sum_{i=1}^n Q_\mu(x_i))\Delta(x)\,dx}{\int_{-R}^R\cdots\int_{-R}^R\exp(-n\sum_{i=1}^n Q_\mu(x_i))\Delta(x)\,dx} \\ &\quad + \lim_{n\to\infty}\frac{1}{n^2}\log\frac{\int_{-R}^R\cdots\int_{-R}^R\exp(-n\sum_{i=1}^n Q_\nu(x_i))\Delta(x)\,dx}{\int_{-R}^R\cdots\int_{-R}^R\exp(-n\sum_{i=1}^n Q_\mu(x_i))\Delta(x)\,dx}. \end{aligned}$$

PROOF. Set $h_n = h := Q_\mu - Q_\nu$. Then the assumption implies $\mu = \nu^h$. The first limit is an explicit expression of $-\nu^h(h)$ from (4.13), and the second limit is a rewriting of $-c(h,\nu)$ from (4.8). □

The free relative entropy $\Sigma(\mu,\nu)$ is symmetric in its two variables unlike the relative entropy, while the formula in Corollary 4.3 does not seem symmetric in μ and ν as it stands. On the other hand, the perturbation via relative entropy is symmetric in the sense that if μ is the perturbation of ν by h, then ν is the perturbation of μ by $-h$. This type of symmetry does not hold in the perturbation via free relative entropy as was verified in Example 2.7, even though the limiting procedure from the perturbation via relative entropy to that via free relative entropy was established in Theorem 4.2.

5. Specialization to free entropy.

In this section let us work on a finite interval $[-R,R]$ as in the previous section. Let σ be the arcsine law on $[-R,R]$ given in (2.14). Then $Q_\sigma(x) \equiv 2\log(R/2)$ and $\Sigma(\sigma) = \log(R/2)$ as remarked in Example 2.7, and

$$\Sigma(\mu,\sigma) = -\Sigma(\mu) + \log\frac{R}{2} \quad \text{for } \mu \in \mathcal{M}([-R,R]). \tag{5.1}$$

On the other hand, if m is the uniform distribution on $[-R,R]$ (i.e. $m = dx/2R$), then

$$S(\mu,m) = -S(\mu) + \log(2R) \quad \text{for } \mu \in \mathcal{M}([-R,R]).$$

Thus, the arcsine law can be considered as the free probabilistic analogue of the uniform distribution, and the minus free entropy is a special case (up to an additive constant) of free relative entropy while the minus Boltzmann-Gibbs entropy is a special case of relative entropy. The aim of this section is to find the exact forms when the previous results for free relative entropy are specialized via (5.1) to free entropy.

Define the Legendre transform of $-\Sigma(\mu)$ for $\mu \in \mathcal{M}([-R,R])$ as

$$\Pi(h) := \sup\{-\mu(h) + \Sigma(\mu) : \mu \in \mathcal{M}([-R,R])\}$$

for each $h \in C_{\boldsymbol{R}}([-R,R])$. The following formula is clear from (5.1):

$$\Pi(h) = c(h,\sigma) + \log\frac{R}{2} \quad \text{for } h \in C_{\boldsymbol{R}}([-R,R]). \tag{5.2}$$

For every $h \in C_{\boldsymbol{R}}([-R,R])$ let σ^h denote the unique maximizer (guaranteed by Lemma 2.1) of $-\mu(h) + \Sigma(\mu)$ for $\mu \in \mathcal{M}([-R,R])$ so that

$$-\sigma^h(h) + \Sigma(\sigma^h) = \Pi(h).$$

In fact, it is obvious from (5.1) that this σ^h is nothing but the perturbed probability measure of σ by h as defined in Section 2. Hence it is straightforward to translate (or specialize) all the results in Sections 2 and 3 to the case of $\Sigma(\mu)$ and $\Pi(h)$. For instance, we have the following:

(i) Π is a decreasing convex function on $C_{\boldsymbol{R}}([-R,R])$ satisfying

$$-\sigma(h) + \log\frac{R}{2} \le \Pi(h) \le \|h\| + \log\frac{R}{2}$$

and

$$|\Pi(h_1) - \Pi(h_2)| \le \|h_1 - h_2\|$$

for all $h, h_1, h_2 \in C_{\boldsymbol{R}}([-R,R])$.

(ii) For every $\mu \in \mathscr{M}([-R,R])$,

$$\Sigma(\mu) = \inf\{\mu(h) + \Pi(h) : h \in C_{\boldsymbol{R}}([-R,R])\}.$$

(iii) For $h \in C_{\boldsymbol{R}}([-R,R])$ and $\mu \in \mathscr{M}([-R,R])$, $\mu = \sigma^h$ if and only if

$$\Pi(h+k) \ge \Pi(h) - \mu(k) \quad \text{for all } k \in C_{\boldsymbol{R}}([-R,R]).$$

We finally adapt the results in Section 4 to the free entropy case. For $n \in \boldsymbol{N}$ let $\Lambda_{n,R}$ denote the restriction of the Lebesgue measure Λ_n (see Section 1) on $(M_n^{sa})_R$. Since $Q_\sigma(x)$ is constant, the probability measure $\lambda_n(\sigma) := \lambda_n(\sigma; R)$ on $(M_n^{sa})_R$ is nothing but the normalization of the restriction of Λ_n on $(M_n^{sa})_R$:

$$\lambda_n(\sigma) = \frac{1}{\Lambda_n((M_n^{sa})_R)} \Lambda_n|_{(M_n^{sa})_R}, \tag{5.3}$$

and it induces the joint eigenvalue distribution on $[-R,R]^n$

$$\tilde{\lambda}_n(\sigma; R) = \frac{1}{Z_n(\sigma; R)} \Delta(x) \prod_{i=1}^n \chi_{[-R,R]}(x_i)\, dx$$

with

$$Z_n(\sigma; R) = \int_{-R}^R \cdots \int_{-R}^R \Delta(x)\, dx.$$

Note ([**13**, p. 240]) that

$$\lim_{n\to\infty} \frac{1}{n^2} \log Z_n(\sigma; R) = \log\frac{R}{2}. \tag{5.4}$$

Let η be a probability measure on $(M_n^{sa})_R$. The *Boltzmann-Gibbs entropy* of η can be defined as

$$S(\eta) := -S(\eta, \Lambda_n), \tag{5.5}$$

where $S(\eta, \Lambda_n)$ is the relative entropy of η with respect to Λ_n. The measure η is also considered as an $n \times n$ selfadjoint random matrix and it induces a distribution on $\boldsymbol{R}^{n^2}$ via the isometry $M_n^{sa} \cong \boldsymbol{R}^{n^2}$ mentioned in Section 1. Then the entropy (5.5) is indeed equal to the usual Boltzmann-Gibbs entropy of the induced distribution on $\boldsymbol{R}^{n^2}$. From (5.3) one can rewrite (5.5) as

$$S(\eta) = -S(\eta, \lambda_n(\sigma)) + \log \Lambda_n((M_n^{sa})_R). \tag{5.6}$$

It is known ([**13**, p. 240]) that the measure on $\boldsymbol{R}^n$ induced by Λ_n is

$$C_n \Delta(x)\, dx \quad \text{with } C_n := \frac{(2\pi)^{n(n-1)/2}}{\prod_{j=1}^n j!},$$

and

$$\lim_{n\to\infty} \left(\frac{1}{n^2} \log C_n + \frac{1}{2} \log n \right) = \frac{1}{2} \log(2\pi) + \frac{3}{4}. \tag{5.7}$$

Under the above preparations we show the next theorem.

THEOREM 5.1. *If $h, h_n \in C_{\boldsymbol{R}}([-R, R])$, $n \in \boldsymbol{N}$, satisfy $\|h_n - h\| \to 0$, then the following hold*:

(i)

$$\Sigma(\sigma^h) + \frac{1}{2}\log(2\pi) + \frac{3}{4} = \lim_{n\to\infty} \left[\frac{1}{n^2} S(\lambda_n(h_n; R)) + \frac{1}{2} \log n \right].$$

(ii)

$$\sigma^h(h) = \lim_{n\to\infty} \frac{\int_{-R}^R \cdots \int_{-R}^R ((1/n)\sum_{i=1}^n h_n(x_i)) \exp(-n \sum_{i=1}^n h_n(x_i)) \Delta(x)\, dx}{\int_{-R}^R \cdots \int_{-R}^R \exp(-n\sum_{i=1}^n h_n(x_i)) \Delta(x)\, dx}.$$

(iii)

$$\Pi(h) = \lim_{n\to\infty} \frac{1}{n^2} \log \int_{-R}^R \cdots \int_{-R}^R \exp\left(-n \sum_{i=1}^n h_n(x_i) \right) \Delta(x)\, dx.$$

PROOF. (i) Since $\lambda_n(\sigma)^{\phi_n(h_n), S} = \lambda_n(h_n; R)$ by Lemma 4.1, it follows from (4.7), (5.1) and (5.6) that

$$-\Sigma(\sigma^h) + \log\frac{R}{2} = \lim_{n\to\infty}\frac{1}{n^2}[-S(\lambda_n(h_n;R)) + \log\Lambda_n((M_n^{sa})_R)].$$

Since by (5.4) and (5.7)

$$\begin{aligned}\lim_{n\to\infty}\left[\frac{1}{n^2}\log\Lambda_n((M_n^{sa})_R) + \frac{1}{2}\log n\right] &= \lim_{n\to\infty}\left[\frac{1}{n^2}\log Z_n(\sigma;R) + \frac{1}{n^2}\log C_n + \frac{1}{2}\log n\right] \\ &= \log\frac{R}{2} + \frac{1}{2}\log(2\pi) + \frac{3}{4},\end{aligned}$$

we have the desired formula.

(ii) By (4.6) and (4.1) we get

$$\begin{aligned}\sigma^h(h) &= \lim_{n\to\infty}\frac{1}{n^2}\lambda_n(h_n;R)(\phi_n(h_n)) \\ &= \lim_{n\to\infty}\int_{-R}^{R}\cdots\int_{-R}^{R}\left(\frac{1}{n}\sum_{i=1}^{n}h_n(x_i)\right)d\tilde{\lambda}_n(h_n;R)(x),\end{aligned}$$

which gives the desired formula.

(iii) By (4.8) and (5.2) we get

$$\begin{aligned}\Pi(h) &= \lim_{n\to\infty}\frac{1}{n^2}\log\int_{-R}^{R}\cdots\int_{-R}^{R}\exp\left(-n\sum_{i=1}^{n}h_n(x_i)\right)d\tilde{\lambda}_n(\sigma;R)(x) + \log\frac{R}{2} \\ &= \lim_{n\to\infty}\frac{1}{n^2}\log\int_{-R}^{R}\cdots\int_{-R}^{R}\exp\left(-n\sum_{i=1}^{n}h_n(x_i)\right)\Delta(x)\,dx\end{aligned}$$

thanks to (5.4). □

In particular, when $h_n = h$ for all $n \in \boldsymbol{N}$, the above (i) says that the free entropy $\Sigma(\sigma^h)$ is the renormalized limit of the Boltzmann-Gibbs entropy of the distribution $\lambda_n(h;R)$ on the matrix space. The asymptotic formulas in (ii) and (iii) together provide that of $\Sigma(\sigma^h) = \sigma^h(h) + \Pi(h)$.

The free entropy $\chi(\mu)$ in the microstates approach is defined by (1.3)–(1.4) so that

$$\chi(\mu) = \Sigma(\mu) + \frac{1}{2}\log(2\pi) + \frac{3}{4}$$

for $\mu \in \mathscr{M}([-R,R])$. Since the limit in the formula of Theorem 5.1 (iii) is re-written as

$$\lim_{n\to\infty}\left[\frac{1}{n^2}\log\int_{(M_n^{sa})_R}\exp(-n^2\operatorname{tr}_n(h(A)))\,d\Lambda_n(A)+\frac{1}{2}\log n\right]-\frac{1}{2}\log(2\pi)-\frac{3}{4},$$

the formula means that the Legendre transform π of $-\chi(\mu)$ for $\mu\in\mathcal{M}([-R,R])$ is given by

$$\begin{aligned}\pi(h)&=\Pi(h)+\frac{1}{2}\log(2\pi)+\frac{3}{4}\\&=\lim_{n\to\infty}\left[\frac{1}{n^2}\log\int_{(M_n^{sa})_R}\exp(-n^2\operatorname{tr}_n(h(A)))\,d\Lambda_n(A)+\frac{1}{2}\log n\right]\end{aligned}$$

for $h\in C_{\boldsymbol{R}}([-R,R])$. The form of this limit has some resemblance to the limit in (1.3).

Example 5.2. For $r>0$, when $h_r\in C_{\boldsymbol{R}}([-R,R])$ is given by $h_r(x):=2x^2/r^2$, we determined σ^{h_r} $(=\nu^{h_r})$ in Example 2.7 and

$$\Sigma(\sigma^{h_r})=\begin{cases}\log(r/2)-1/4 & \text{if } 0<r\le R,\\ \log(R/2)-R^4/4r^4 & \text{if } r\ge R,\end{cases}$$

$$\Pi(h_r)=\begin{cases}\log(r/2)-3/4 & \text{if } 0<r\le R,\\ \log(R/2)+R^4/4r^4-R^2/r^2 & \text{if } r\ge R,\end{cases}$$

When $r=2$, by Theorem 5.1 (iii) the latter estimate supplies the asymptotic limit of integrals

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n^2}\log\int_{-R}^{R}\cdots\int_{-R}^{R}\exp\left(-\frac{n}{2}\sum_{i=1}^{n}x_i^2\right)\Delta(x)\,dx\\&\quad=\begin{cases}-3/4 & \text{if } R\ge 2,\\ \log(R/2)+R^4/64-R^2/4 & \text{if } 0<R\le 2.\end{cases}\end{aligned}\tag{5.8}$$

Next, for $r>0$ set $k_r\in C_{\boldsymbol{R}}([-R,R])$ by $k_r(x):=2x/r$. The *free Poisson distribution* μ_1 (see [**21**, p. 34–35], [**13**, 3.3.5]) is given by

$$\mu_1=\frac{\sqrt{4x-x^2}}{2\pi x}\chi_{(0,4]}(x)\,dx.$$

It is known ([**13**, 5.3.7]) that

$$Q_{\mu_1}(x)\begin{cases}=x-2 & \text{if } 0\le x\le 4,\\ <x-2 & \text{if } x<0 \text{ or } x>4.\end{cases}$$

Let $\tilde{\mu}_r$ be the transform of μ_1 by the affine transformation $x \mapsto (r/2)(x-2)$ so that

$$\tilde{\mu}_r = \frac{2}{r} \cdot \frac{\sqrt{4((2x/r)+2) - ((2x/r)+2)^2}}{2\pi((2x/r)+2)} \chi_{(-r,r]}(x)\, dx = \frac{\sqrt{r^2 - x^2}}{\pi r(x+r)} \chi_{(-r,r]}(x)\, dx.$$

Then we have

$$Q_{\tilde{\mu}_r}(x) = 2\int \log\left|x - \frac{r}{2}(y-2)\right| d\mu_1(y) = Q_{\mu_1}\left(\frac{2x}{r} + 2\right) + 2\log\frac{r}{2}$$

so that

$$Q_{\tilde{\mu}_r}(x) \begin{cases} = (2x/r) + 2\log(r/2) & \text{if } |x| \le r, \\ < (2x/r) + 2\log(r/2) & \text{if } |x| > r. \end{cases}$$

Therefore, we notice that $\sigma^{k_r} = \tilde{\mu}_r$ when $0 < r \le R$. On the other hand, if $r \ge R$ and

$$\hat{\mu}_r := \left(1 - \frac{R}{r}\right)\sigma + \frac{R}{r}\tilde{\mu}_R,$$

then

$$Q_{\hat{\mu}_r}(x) = \frac{2x}{r} + 2\log\frac{R}{2} \quad \text{for } |x| \le R,$$

so we have $\sigma^{k_r} = \hat{\mu}_r$. Moreover, since $\int x\, d\tilde{\mu}_r(x) = -r/2$, the following can be easily computed:

$$\Sigma(\sigma^{k_r}) = \begin{cases} \log(r/2) - 1/2 & \text{if } 0 < r \le R, \\ \log(R/2) - R^2/2r^2 & \text{if } r \ge R, \end{cases}$$

$$\Pi(k_r) = \begin{cases} \log(r/2) + 1/2 & \text{if } 0 < r \le R, \\ \log(R/2) + R^2/2r^2 & \text{if } r \ge R. \end{cases}$$

When $r = 2$, Theorem 5.1 (iii) gives

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n^2} \log \int_{-R}^{R} \cdots \int_{-R}^{R} \exp\left(-n\sum_{i=1}^n x_i\right) \Delta(x)\, dx \\ = \begin{cases} 1/2 & \text{if } R \ge 2, \\ \log(R/2) + R^2/8 & \text{if } 0 < R \le 2. \end{cases} \end{aligned} \tag{5.9}$$

It is worthwhile to note that a phase transition occurs at $R = 2$ in the asymptotics of the integrals in (5.8) and (5.9).

In terms of statistical thermodynamics ([**17**])

$$\int_{-R}^{R} \cdots \int_{-R}^{R} \exp\left(-n \sum_{i=1}^{n} h(x_i)\right) \Delta(x)\, dx$$

is the *partition function* of n logarithmically interacting particle in an outer field h. So

$$\Pi(h) = \lim_{n\to\infty} \frac{1}{n^2} \log \int_{-R}^{R} \cdots \int_{-R}^{R} \exp\left(-n \sum_{i=1}^{n} h(x_i)\right) \Delta(x)\, dx$$

is nothing else but the *pressure* in a one-dimensional Coulomb gas model.

6. The case of measures on the unit circle.

Let μ be a probability measure on the unit circle $\boldsymbol{T}$. The free entropy of μ is given as

$$\Sigma(\mu) := \iint_{\boldsymbol{T}^2} \log|\zeta - \eta|\, d\mu(\zeta)\, d\mu(\eta).$$

For $n, r \in \boldsymbol{N}$ and $\varepsilon > 0$ define

$$\Gamma_u(\mu; n, r, \varepsilon) := \{U \in \mathscr{U}(n) : |\mathrm{tr}_n(U^k) - m_k(\mu)| \le \varepsilon, -r \le k \le r\},$$

where $\mathscr{U}(n)$ is the unitary group of order n and $m_k(\mu) := \int_{\boldsymbol{T}} \zeta^k\, d\mu(\zeta)$. Then $\Sigma(\mu)$ has the asymptotic expression ([**11**, Proposition 1.4]):

$$\Sigma(\mu) = \lim_{r\to\infty, \varepsilon\to+0} \lim_{n\to\infty} \frac{1}{n^2} \log \gamma_n(\Gamma_u(\mu; n, r, \varepsilon)), \tag{6.1}$$

where γ_n is the Haar probability measure on the group $\mathscr{U}(n)$.

Let ν be another probability measure on $\boldsymbol{T}$. The *free relative entropy* of μ with respect to ν is defined as (1.8) by

$$\Sigma(\mu, \nu) := \lim_{\varepsilon\to+0} \left[-\iint_{\boldsymbol{T}^2} \log(|\zeta - \eta| + \varepsilon)\, d(\mu - \nu)(\zeta)\, d(\mu - \nu)(\eta)\right].$$

To obtain the asymptotic expression of $\Sigma(\mu, \nu)$ via the matricial approximation, we need to introduce a probability distribution on $\mathscr{U}(n)$ corresponding to ν. Now assume that

$$Q_\nu(\zeta) := 2\int_{\boldsymbol{T}} \log|\zeta - \eta|\, d\nu(\eta)$$

is continuous on $\boldsymbol{T}$, and define a probability distribution $\gamma_n(\nu)$ on $\mathscr{U}(n)$ by

$$\gamma_n(\nu) := \frac{1}{Z_n(\nu)} \exp(-n^2 \operatorname{tr}_n(Q_\nu(U)))\, d\gamma_n(U),$$

where $Z_n(\nu)$ is for normalization. This distribution is invariant under unitary conjugation and its joint eigenvalue distribution on $\boldsymbol{T}^n$ has the density

$$\frac{1}{\tilde{Z}_n(\nu)} \exp\left(-n\sum_{i=1}^n Q_\nu(\zeta_i)\right) \prod_{i<j} |\zeta_i - \zeta_j|^2$$

with respect to $d\zeta_1 \cdots d\zeta_n$ where $d\zeta_j = (1/2\pi)\, d\theta_j$ $(\zeta_j = e^{i\theta_j})$. Then the following expression similar to (1.15) is proven by use of the large deviation theorem ([**12**]):

$$-\Sigma(\mu,\nu) = \lim_{r\to\infty,\,\varepsilon\to+0} \lim_{n\to\infty} \frac{1}{n^2} \log \gamma_n(\nu)(\Gamma_u(\mu; n, r, \varepsilon)). \tag{6.2}$$

The above (6.1) is a special case of (6.2) where ν is the uniform distribution ν_0 on $\boldsymbol{T}$ (having the maximal free entropy 0); notice $\Sigma(\mu) = -\Sigma(\mu,\nu_0)$ and $\gamma_n = \gamma_n(\nu_0)$.

One can see that the perturbation theory of measures in Section 2 is similarly valid also for measures on $\boldsymbol{T}$. Furthermore, one can perform the whole discussions in Sections 3 and 4 in the case of measures on $\boldsymbol{T}$ by replacing $\mathscr{M}([-R,R])$, $C_{\boldsymbol{R}}([-R,R])$ and $(M_n^{sa})_R$ by $\mathscr{M}(\boldsymbol{T})$, $C_{\boldsymbol{R}}(\boldsymbol{T})$ and $\mathscr{U}(n)$, respectively. The details are left to the reader, and we here remark just a few points. First, for $\mu_1, \mu_2 \in \mathscr{M}_\Sigma(\boldsymbol{T})$ $(:=\{\mu \in \mathscr{M}_\Sigma(\boldsymbol{T}) : \Sigma(\mu) > -\infty\})$ the variant of (3.2) is

$$\Sigma(\mu_1,\mu_2) = \sum_{n=1}^\infty \frac{1}{n} \left|\int_{\boldsymbol{T}} \zeta^n\, d(\mu_1 - \mu_2)(\zeta)\right|^2,$$

which shows that Lemma 3.1 remains valid for $\mathscr{M}_\Sigma(\boldsymbol{T})$. Thus, the convergence properties in Theorem 3.3 hold for perturbed measures on $\boldsymbol{T}$ too.

Next, the Legendre transform of $-\Sigma(\mu)$ is given as

$$\pi(h) := \sup\{-\mu(h) + \Sigma(\mu) : \mu \in \mathscr{M}(\boldsymbol{T})\}$$

for $h \in C_{\boldsymbol{R}}(\boldsymbol{T})$. Then we have

$$\begin{aligned}\pi(h) &= \lim_{n\to\infty} \frac{1}{n^2} \log \int_{\mathscr{U}(n)} \exp(-n^2 \operatorname{tr}_n(h(U)))\, d\gamma_n(U) \\ &= \lim_{n\to\infty} \frac{1}{n^2} \log \int_{\boldsymbol{T}} \cdots \int_{\boldsymbol{T}} \exp\left(-n\sum_{i=1}^n h(\zeta_i)\right) \prod_{i<j} |\zeta_i - \zeta_j|^2\, d\zeta_1 \cdots d\zeta_n\end{aligned}$$

and

$$\Sigma(\mu) = \inf\{\mu(h) + \pi(h) : h \in C_{\boldsymbol{R}}(\boldsymbol{T})\}$$

for $\mu \in \mathscr{M}(\boldsymbol{T})$. The "pressure" $\pi(h)$ was computed, for instance, in [**12**] for $h(\zeta) = \log|\zeta - \alpha|^2$ $(\alpha \in \boldsymbol{C}, |\alpha| < 1)$ and for $h(e^{i\theta}) = -(2/\lambda)\cos\theta$ $(\lambda > 0)$.

Finally, for every $h \in C_{\boldsymbol{R}}(\boldsymbol{T})$ let ν_0^h be the unique maximizer of $-\mu(h) + \Sigma(\mu)$ (i.e. the perturbed measure of ν_0 by h via free relative entropy), and let $\gamma_n(h)$ be a distribution on $\mathscr{U}(n)$ (or a unitary random matrix) defined by

$$\gamma_n(h) := \frac{1}{Z_n(h)} \exp(-n^2 \operatorname{tr}_n(h(U)))\, d\gamma_n(U).$$

Then we have

$$-\Sigma(\nu_0^h) = \Sigma(\nu_0^h, \nu_0) = \lim_{n\to\infty} \frac{1}{n^2} S(\gamma_n(h), \gamma_n).$$

References

[1] H. Araki, Relative entropy for states of von Neumann algebras II, Publ. Res. Inst. Math. Sci., **13** (1977), 173–192.

[2] G. Ben Arous and A. Guionnet, Large deviation for Wigner's law and Voiculescu's non-commutative entropy, Probab. Theory Related Fields, **108** (1997), 517–542.

[3] P. Biane and R. Speicher, Free diffusions, free entropy and free Fisher information, Ann. Inst. H. Poincaré Probab. Statist., **37** (2001), 581–606.

[4] I. Ciszár, Information-type measures of difference of probability distributions and indirect observations, Studia Sci. Math. Hungar., **2** (1967), 299–318.

[5] A. Dembo and O. Zeitouni, Large Deviation Techniques and Applications, Second edition, Springer, New York, 1998.

[6] J. D. Deuschel and D. W. Stroock, Large Deviations, Academic Press, Boston, 1989.

[7] M. J. Donald, Relative hamiltonian which are not bounded from above, J. Funct. Anal., **91** (1990), 143–173.

[8] U. Haagerup, unpublished notes.

[9] F. Hiai, M. Ohya and M. Tsukada, Sufficiency, KMS condition and relative entropy in von Neumann algebras, Pacific J. Math., **96** (1981), 99–109.

[10] F. Hiai and D. Petz, Maximizing free entropy, Acta Math. Hungar., **80** (1998), 335–356.

[11] F. Hiai and D. Petz, Properties of free entropy related to polar decomposition, Comm. Math. Phys., **202** (1999), 421–444.

[12] F. Hiai and D. Petz, A large deviation theorem for the empirical eigenvalue distribution of random unitary matrices, Ann. Inst. H. Poincaré Probab. Statist., **36** (2000), 71–85.

[13] F. Hiai and D. Petz, The Semicircle Law, Free Random Variables and Entropy, Mathematical Surveys and Monographs, Vol. 77, Amer. Math. Soc., Providence, 2000.

[14] M. Mizuo, Large deviations and microstate free relative entropy, Interdiscip. Inform. Sci., **6** (2000), 111–121.

[15] M. Ohya and D. Petz, Quantum Entropy and Its Use, Springer, Berlin-Heidelberg, 1993.

[16] D. Petz, A variational expression for the relative entropy, Comm. Math. Phys., **114** (1988), 345–349.

[17] D. Ruelle, Thermodynamic Formalism, Encyclopedia Math. Appl., Vol. 5, Addison-Wesley, London, 1978.

[18] E. B. Saff and V. Totik, Logarithmic Potentials with External Fields, Springer, Berlin-Heidelberg-New York, 1997.
[19] D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory, I, Comm. Math. Phys., **155** (1993), 71–92.
[20] D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory, II, Invent. Math., **118** (1994), 411–440.
[21] D. V. Voiculescu, K. J. Dykema and A. Nica, Free Random Variables, CRM Monogr. Ser., Vol. 1, Amer. Math. Soc., 1992.

Fumio HIAI
Graduate School of Information Sciences
Tohoku University
Aoba-ku, Sendai 980-8579
Japan
E-mail: hiai@math.is.tohoku.ac.jp

Masaru MIZUO
Graduate School of Information Sciences
Tohoku University
Aoba-ku, Sendai 980-8579
Japan
E-mail: mizuo@ims.is.tohoku.ac.jp

Dénes PETZ
Department for Mathematical Analysis
Budapest University
of Technology and Economics
H-1521 Budapest XI. Sztoczek u. 2
Hungary
E-mail: petz@math.bme.hu