FREE TRANSPORTATION COST INEQUALITIES VIA RANDOM MATRIX APPROXIMATION

FUMIO HIAI [1], DÉNES PETZ [2], AND YOSHIMICHI UEDA [3]

Abstract. By means of random matrix approximation procedure, we reprove Biane and Voiculescu's free analog of Talagrand's transportation cost inequality for measures on $\mathbf{R}$ in a more general setup. Furthermore, we prove the free transportation cost inequality for measures on $\mathbf{T}$ as well by extending the method to special unitary random matrices.

Introduction

In 1996, M. Talagrand [26] obtained an interesting inequality, called the transportation cost inequality (TCI), comparing the (quadratic) Wasserstein distance $W(\mu,\nu)$ between probability measures μ,ν (for the definition see (2.1) in §2 of this paper) with $\sqrt{S(\mu,\nu)}$, the square root of the relative entropy. Indeed, in [26] Talagrand proved the inequality $W(\mu,\nu) \le \sqrt{S(\mu,\nu)}$ when ν is the standard Gaussian measure on $\mathbf{R}^n$, and an exposition in the case of more general ν can be found in [19] for example (see also Theorem 2.2 in §2). Furthermore, in [23] F. Otto and C. Villani succeeded in discovering links between the TCI and the logarithmic Sobolev inequality (LSI) in the Riemannian manifold setting. This, combined with the celebrated LSI due to D. Bakry and M. Emery [1], implies the TCI in the same situation as in [1] (see Theorem 2.3 in §2). In [18, 19, 27] the interested reader will find more about the subject matter as well as related topics.

The relative free entropy $\widetilde{\Sigma}_Q(\mu)$ was introduced by Ph. Biane and R. Speicher [5] for $\mu \in \mathcal{M}(\mathbf{R})$, the probability measures on $\mathbf{R}$, relative to a real continuous function Q on $\mathbf{R}$ with a certain growth condition. Note that $\widetilde{\Sigma}_Q(\mu)$ is regarded as the relative version of the free entropy $\Sigma(\mu)$ introduced by D. Voiculescu [28] as the classical relative entropy is the relative version of the Boltzmann-Gibbs entropy. (The "free relative entropy" $\Sigma(\mu,\nu)$ for two measures was introduced in [11] from a slightly different viewpoint.) In this paper the relative free entropy $\widetilde{\Sigma}_Q(\mu)$ is also introduced for $\mu \in \mathcal{M}(\mathbf{T})$, the probability measures on the 1-dimensional torus $\mathbf{T}$, relative to a real continuous function Q on $\mathbf{T}$. An important fact is that the relative free entropy $\widetilde{\Sigma}_Q(\mu)$ is the rate function (or the so-called weighted logarithmic integral up to an additive constant) of a large deviation for the empirical eigenvalue distribution of a certain random matrix. Indeed, $\widetilde{\Sigma}_Q(\mu)$ for $\mu \in \mathcal{M}(\mathbf{R})$ is the good rate function of large deviation principle for the $n \times n$ selfadjoint random matrix determined by the function Q, while $\widetilde{\Sigma}_Q(\mu)$ for $\mu \in \mathcal{M}(\mathbf{T})$ is that for the $n \times n$ (special) unitary random matrix associated with Q. The definitions of these quantities as well as related matters are collected in the first §1 of this paper.

---

[1] Supported in part by Grant-in-Aid for Scientific Research (C)14540198 and by the program "R&D support scheme for funding selected IT proposals" of the Ministry of Public Management, Home Affairs, Posts and Telecommunications.

[2] Supported in part by MTA-JSPS project (Quantum Probability and Information Theory) and by OTKA T032662.

[3] Supported in part by Grant-in-Aid for Young Scientists (B)14740118.

In [6] Biane and Voiculescu obtained the free analog of Talagrand's TCI for compactly supported $\mu \in \mathcal{M}(\mathbf{R})$ as follows:

$$W(\mu, \gamma_{0,2}) \le \sqrt{-\Sigma(\mu) + \int_{\mathbf{R}} \frac{x^2}{2}\, d\mu(x) - \frac{3}{4}},$$

where $\gamma_{0,2}$ denotes the standard semicircular distribution (with radius 2). Their proof involves the free process and the complex Burgers' equation, and one of its keys is Voiculescu's free LSI in [30, Proposition 7.9]. The work [6] is a realization of free probability parallel of not only the result itself but also the proof in [23], and their proof itself justifies the above inequality to be the right free analog of Talagrand's TCI.

In §2 of this paper we reprove Biane and Voiculescu's free TCI in a slightly more general setting by making use of random matrix approximation and furthermore give a free TCI for measures on $\mathbf{T}$ in a similar way. Our initial motivation is to find another proof to Biane and Voiculescu's TCI by use of random matrix approximation on the lines of so-called Voiculescu's heuristics in [28] and to justify their TCI as the right free analog from the viewpoint of random matrix theory. In §§2.1 we prove the free TCI

$$W(\mu, \mu_Q) \le \sqrt{\frac{1}{\rho}\widetilde{\Sigma}_Q(\mu)} \quad \text{for compactly supported } \mu \in \mathcal{M}(\mathbf{R})$$

if Q is a real function on $\mathbf{R}$ such that $Q(x) - \frac{\rho}{2}x^2$ is convex with a constant $\rho > 0$ and μ_Q is the equilibrium measure associated with Q (or the unique minimizer of $\widetilde{\Sigma}_Q(\mu)$ for $\mu \in \mathcal{M}(\mathbf{R})$). When $Q(x) = x^2/2$ and $\rho = 1$, this becomes Biane and Voiculescu's TCI. To prove this, we first suppose that μ is supported in $[-R, R]$ and that $Q_\mu(x) := 2\int_{\mathbf{R}} \log|x-y|\, d\mu(y)$ is continuous on $\mathbf{R}$. We consider two $n \times n$ selfadjoint random matrices; one is associated with Q, and the other is associated with Q_μ and restricted on the $n \times n$ selfadjoint matrices with the operator norm $\le R$. Then, these random matrices are probability measures on the space of $n \times n$ selfadjoint matrices ($\cong \mathbf{R}^{n^2}$), and the classical TCI for these measures asymptotically approaches, as n goes to ∞, to the free TCI we want. The case of general compactly supported $\mu \in \mathcal{M}(\mathbf{R})$ can be treated by an approximation technique.

Furthermore, in §§2.2 we apply a similar method using special unitary random matrices to prove the free TCI

$$W(\mu, \mu_Q) \le \sqrt{\frac{2}{1+2\rho}\widetilde{\Sigma}_Q(\mu)} \quad \text{for } \mu \in \mathcal{M}(\mathbf{T})$$

if Q is a real function on $\mathbf{T}$ such that $Q\left(e^{\mathrm{i}t}\right) - \frac{\rho}{2}t^2$ is convex on $\mathbf{R}$ with $\rho > -1/2$. Here, $W(\mu, \mu_Q)$ is the Wasserstein distance with respect to the geodesic distance (or the angular distance) on $\mathbf{T}$. In the particular case where $Q \equiv 0$ and $\rho = 0$, we have

$$W\left(\mu, \frac{d\theta}{2\pi}\right) \le \sqrt{-2\Sigma(\mu)} \quad \text{for } \mu \in \mathcal{M}(\mathbf{T}).$$

The final §3 is a collection of remarks and examples of computations.

It is worthwhile to note that the random matrix approximation method can be also used to obtain the free probabilistic analog of the classical LSI, which compares the relative free Fisher information (see [5]) with the relative free entropy. Indeed, in [4] Biane obtained the free LSI (extending Voiculescu's one in [30]) for measures on $\mathbf{R}$, and in our forthcoming notes [15] we will treat its variant for measures on $\mathbf{T}$ based on the special unitary random matrix approximation together with some related aspects.

1. Preliminaries

The purpose of this preliminary section is to summarize, for the convenience of the reader, the basic notions and the results which will be needed later. We will use them with no explicit explanation in the main part of this paper.

1.1. **Notations.**

The set of all Borel probability measures on a Polish space $\mathcal{X}$ is denoted by $\mathcal{M}(\mathcal{X})$. The Dirac measure at a point $x \in \mathcal{X}$ is denoted by δ_x as usual. For $\mu, \nu \in \mathcal{M}(\mathcal{X})$, the *relative entropy* of μ with respect to ν is denoted by $S(\mu, \nu)$, which is defined by

$$S(\mu,\nu) := \int_{\mathcal{X}} \log \frac{d\mu}{d\nu}\, d\mu = \int_{\mathcal{X}} \frac{d\mu}{d\nu} \log \frac{d\mu}{d\nu}\, d\nu \tag{1.1}$$

when μ is absolutely continuous with respect to ν; otherwise $S(\mu,\nu) := +\infty$.

The usual trace on $M_n(\mathbf{C})$, the $n \times n$ complex matrices, is denoted by Tr_n. The Hilbert-Schmidt norm on $M_n(\mathbf{C})$ induced from Tr_n is denoted by $\|\cdot\|_{HS}$, i.e., $\|A\|_{HS} := \mathrm{Tr}_n(A^*A)^{1/2}$ for $A \in M_n(\mathbf{C})$. Let M_n^{sa} denote the set of all $n \times n$ self-adjoint matrices, $\mathrm{U}(n)$ the group of all $n \times n$ unitaries, and $\mathrm{SU}(n)$ the special unitary group of order n, i.e., the group of all $n \times n$ unitaries whose determinants are equal to one.

1.2. **Free entropy for measures.**

The notion of free entropy is the free probabilistic analog of the Boltzmann-Gibbs entropy in classical theory. For each $\mu \in \mathcal{M}(\mathbf{R})$, Voiculescu [28] introduced the *free entropy* of μ

$$\Sigma(\mu) := \iint_{\mathbf{R}^2} \log|x-y|\, d\mu(x)\, d\mu(y),$$

which is the minus of the so-called *logarithmic energy* of μ useful in potential theory (see [24]). It is the "main component" of the free entropy $\chi(\mu)$ of μ introduced in [29]:

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

For each $\mu \in \mathcal{M}(\mathbf{T})$, the *free entropy* $\Sigma(\mu)$ of μ is defined in the same manner as in the real line case; that is,

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

([31, §§10.7], [12]). For its justification to be a right quantity, see [31, Proposition 10.8] in relation to the free Fisher information as well as [12, Proposition 1.4], [13] from the microstate approach or large deviation principle.

1.3. **Large deviations for self-adjoint random matrices.**

Let Q be a real-valued continuous function on $\mathbf{R}$ such that

$$\lim_{|x|\to+\infty} |x|\exp(-\varepsilon Q(x)) = 0 \quad \text{for every } \varepsilon > 0. \tag{1.2}$$

The *weighted energy integral* associated with Q is defined by

$$E_Q(\mu) := -\Sigma(\mu) + \int_{\mathbf{R}} Q(x)\, d\mu(x) \quad \text{for } \mu \in \mathcal{M}(\mathbf{R}).$$

According to a fundamental result in the theory of weighted potentials (see [24, I.1.3]), there exists a unique $\mu_Q \in \mathcal{M}(\mathbf{R})$ such that

$$E_Q(\mu_Q) = \inf\left\{E_Q(\mu) : \mu \in \mathcal{M}(\mathbf{R})\right\},$$

and $E_Q(\mu_Q)$ is finite (hence so is $\Sigma(\mu_Q)$). Moreover, μ_Q is known to be compactly supported. The minimizer μ_Q is sometimes called the *equilibrium measure* associated with Q. Set $B(Q) := -E_Q(\mu_Q)$ so that the function

$$-\Sigma(\mu) + \int_{\mathbf{R}} Q(x)\,d\mu(x) + B(Q) \quad \text{for } \mu \in \mathcal{M}(\mathbf{R}) \tag{1.3}$$

is non-negative and is zero only when $\mu = \mu_Q$. It is well known that if $Q(x) = 2x^2/r^2$ with $r > 0$, then the equilibrium measure (or the unique minimizer) μ_Q is the $(0, r^2/4)$-*semicircular distribution* $\gamma_{0,r}$ (with variance $r^2/4$):

$$d\gamma_{0,r}(x) := \frac{2}{\pi r^2}\sqrt{r^2 - x^2}\,\chi_{[-r,r]}(x)\,dx. \tag{1.4}$$

For each $n \in \mathbf{N}$ define $\lambda_n(Q) \in \mathcal{M}(M_n^{sa})$, the $n \times n$ *self-adjoint random matrix* associated with Q, by

$$d\lambda_n(Q)(A) := \frac{1}{Z_n(Q)}\exp\bigl(-n\mathrm{Tr}_n(Q(A))\bigr)\,dA,$$

where dA means the "Lebesgue measure" on $M_n^{sa} \cong \mathbf{R}^{n^2}$, i.e.,

$$dA := \prod_{i=1}^n dA_{ii} \prod_{i<j} d\,(\mathrm{Re}\,A_{ij})\,d\,(\mathrm{Im}\,A_{ij}) \quad \text{with } A = [A_{ij}],$$

$Q(A)$ is the usual functional calculus and $Z_n(Q)$ is a normalization constant. It is known (see [20, 14] for example) that the *joint eigenvalue distribution* on $\mathbf{R}^n$ of $\lambda_n(Q)$ is given as

$$d\tilde{\lambda}_n(Q)(x_1, \dots, x_n) := \frac{1}{\widetilde{Z}_n(Q)}\exp\left(-n\sum_{i=1}^n Q(x_i)\right)\prod_{i<j}(x_i - x_j)^2 \prod_{i=1}^n dx_i$$

with a new normalization constant $\widetilde{Z}_n(Q)$. Moreover, the *mean eigenvalue distribution* on $\mathbf{R}$ of $\lambda_n(Q)$ is defined by

$$\hat{\lambda}_n(Q) := \int\cdots\int_{\mathbf{R}^n} \frac{1}{n}\,(\delta_{x_1} + \cdots + \delta_{x_n})\,d\tilde{\lambda}_n(Q)(x_1, \dots, x_n).$$

In [2] Ben Arous and Guionnet showed the large deviation principle for the empirical eigenvalue distribution of the standard self-adjoint Gaussian random matrix (i.e., $\lambda_n(Q)$ with $Q(x) = x^2/2$). The following is its slight generalization given in [14, 5.4.3]: When $(x_1, \dots, x_n)$ is distributed according to $\tilde{\lambda}_n(Q)$, the *empirical eigenvalue distribution*

$$\frac{1}{n}\,(\delta_{x_1} + \cdots + \delta_{x_n}) \tag{1.5}$$

satisfies the large deviation principle in the scale $1/n^2$ and the good rate function is given by (1.3). Furthermore, one has $B(Q) = \lim_{n\to\infty} \frac{1}{n^2}\log \widetilde{Z}_n(Q)$, i.e.,

$$B(Q) = \lim_{n\to\infty}\frac{1}{n^2}\log\int\cdots\int_{\mathbf{R}^n}\exp\left(-n\sum_{i=1}^n Q\left(x_i\right)\right)\prod_{i<j}(x_i - x_j)^2\prod_{i=1}^n dx_i.$$

See [8, 9] for general theory of large deviations. Since μ_Q is the unique minimizer of (1.3), the random measure (1.5) converges in the weak topology to μ_Q almost surely, and hence $\hat{\lambda}_n(Q) \to \mu_Q$ weakly (see [14, p. 211] and also [7]). From the viewpoint of the large deviation theory of level-2 (see [8, 9]), the function (1.3) can be regarded as a kind of free analog of the relative entropy with respect to its unique minimizer μ_Q. Thus, following Biane and Speicher

[5, §6] and Biane [4, §3], we call the function (1.3) the *relative free entropy* (or *modified free entropy*) of μ relative to Q, which is denoted by $\widetilde{\Sigma}_Q(\mu)$; that is,

$$\widetilde{\Sigma}_Q(\mu) := -\Sigma(\mu) + \int_{\mathbf{R}} Q(x)\,d\mu(x) + B(Q) \quad \text{for } \mu \in \mathcal{M}(\mathbf{R}). \tag{1.6}$$

We do *not* call this the "free relative entropy" introduced in [11], a slightly different relative entropy-like quantity $\Sigma(\mu,\nu)$ for two probability measures in the framework of free probability. Indeed, the free relative entropy $\Sigma(\mu,\nu)$ for $\mu,\nu \in \mathcal{M}(\mathbf{R})$ is defined as

$$\Sigma(\mu,\nu) := \iint_{\mathbf{R}^2} \log|x-y|\,d(\mu-\nu)(x)\,d(\mu-\nu)(y).$$

But it is known (see [11, (2.7)]) that $\Sigma(\mu,\mu_Q) = \widetilde{\Sigma}_Q(\mu)$ if the support of μ is included in that of μ_Q.

1.4. **Large deviations for restricted self-adjoint random matrices.** In the course of finding a right free analog of relative entropy, another random matrix model associated with Q and $R > 0$ was introduced in [11]. Here, Q is an arbitrary real-valued continuous function whose domain includes $[-R,R]$. The self-adjoint random matrix $\lambda_n(Q;R) \in \mathcal{M}\left(M_n^{sa}\right)$ restricted on a compact subset $\{A \in M_n^{sa} : \|A\|_\infty \le R\}$ is defined by

$$d\lambda_n(Q;R)(A) := \frac{1}{Z_n(Q;R)} \exp\bigl(-n\mathrm{Tr}_n(Q(A))\bigr)\chi_{\{\|A\|_\infty \le R\}}(A)\,dA$$

with a normalization constant $Z_n(Q;R)$. In the above, $\|\cdot\|_\infty$ means the operator norm. The joint eigenvalue distribution supported in $[-R,R]^n$ of $\lambda_n(Q;R)$ is given as

$$\begin{aligned} &d\tilde{\lambda}_n(Q;R)(x_1,\dots,x_n) \\ &\quad := \frac{1}{\widetilde{Z}_n(Q;R)} \exp\left(-n\sum_{i=1}^n Q(x_i)\right) \prod_{i<j}(x_i-x_j)^2 \prod_{i=1}^n \chi_{[-R,R]}(x_i)\,dx_i \end{aligned}$$

with a new normalization constant $\tilde{Z}_n(Q;R)$. Its mean eigenvalue distribution $\hat{\lambda}_n(Q;R)$ supported in $[-R,R]$ is defined as in §§1.3. As in the case of $\lambda_n(Q)$, the following large deviation theorem holds: The finite limit $B(Q;R) := \lim_{n\to\infty} \dfrac{1}{n^2} \log \tilde{Z}_n(Q;R)$ exists, and when $(x_1,\dots,x_n)$ is distributed according to $\tilde{\lambda}_n(Q;R)$, the empirical eigenvalue distribution (1.5) satisfies the large deviation principle in the scale $1/n^2$ with the rate function

$$-\Sigma(\mu) + \int_{[-R,R]} Q(x)\,d\mu(x) + B(Q;R) \quad \text{for } \mu \in \mathcal{M}([-R,R]). \tag{1.7}$$

The proof of this large deviation principle is similar to [14, 5.4.3 and 5.5.1] as noticed in [11]. In this setting, there also exists a unique minimizer $\mu_{Q,R} \in \mathcal{M}([-R,R])$ of the rate function (1.7), whose value at $\mu_{Q,R}$ is zero. If $R > 0$ is chosen so that μ_Q in §§1.3 is supported in $[-R,R]$, then $\mu_Q = \mu_{Q,R}$ is seen by comparing the two rate functions, and hence $B(Q) = B(Q;R)$; this assertion is essentially same as in [29, Proposition 2.4] in the single variable case.

1.5. **Large deviations for special unitary random matrices.** Let Q be a real-valued continuous function on $\mathbf{T}$. Similarly to the real line case in §§1.3, the weighted energy integral

$$-\Sigma(\mu) + \int_{\mathbf{T}} Q(\zeta)\,d\mu(\zeta) \quad \text{for } \mu \in \mathcal{M}(\mathbf{T})$$

admits a unique minimizer $\mu_Q \in \mathcal{M}(\mathbf{T})$ (or the equilibrium measure associated with Q). Set $B(Q) := \Sigma(\mu_Q) - \int_{\mathbf{T}} Q(\zeta)\,d\mu_Q(\zeta)$. It is known ([13]) that the function

$$-\Sigma(\mu) + \int_{\mathbf{T}} Q(\zeta)\,d\mu(\zeta) + B(Q) \quad \text{for } \mu \in \mathcal{M}(\mathbf{T})$$

is the rate function of the large deviation for the empirical eigenvalue distribution of an $n \times n$ *unitary random matrix*

$$d\lambda_n^{\mathrm{U}}(Q)(U) := \frac{1}{Z_n^{\mathrm{U}}(Q)} \exp\Big(-n\mathrm{Tr}_n(Q(U))\Big)\,dU,$$

where dU is the Haar probability measure on $\mathrm{U}(n)$, $Q(U)$ is defined via functional calculus and $Z_n^{\mathrm{U}}(Q)$ is a normalization constant. Furthermore,

$$B(Q) = \lim_{n\to\infty} \frac{1}{n^2} \log \int\cdots\int_{\mathbf{T}^n} \exp\left(-n\sum_{i=1}^n Q(\zeta_i)\right) \prod_{1\le i<j\le n} |\zeta_i - \zeta_j|^2 \prod_{i=1}^n d\zeta_i$$

where $d\zeta_i = d\theta_i/2\pi$ for $\zeta_i = e^{\mathrm{i}\theta_i}$. However, the above unitary random matrix $\lambda_n^{\mathrm{U}}(Q)$ is not suitable for our present purpose as will be explained in §1.6. Thus, we need to modify the above large deviation to the setup of $\mathrm{SU}(n)$.

It seems a folklore for specialists that the joint eigenvalue distribution on $\mathbf{T}^{n-1}$ of the Haar probability measure on $\mathrm{SU}(n)$ is

$$\frac{1}{n!} \prod_{1\le i<j\le n} |\zeta_i - \zeta_j|^2 \prod_{i=1}^{n-1} d\zeta_i \quad \text{with } \zeta_n = (\zeta_1\cdots\zeta_{n-1})^{-1}, \tag{1.8}$$

or

$$\frac{1}{n!(2\pi)^{n-1}} \prod_{1\le i<j\le n} \left|e^{\mathrm{i}\theta_i} - e^{\mathrm{i}\theta_j}\right|^2 \prod_{i=1}^{n-1} d\theta_i$$
$$\text{with } \theta_n = -(\theta_1 + \cdots + \theta_{n-1}) \pmod{2\pi}.$$

In fact, this distribution is easily derived from the Weyl integration formula familiar in representation theory (see [16, p. 104] for example). Let Q be a real-valued continuous function on $\mathbf{T}$. For each $n \in \mathbf{N}$ define $\lambda_n(Q) \in \mathcal{M}(\mathrm{SU}(n))$, the $n \times n$ *special unitary random matrix* associated with Q, by

$$d\lambda_n^{\mathrm{SU}}(Q)(U) := \frac{1}{Z_n^{\mathrm{SU}}(Q)} \exp\big(-n\mathrm{Tr}_n(Q(U))\big)\,dU,$$

where dU is the Haar probability measure on $\mathrm{SU}(n)$ and $Z_n^{\mathrm{SU}}(Q)$ is a normalization constant. By (1.8) the joint eigenvalue distribution on $\mathbf{T}^{n-1}$ of $\lambda_n^{\mathrm{SU}}(Q)$ is given as

$$d\tilde{\lambda}_n^{\mathrm{SU}}(Q)(\zeta_1,\dots,\zeta_{n-1}) = \frac{1}{\widetilde{Z}_n^{\mathrm{SU}}(Q)} \exp\left(-n\sum_{i=1}^n Q(\zeta_i)\right) \prod_{1\le i<j\le n} |\zeta_i - \zeta_j|^2 \prod_{i=1}^n d\zeta_i$$
$$\text{with} \quad \zeta_n = (\zeta_1\cdots\zeta_{n-1})^{-1}.$$

The next theorem is the large deviation principle for the empirical eigenvalue distribution of $\lambda_n^{\mathrm{SU}}(Q)$, whose proof, based on the explicit form of the density of $\tilde{\lambda}_n^{\mathrm{SU}}(Q)$, will be explained in our forthcoming notes on the related topics of this paper.

Theorem 1.1. *The finite limit* $B(Q) := \lim_{n\to\infty} \frac{1}{n^2} \log \widetilde{Z}_n^{\mathrm{SU}}(Q)$ *exists. When* $(\zeta_1,\dots,\zeta_{n-1})$ *is distributed on* $\mathbf{T}^{n-1}$ *according to* $\tilde{\lambda}_n^{\mathrm{SU}}(Q)$*, the empirical distribution* $\frac{1}{n}(\delta_{\zeta_1} + \cdots + \delta_{\zeta_{n-1}} + \delta_{\zeta_n})$

with $\zeta_n = (\zeta_1 \cdots \zeta_{n-1})^{-1}$ *satisfies the large deviation principle in the scale* $1/n^2$ *with the rate function*

$$\widetilde{\Sigma}_Q(\mu) := -\Sigma(\mu) + \int_{\mathbf{T}} Q(\zeta)\,d\mu(\zeta) + B(Q) \quad \textit{for } \mu \in \mathcal{M}(\mathbf{T}). \tag{1.9}$$

Furthermore, there exists a unique minimizer $\mu_Q \in \mathcal{M}(\mathbf{T})$ *of the rate function so that* $\widetilde{\Sigma}_Q(\mu_Q) = 0$.

As before, we call the rate function (1.9) the *relative free entropy* of μ with respect to Q, which is denoted by $\widetilde{\Sigma}_Q(\mu)$ as in (1.6).

1.6. **Ricci curvature tensor of** $\mathrm{SU}(n)$**.** Let M be a smooth complete Riemannian manifold of dimension m, and let $\mathrm{Ric}(M)$ denote the *Ricci curvature tensor* of M. For a real-valued C^2 function Ψ on M, the *Hessian* of Ψ is denoted by $\mathrm{Hess}(\Psi)$. Our arguments in §§2.2 will need to verify the so-called *Bakry and Emery criterion* with a positive constant ρ:

$$\mathrm{Ric}(M) + \mathrm{Hess}(\Psi) \geq \rho I_m; \tag{1.10}$$

see [1] and also Theorem 2.3 below.

The Ricci curvature tensor of $\mathrm{U}(n)$ is known to be degenerate, while that of $\mathrm{SU}(n)$ to be of positive constant (see [21], a nice reference for the topic) and a straightforward computation shows that the Ricci curvature tensor of $\mathrm{SU}(n)$ with respect to the Riemannian structure associated with Tr_n is

$$\mathrm{Ric}\,(\mathrm{SU}(n)) = \frac{n}{2} I_{n^2-1}. \tag{1.11}$$

This is the reason why we have presented Theorem 1.1 with use of $\mathrm{SU}(n)$ instead of $\mathrm{U}(n)$.

1.7. **Hessian of trace functions.** As explained just above, we will have to verify the Bakry and Emery criterion (1.10) for a trace function on $\mathrm{SU}(Q)$. Thus, the higher (especially, the second) order differentiability of a certain kind of trace functions will be needed as a prerequisite to discuss the Hessian. The topic seems rather familiar to specialists, however we can find no appropriate literature.

Let $f(t)$ be a real-valued function on an interval (a, b), and let $\lambda_1, \lambda_2, \ldots$ be distinct points in (a, b). The *divided differences* $f^{[r]}$ for $r = 0, 1, 2, \ldots$ are recursively introduced as follows: $f^{[0]}(\lambda_1) := f(\lambda_1)$ and

$$f^{[r]}(\lambda_1, \lambda_2, \ldots, \lambda_{r+1}) := \frac{f^{[r-1]}(\lambda_1, \lambda_2, \ldots, \lambda_r) - f^{[r-1]}(\lambda_2, \ldots, \lambda_r, \lambda_{r+1})}{\lambda_1 - \lambda_{r+1}}.$$

When λ_i's are not necessarily distinct, $f^{[r]}(\lambda_1, \lambda_2, \ldots, \lambda_{r+1})$ can be defined by continuity as long as $f \in C^r(a, b)$; for instance, $f^{[1]}(\lambda, \lambda) = f'(\lambda)$ and $f^{[2]}(\lambda, \lambda, \lambda) = f''(\lambda)/2$. See [10, §II.2] for basic properties of divided differences. Let $A \in M_n^{sa}$, all of whose eigenvalues are in (a, b), and $A = \sum_{i=1}^{l} \lambda_i P_i$ be the spectral decomposition with distinct eigenvalues $\lambda_1, \ldots, \lambda_l$ in (a, b). For each $H_1, H_2, \ldots, H_r \in M_n^{sa}$ we define

$$\begin{aligned} &f^{[r]}(A) \circ (H_1, H_2, \ldots, H_r) \\ &:= \sum_{\sigma \in S_r} \sum_{i_1, \ldots, i_{r+1}=1}^{l} f^{[r]}(\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_{r+1}}) P_{i_1} H_{\sigma(1)} P_{i_2} H_{\sigma(2)} \cdots P_{i_r} H_{\sigma(r)} P_{i_{r+1}}, \end{aligned}$$

where S_r is the set of all permutations on $\{1, \ldots, r\}$. In particular, note ([3, V.3.3]) that if $f \in C^1(a, b)$ and $A = U\mathrm{diag}(\lambda_1, \ldots, \lambda_n)U^*$ is a diagonalization, then

$$\left.\frac{d}{dt}\right|_{t=0} f(A + tH_1) = f^{[1]}(A) \circ H_1 = U\Big(\Big[f^{[1]}(\lambda_i, \lambda_j)\Big]_{ij} \circ U^* H_1 U\Big)U^*,$$

where $\circ$ stands for the Schur product. The next lemma can be shown in an essentially same way as in the proof of [3, V.3.3].

Lemma 1.2. *Let* $A, H_1, \dots, H_m \in M_n^{sa}$ *and set* $G(x) := A + \sum_{k=1}^m x_k H_k$ *for* $x = (x_1, \dots, x_m) \in \mathbf{R}^m$. *Let* f *be a real-valued* C^r *function on* (a, b) *for some* $r \in \mathbf{N}$. *If the eigenvalues of* $G(x)$ *are in* (a, b) *for all* x *in an open domain* D *of* $\mathbf{R}^m$, *then the function* $\mathrm{Tr}_n(f(G(x)))$ *is* C^r *on* D *and*

$$\begin{aligned}\frac{\partial^r}{\partial x_{k_1}\partial x_{k_2}\cdots\partial x_{k_r}}\mathrm{Tr}_n(f(G(x))) &= \mathrm{Tr}_n\Big(f^{[r]}(G(x))\circ(H_{k_1}, H_{k_2}, \dots, H_{k_r})\Big) \\ &= \mathrm{Tr}_n\Big(\Big((f')^{[r-1]}(G(x))\circ(H_{k_1}, \dots, H_{k_{r-1}})\Big)H_{k_r}\Big)\end{aligned}$$

for all $1 \le k_1, k_2, \dots, k_r \le m$ *and* $x \in D$. *In particular,*

$$\frac{\partial}{\partial x_k}\mathrm{Tr}_n(f(G(x))) = \mathrm{Tr}_n(f'(G(x))H_k)$$

for all $1 \le k \le m$ *and* $x \in D$.

The next lemma is what we will actually need in §§2.2.

Lemma 1.3. *Let* Q *be a harmonic function on a neighborhood of the unit disk* $\{\zeta \in \mathbf{C} : |\zeta| \le 1\}$. *For each* $n \in \mathbf{N}$ *and each* $U \in \mathrm{SU}(n)$ *define* $Q(U)$ *via the functional calculus and set* $\Psi(U) := \mathrm{Tr}_n(Q(U))$. *Then one has*

(i) *The function* $\Psi(U)$ *on* $\mathrm{SU}(n)$ *is* C^∞.

(ii) *If* $Q\Big(e^{\mathrm{i}t}\Big) - \frac{\rho}{2}t^2$ *is convex on* $\mathbf{R}$ *for some constant* $\rho \in \mathbf{R}$, *then* $\mathrm{Hess}(\Psi) \ge \rho I_{n^2-1}$.

Proof. Set $f(t) := Q\Big(e^{\mathrm{i}t}\Big)$ for $t \in \mathbf{R}$, and let $Y_k := \mathrm{i}X_k$ with $X_k = X_k^*$, $1 \le k \le n^2-1$, be a basis of the Lie algebra $\mathfrak{su}(n) = \{T \in M_n(\mathbf{C}) : T + T^* = 0, \mathrm{Tr}_n(T) = 0\}$ $(\cong \mathbf{R}^{n^2-1})$. For any $U_0 = e^{\mathrm{i}A_0} \in \mathrm{SU}(n)$ with $\mathrm{i}A_0 \in \mathfrak{su}(n)$ and for $x = (x_1, \dots, x_{n^2-1}) \in \mathbf{R}^{n^2-1}$, we write

$$\Psi\left(\exp\left(\mathrm{i}A_0 + \sum_{k=1}^{n^2-1} x_k Y_k\right)\right) = \mathrm{Tr}_n\left(f\left(A_0 + \sum_{k=1}^{n^2-1} x_k X_k\right)\right).$$

The C^∞-property of f on $\mathbf{R}$ immediately follows from the assumption of Q. In fact, for each $t_0 \in \mathbf{R}$, the function $f(t_0 + t)$ has a power series expansion for t near 0. Thus, Lemma 1.2 implies (i).

Set $F(t) := Q\Big(e^{\mathrm{i}t}\Big) - \frac{\rho}{2}t^2$ for $t \in \mathbf{R}$. For any $U_0 = e^{\mathrm{i}A_0} \in \mathrm{SU}(n)$ with $\mathrm{i}A_0 \in \mathfrak{su}(n)$ and for $(x_1, \dots, x_{n^2-1}) \in \mathbf{R}^{n^2-1}$, we have

$$\begin{aligned}&\Psi\left(\exp\left(\mathrm{i}A_0 + \sum_{k=1}^{n^2-1} x_k Y_k\right)\right) \\ &\quad = \mathrm{Tr}_n\left(F\left(A_0 + \sum_{k=1}^{n^2-1} x_k X_k\right)\right) + \frac{\rho}{2}\mathrm{Tr}_n\left(\left(A_0 + \sum_{k=1}^{n^2-1} x_k X_k\right)^2\right) \\ &\quad = \mathrm{Tr}_n\left(F\left(A_0 + \sum_{k=1}^{n^2-1} x_k X_k\right)\right) + \frac{\rho}{2}\mathrm{Tr}_n(A_0^2) + \rho\sum_{k=1}^{n^2-1}\mathrm{Tr}_n(A_0X_k)x_k + \frac{\rho}{2}\sum_{k=1}^{n^2-1} x_k^2.\end{aligned}$$

Since $F(t)$ is convex on $\mathbf{R}$, it is known ([22, 3.1]) that $\mathrm{Tr}_n\big(F(A_0 + \sum_{k=1}^{n^2-1} x_k X_k)\big)$ is convex in $(x_1, \dots, x_{n^2-1})$ so that (ii) follows. □

2. Main Results

The aim of this paper is to obtain the free analog of transportation cost inequalities for measures on $\mathbf{R}$ and on $\mathbf{T}$. We deal with probability measures on $\mathbf{R}$ in the first half of this section and those on $\mathbf{T}$ in the latter. The (classical) transportation cost inequalities compare the Wasserstein distance with the relative entropy (see (1.1)) for two given probability measures. Let us first recall the definition of the Wasserstein distance. Let $\mathcal{X}$ be a Polish space with a metric d. The (quadratic) *Wasserstein distance* between $\mu, \nu \in \mathcal{M}(\mathcal{X})$ is defined by

$$W(\mu,\nu) := \inf_{\pi \in \Pi(\mu,\nu)} \sqrt{\iint_{\mathcal{X}\times\mathcal{X}} \frac{1}{2} d(x,y)^2 \, d\pi(x,y)}, \tag{2.1}$$

where $\Pi(\mu,\nu)$ denotes the set of all probability measures on $\mathcal{X}\times\mathcal{X}$ with marginals μ and ν, i.e., $\pi(\,\cdot\times\mathcal{X}) = \mu$ and $\pi(\mathcal{X}\times\cdot\,) = \nu$. The reader should note that the Wasserstein distance is sometimes defined with the integral of $d(x,y)^2$ instead of $\frac{1}{2}d(x,y)^2$. The next lemma is well known and easy to show.

Lemma 2.1. *$W(\mu,\nu)$ is weakly lower semicontinuous in $\mu,\nu \in \mathcal{M}(\mathcal{X})$; namely, if $\mu_n, \nu_n \in \mathcal{M}(\mathcal{X})$, $\mu_n \to \mu$ and $\nu_n \to \nu$ in the weak topology, then*

$$W(\mu,\nu) \le \liminf_{n\to\infty} W(\mu_n,\nu_n).$$

In the typical case where $\mathcal{X} = \mathbf{R}^n$ and $d(x,y) = \|x-y\|$, the usual Euclidean metric, let g_n be the standard Gaussian measure, i.e., $dg_n(x) := (2\pi)^{-n/2} e^{-\|x\|^2/2}\,dx$ (dx means the Lebesgue measure on $\mathbf{R}^n$). The celebrated *transportation cost inequality* (*TCI* in short) of Talagrand [26] is

$$W(\mu, g_n) \le \sqrt{S(\mu, g_n)}, \qquad \mu \in \mathcal{M}(\mathbf{R}^n).$$

This inequality is a bit extended as follows (see [19]):

Theorem 2.2. *Let $\Psi : \mathbf{R}^n \to \mathbf{R}$ and assume that $\Psi(x) - \frac{\rho}{2}\|x\|^2$ is convex on $\mathbf{R}^n$ with a constant $\rho > 0$. If $d\nu(x) := \frac{1}{Z} e^{-\Psi(x)}\,dx \in \mathcal{M}(\mathbf{R}^n)$ with a normalization constant Z, then*

$$W(\mu,\nu) \le \sqrt{\frac{1}{\rho} S(\mu,\nu)}, \qquad \mu \in \mathcal{M}(\mathbf{R}^n).$$

In [23] Otto and Villani established the interrelation between TCI and logarithmic Sobolev inequalities (see [18, 19] for example) by a technique using partial differential equations. Their result, combined with Bakry and Emery's LSI (see [1] for details), implies the following TCI in a setup on Riemannian manifolds, which will play a crucial role in deriving our free analog of TCI for measures on $\mathbf{T}$. In the theorem, let M be an m-dimensional smooth complete Riemannian manifold equipped with the geodesic distance $d(x,y)$ and the volume measure dx.

Theorem 2.3. (Bakry and Emery [1] and Otto and Villani [23]) *Let Ψ be a real-valued C^2 function on M and set $d\nu(x) := \frac{1}{Z} e^{-\Psi(x)}\,dx \in \mathcal{M}(M)$ with a normalization constant Z. If the Bakry and Emery criterion $\mathrm{Ric}(M) + \mathrm{Hess}(\Psi) \ge \rho I_m$ holds with a constant $\rho > 0$, then*

$$W(\mu,\nu) \le \sqrt{\frac{1}{\rho} S(\mu,\nu)}, \qquad \mu \in \mathcal{M}(M).$$

On the other hand, the following free analog of Talagrand's TCI is shown by Biane and Voiculescu [6]. Recall that $\gamma_{0,2}$ is the standard semicircular measure (see (1.4)).

Theorem 2.4. (Biane and Voiculescu [6]) *For every compactly supported $\mu \in \mathcal{M}(\mathbf{R})$,*

$$W(\mu, \gamma_{0,2}) \le \sqrt{-\Sigma(\mu) + \int_{\mathbf{R}} \frac{x^2}{2}\, d\mu(x) - \frac{3}{4}}. \tag{2.2}$$

In the rest of this section we will present a new proof of the above free TCI for measures on $\mathbf{R}$ in a more general situation by using a random matrix technique, and then prove its counterpart in the case of measures on $\mathbf{T}$. Our essential idea is that the classical TCI on the matrix space M_n^{sa} (resp. on SU(n)) asymptotically approaches to the right free analog when the matrix size goes to ∞.

2.1. **The real line case.** Let us prove the following free TCI for measures on $\mathbf{R}$:

Theorem 2.5. *Let Q be a real-valued function on $\mathbf{R}$. If $Q(x) - \frac{\rho}{2}x^2$ is convex on $\mathbf{R}$ with a constant $\rho > 0$, then*

$$W(\mu, \mu_Q) \le \sqrt{\frac{1}{\rho}\widetilde{\Sigma}_Q(\mu)} \tag{2.3}$$

for every compactly supported $\mu \in \mathcal{M}(\mathbf{R})$.

In particular, when $Q(x) = x^2/2$ and so $\rho = 1$, the relative free entropy $\widetilde{\Sigma}_Q(\mu)$ is the inside of the square root in (2.2) and its minimizer is $\gamma_{0,2}$ so that Theorem 2.5 is a generalization of Theorem 2.4.

The next lemma will play a key role in our proof of the theorem.

Lemma 2.6. *Let $\tilde\mu, \tilde\nu \in \mathcal{M}(M_n^{sa})$ and $\hat\mu, \hat\nu$ be the mean eigenvalue distributions on $\mathbf{R}$ of $\tilde\mu, \tilde\nu$, respectively. Then*

$$W(\hat\mu, \hat\nu) \le \frac{1}{\sqrt{n}} W(\tilde\mu, \tilde\nu),$$

where $W(\tilde\mu, \tilde\nu)$ is the Wasserstein distance with respect to the distance induced by the Hilbert-Schmidt norm $\|\cdot\|_{HS}$ on M_n^{sa}.

Proof. For $A \in M_n^{sa}$ let $\lambda_1(A), \dots, \lambda_n(A)$ be the eigenvalues of A in increasing order with counting multiplicities. The mean eigenvalue distribution $\hat\mu$ is written as

$$\hat\mu = \int_{M_n^{sa}} \frac{1}{n}\left(\delta_{\lambda_1(A)} + \cdots + \delta_{\lambda_n(A)}\right) d\tilde\mu(A).$$

For each $\tilde\pi \in \Pi(\tilde\mu, \tilde\nu)$ define $\hat\pi \in \mathcal{M}(\mathbf{R} \times \mathbf{R})$ by

$$\hat\pi(G) := \iint_{M_n^{sa} \times M_n^{sa}} \frac{1}{n} \#\{i : (\lambda_i(A), \lambda_i(B)) \in G\}\, d\tilde\pi(A, B)$$

for Borel sets $G \subset \mathbf{R} \times \mathbf{R}$. Since

$$\hat\pi(F \times \mathbf{R}) = \int_{M_n^{sa}} \frac{1}{n} \{\#\{i : \lambda_i(A) \in F\}\, d\tilde\mu(A) = \hat\mu(F)$$

and similarly $\hat\pi(\mathbf{R} \times F) = \hat\nu(F)$ for $F \subset \mathbf{R}$, we get $\hat\pi \in \Pi(\hat\mu, \hat\nu)$ so that

$$\begin{aligned}
W(\hat\mu, \hat\nu)^2 &\le \iint_{\mathbf{R}\times\mathbf{R}} \frac{1}{2}(x-y)^2\, d\hat\pi(x, y) \\
&= \iint_{M_n^{sa} \times M_n^{sa}} \left\{ \iint_{\mathbf{R}\times\mathbf{R}} \frac{1}{2}(x-y)^2\, d\left(\frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i(A)} \otimes \delta_{\lambda_i(B)}\right)\right\} d\tilde\pi(A, B) \\
&= \frac{1}{n} \iint_{M_n^{sa} \times M_n^{sa}} \frac{1}{2} \sum_{i=1}^n \bigl(\lambda_i(A) - \lambda_i(B)\bigr)^2\, d\tilde\pi(A, B).
\end{aligned}$$

The famous Lidskii-Wielandt majorization for Hermitian matrices (see [3]) implies that

$$\sum_{i=1}^{n}\bigl(\lambda_i(A)-\lambda_i(B)\bigr)^2 \le \sum_{i=1}^{n}\lambda_i(A-B)^2 = \|A-B\|_{HS}^2$$

for all $A, B \in M_n^{sa}$. Therefore,

$$W(\hat{\mu},\hat{\nu})^2 \le \frac{1}{n}\iint_{M_n^{sa}\times M_n^{sa}} \frac{1}{2}\|A-B\|_{HS}^2\, d\tilde{\pi}(A,B),$$

and taking the infimum over $\tilde{\pi} \in \Pi(\tilde{\mu},\tilde{\nu})$ gives $W(\hat{\mu},\hat{\nu})^2 \le \frac{1}{n}W(\tilde{\mu},\tilde{\nu})^2$. $\square$

Proof of Theorem 2.5. First, let $\mu \in \mathcal{M}(\mathbf{R})$ be compactly supported, and suppose that the function $Q_\mu(x) := 2\int \log|x-y|\,d\mu(y)$ is finite and continuous on the whole $\mathbf{R}$. Choose $R > 0$ so that μ is supported in $[-R, R]$. For each $n \in \mathbf{N}$ consider the $n \times n$ self-adjoint random matrix $\lambda_n(Q_\mu; R) \in \mathcal{M}(M_n^{sa})$ supported in $\{A \in M_n^{sa} : \|A\|_\infty \le R\}$ as well as $\lambda_n(Q) \in \mathcal{M}(M_n^{sa})$ (see §§1.3 and §§1.4). Here, note that the condition (1.2) is automatically satisfied under the convexity assumption of Q. Since the corresponding large deviation principle guarantees the weak convergence of the mean eigenvalue distribution $\hat{\lambda}_n(Q)$ (resp. $\hat{\lambda}_n(Q_\mu; R)$) to μ_Q (resp. μ), Lemma 2.1 gives

$$W(\mu,\mu_Q) \le \liminf_{n\to\infty} W\bigl(\hat{\lambda}_n(Q_\mu;R),\hat{\lambda}_n(Q)\bigr). \tag{2.4}$$

By Lemma 2.6 we get

$$W\bigl(\hat{\lambda}_n(Q_\mu;R),\hat{\lambda}_n(Q)\bigr) \le \frac{1}{\sqrt{n}} W\bigl(\lambda_n(Q_\mu;R),\lambda_n(Q)\bigr). \tag{2.5}$$

Set $\Psi_n(A) := n\mathrm{Tr}_n(Q(A))$ for $A \in M_n^{sa}$; then $d\lambda_n(Q)(A) = \frac{1}{Z_n(Q)}e^{-\Psi_n(A)}\,dA$. Since $Q(x) - \frac{\rho}{2}x^2$ is convex on $\mathbf{R}$, so is

$$\Psi_n(A) - \frac{\rho n}{2}\|A\|_{HS}^2 = n\mathrm{Tr}_n\Bigl(Q(A) - \frac{\rho}{2}A^2\Bigr) \quad \text{on } M_n^{sa}.$$

Also, note that $\|\cdot\|_{HS}$ corresponds to the Euclidean norm on $\mathbf{R}^{n^2}$ under the isometry $A = [A_{ij}] \in M_n^{sa} \mapsto \bigl((A_{ii})_{1\le i\le n}, (\sqrt{2}A_{ij})_{i<j}\bigr) \in \mathbf{R}^{n^2}$. Hence, Theorem 2.2 implies that

$$W\bigl(\lambda_n(Q_\mu;R),\lambda_n(Q)\bigr) \le \sqrt{\frac{1}{\rho n} S\bigl(\lambda_n(Q_\mu;R),\lambda_n(Q)\bigr)}. \tag{2.6}$$

We now estimate $\frac{1}{n^2}S\bigl(\lambda_n(Q_\mu;R),\lambda_n(Q)\bigr)$. Notice

$$\frac{d\lambda_n(Q_\mu;R)}{d\lambda_n(Q)}(A) = \frac{\widetilde{Z}_n(Q)}{\widetilde{Z}_n(Q_\mu;R)}\exp\bigl(-n\mathrm{Tr}_n(Q_\mu(A)) + n\mathrm{Tr}_n(Q(A))\bigr)$$

on $(M_n^{sa})_R := \{A \in M_n^{sa} : \|A\|_\infty \le R\}$. Here, $\widetilde{Z}_n(Q)$ and $\widetilde{Z}_n(Q_\mu;R)$ are the normalization constants of the joint eigenvalue distributions of the random matrices $\lambda_n(Q)$ and $\lambda_n(Q_\mu;R)$, respectively (see §§1.3 and §§1.4). Hence, it follows that

$$
\begin{aligned}
&\frac{1}{n^2} S\bigl(\lambda_n(Q_\mu; R), \lambda_n(Q)\bigr) \\
&\quad = \frac{1}{n^2} \int_{(M_n^{sa})_R} \log \frac{d\lambda_n(Q_\mu; R)}{d\lambda_n(Q)}(A)\, d\lambda_n(Q_\mu; R)(A) \\
&\quad = \frac{1}{n^2} \log \widetilde{Z}_n(Q) - \frac{1}{n^2} \log \widetilde{Z}_n(Q_\mu) - \int_{(M_n^{sa})_R} \frac{1}{n} \mathrm{Tr}_n(Q_\mu(A))\, d\lambda_n(Q_\mu; R)(A) \\
&\qquad\qquad + \int_{(M_n^{sa})_R} \frac{1}{n} \mathrm{Tr}_n(Q(A))\, d\lambda_n(Q_\mu; R)(A) \\
&\quad = \frac{1}{n^2} \log \widetilde{Z}_n(Q) - \frac{1}{n^2} \log \widetilde{Z}_n(Q_\mu) - \int_{[-R,R]} Q_\mu(x)\, d\hat{\lambda}_n(Q_\mu; R)(x) \\
&\qquad\qquad + \int_{[-R,R]} Q(x)\, d\hat{\lambda}_n(Q_\mu; R)(x) \\
&\quad \longrightarrow B(Q) - B(Q_\mu; R) - \int_{[-R,R]} Q_\mu(x)\, d\mu(x) + \int_{\mathbf{R}} Q(x)\, d\mu(x) = \widetilde{\Sigma}_Q(\mu) \qquad (2.7)
\end{aligned}
$$

thanks to the fact that μ is the minimizer of the rate function (1.7) with Q_μ in place of Q, i.e.,

$$\int_{[-R,R]} Q_\mu(x)\, d\mu(x) + B(Q_\mu; R) = \Sigma(\mu).$$

Combining (2.4)–(2.7) altogether implies the inequality (2.3) under the continuity assumption of $Q_\mu(x)$.

Finally, let $\mu \in \mathcal{M}(\mathbf{R})$ be a general compactly supported measure. By the regularization method in [14, p. 216] we can choose a sequence $\{\mu_k\}$ of measures in $\mathcal{M}(\mathbf{R})$ with compact supports uniformly bounded such that $Q_{\mu_k}(x)$ is continuous on $\mathbf{R}$ for each k, $\mu_k \to \mu$ weakly and $\Sigma(\mu_k) \ge \Sigma(\mu)$ for all k. Hence, by Lemma 2.1 and the first case we have

$$
\begin{aligned}
W(\mu, \mu_Q) &\le \liminf_{n\to\infty} W(\mu_k, \mu_Q) \\
&\le \liminf_{k\to\infty} \sqrt{\frac{1}{\rho} \widetilde{\Sigma}_Q(\mu_k)} \le \sqrt{\frac{1}{\rho} \widetilde{\Sigma}_Q(\mu)},
\end{aligned}
$$

completing the proof. □

2.2. **The 1-dimensional torus case.** Next, we will present the free analog of transportation cost inequalities for measures on $\mathbf{T}$. The idea with use of special unitary random matrices is essentially the same as before. In the following we consider two kinds of Wasserstein distances between probability measures $\mu, \nu \in \mathcal{M}(\mathbf{T})$. The one is the Wasserstein distance with respect to the usual metric $|\zeta - \eta|$, $\zeta, \eta \in \mathbf{T}$, and the other is with respect to the geodesic distance (i.e., the angular distance) on $\mathbf{T}$. We write $W_{|\cdot|}(\mu, \nu)$ for the former and $W(\mu, \nu)$ for the latter. Of course, one has

$$W_{|\cdot|}(\mu, \nu) \le W(\mu, \nu), \qquad \mu, \nu \in \mathcal{M}(\mathbf{T}). \tag{2.8}$$

The theorem below is the free TCI for measures on $\mathbf{T}$ comparing the Wasserstein distance with the relative free entropy (1.9).

Theorem 2.7. *Let Q be a real-valued function on $\mathbf{T}$. If there exists a constant $\rho > -\frac{1}{2}$ such that $Q(e^{\mathrm{i}t}) - \frac{\rho}{2}t^2$ is convex on $\mathbf{R}$, then*

$$W_{|\cdot|}(\mu,\mu_Q) \le W(\mu,\mu_Q) \le \sqrt{\frac{2}{1+2\rho}\widetilde{\Sigma}_Q(\mu)} \tag{2.9}$$

for every $\mu \in \mathcal{M}(\mathbf{T})$.

The special case where $Q \equiv 0$ and $\rho = 0$ is

$$W_{|\cdot|}\left(\mu, \frac{d\theta}{2\pi}\right) \le W\left(\mu, \frac{d\theta}{2\pi}\right) \le \sqrt{-2\Sigma(\mu)}, \qquad \mu \in \mathcal{M}(\mathbf{T}). \tag{2.10}$$

We need the next lemma to prove the theorem. Note that the lemma and the proof remain valid when $\mathrm{SU}(n)$ is replaced by $\mathrm{U}(n)$.

Lemma 2.8. *Let $\tilde{\mu}, \tilde{\nu} \in \mathcal{M}(\mathrm{SU}(n))$ and $W(\tilde{\mu},\tilde{\nu})$ be the Wasserstein distance between $\tilde{\mu},\tilde{\nu}$ with respect to the geodesic distance on $\mathrm{SU}(n)$. Let $\hat{\mu},\hat{\nu}$ be the mean eigenvalue distributions on $\mathbf{T}$ of $\tilde{\mu},\tilde{\nu}$, respectively. Then*

$$W(\hat{\mu},\hat{\nu}) \le \frac{1}{\sqrt{n}} W(\tilde{\mu},\tilde{\nu}).$$

Proof. We use the symbol d for the geodesic distance on $\mathrm{SU}(n)$ as well as for that on $\mathbf{T}$. Define the optimal matching distance on $\mathbf{T}^n$ by

$$\delta(\zeta,\eta) := \min_{\sigma\in S_n} \sqrt{\sum_{i=1}^n d(\zeta_i, \eta_{\sigma(i)})^2}$$

for $\zeta = (\zeta_1,\dots,\zeta_n), \eta = (\eta_1,\dots,\eta_n) \in \mathbf{T}^n$. For $U \in \mathrm{SU}(n)$ let $\lambda(U) := (\lambda_1(U),\dots,\lambda_n(U))$ denote the element of $\mathbf{T}^n$ consisting of the eigenvalues of U with multiplicities and in counterclockwise order (i.e., $0 \le \arg\lambda_1(U) \le \cdots \le \arg\lambda_n(U) < 2\pi$). First, we prove

$$\delta(\lambda(U),\lambda(V)) \le d(U,V), \qquad U,V \in \mathrm{SU}(n). \tag{2.11}$$

For $U,V \in \mathrm{SU}(n)$ let $U(t)$ $(0 \le t \le 1)$ be the geodesic curve in $\mathrm{SU}(n)$ connecting U and V. By dividing the curve into several small pieces if necessary, we may assume that there is a smooth curve $A(t)$ $(0 \le t \le 1)$ in $\{A \in M_n^{sa} : \mathrm{Tr}_n(A) = 0\}$ such that $U(t) = e^{\mathrm{i}A(t)}$ for $0 \le t \le 1$. Let $0 = t_0 < t_1 < \cdots < t_K = 1$ be any partition of $A(t)$. For $1 \le k \le K$ we have

$$\begin{aligned}
\delta\bigl(\lambda(U(t_{k-1})),\lambda(U(t_k))\bigr) &\le \left\{\sum_{i=1}^n d\Bigl(e^{\mathrm{i}\lambda_i(A(t_{k-1}))}, e^{\mathrm{i}\lambda_i(A(t_k))}\Bigr)^2\right\}^{1/2} \\
&\le \left\{\sum_{i=1}^n |\lambda_i(A(t_{k-1})) - \lambda_i(A(t_k))|^2\right\}^{1/2} \\
&\le \|A(t_{k-1}) - A(t_k)\|_{HS} \\
&= d(U(t_{k-1}),U(t_k)) + o(t_k - t_{k-1}).
\end{aligned}$$

In the above, $\lambda_1(A_k),\dots,\lambda_n(A_k)$ are the eigenvalues of A_k in increasing order, and the third inequality is due to the Lidskii-Wielandt majorization. Therefore,

$$\delta(\lambda(U),\lambda(V)) \le \sum_{k=1}^K \delta\bigl(\lambda(U(t_{k-1})),\lambda(U(t_k))\bigr) \le d(U,V) + o(1)$$

so that (2.11) follows because $o(1) \to 0$ as $\max_k(t_k - t_{k-1}) \to 0$.

Now, for each $U, V \in \mathrm{SU}(n)$ let $\sigma_{U,V} \in S_n$ be such that

$$\delta(\lambda(U), \lambda(V)) = \left\{ \sum_{i=1}^{n} d\bigl(\lambda_i(U), \lambda_{\sigma_{U,V}(i)}(V)\bigr)^2 \right\}^{1/2}.$$

Of course, we can let $(U, V) \in \mathrm{SU}(n) \times \mathrm{SU}(n) \mapsto \sigma_{U,V} \in S_n$ measurable. For every $\tilde{\mu}, \tilde{\nu} \in \mathcal{M}(\mathrm{SU}(n))$ and $\tilde{\pi} \in \Pi(\tilde{\mu}, \tilde{\nu})$, define $\hat{\pi} \in \mathcal{M}(\mathbf{T} \times \mathbf{T})$ by

$$\hat{\pi}(G) := \iint_{\mathrm{SU}(n)\times\mathrm{SU}(n)} \frac{1}{n} \#\{i : (\lambda_i(U), \lambda_{\sigma_{U,V}(i)}(V)) \in G\}\, d\tilde{\pi}(U, V)$$

for Borel sets $G \subset \mathbf{T} \times \mathbf{T}$. Since for $F \subset \mathbf{T}$

$$\begin{aligned} \hat{\pi}(F \times \mathbf{T}) &= \int_{\mathrm{SU}(n)} \frac{1}{n} \#\{i : \lambda_i(U) \in F\}\, d\tilde{\mu}(U) = \hat{\mu}(F), \\ \hat{\pi}(\mathbf{T} \times F) &= \int_{\mathrm{SU}(n)} \frac{1}{n} \#\{i : \lambda_i(V) \in F\}\, d\tilde{\nu}(V) = \hat{\nu}(F), \end{aligned}$$

we have $\hat{\pi} \in \Pi(\hat{\mu}, \hat{\nu})$ so that

$$\begin{aligned} W(\hat{\mu}, \hat{\nu})^2 &\le \iint_{\mathbf{T}\times\mathbf{T}} \frac{1}{2} d(\zeta, \eta)^2\, d\hat{\pi}(\zeta, \eta) \\ &= \frac{1}{n} \iint_{\mathrm{SU}(n)\times\mathrm{SU}(n)} \frac{1}{2} \sum_{i=1}^{n} d\bigl(\lambda_i(U), \lambda_{\sigma_{U,V}(i)}(V)\bigr)^2\, d\tilde{\pi}(U, V) \\ &= \frac{1}{n} \iint_{\mathrm{SU}(n)\times\mathrm{SU}(n)} \frac{1}{2} \delta(\lambda(U), \lambda(V))^2\, d\tilde{\pi}(U, V) \\ &\le \frac{1}{n} \iint_{\mathrm{SU}(n)\times\mathrm{SU}(n)} \frac{1}{2} d(U, V)^2\, d\tilde{\pi}(U, V) \end{aligned}$$

thanks to (2.11). This implies $W(\hat{\mu}, \hat{\nu})^2 \le \frac{1}{n} W(\tilde{\mu}, \tilde{\nu})^2$. □

Proof of Theorem 2.7. The first inequality of (2.9) is obvious as noted in (2.8). To prove the second, we first assume:

(a) Q is harmonic on a neighborhood of the unit disk;
(b) the function $Q_\mu(\zeta) := 2 \int_{\mathbf{T}} \log|\zeta - \eta|\, d\mu(\eta)$ is finite and continuous on $\mathbf{T}$.

For each $n \in \mathbf{N}$ consider $n \times n$ special unitary random matrices $\lambda_n^{\mathrm{SU}}(Q)$ and $\lambda_n^{\mathrm{SU}}(Q_\mu)$ (see §§1.5). According to Theorem 1.1, the empirical eigenvalue distribution of $\lambda_n^{\mathrm{SU}}(Q)$ (resp. $\lambda_n^{\mathrm{SU}}(Q_\mu)$) satisfies the large deviation principle in the scale $1/n^2$ whose rate function is $\widetilde{\Sigma}_Q(\nu)$ (resp. $\widetilde{\Sigma}_{Q_\mu}(\nu)$). Moreover, note ([24, Theorem I.3.1]) that the equilibrium measure associated with Q_μ (or the minimizer of $\widetilde{\Sigma}_{Q_\mu}$) is the given μ. Thus, the large deviation principle (Theorem 1.1) guarantees that the weak convergence of the mean eigenvalue distribution $\hat{\lambda}_n^{\mathrm{SU}}(Q)$ (resp. $\hat{\lambda}_n^{\mathrm{SU}}(Q_\mu)$) to the equilibrium measure μ_Q (resp. μ), and hence Lemma 2.1 implies that

$$W(\mu, \mu_Q) \le \liminf_{n\to\infty} W\bigl(\hat{\lambda}_n^{\mathrm{SU}}(Q_\mu), \hat{\lambda}_n^{\mathrm{SU}}(Q)\bigr). \tag{2.12}$$

On the other hand, Lemma 2.8 gives

$$W\bigl(\hat{\lambda}_n^{\mathrm{SU}}(Q_\mu), \hat{\lambda}_n^{\mathrm{SU}}(Q)\bigr) \le \frac{1}{\sqrt{n}} W\bigl(\lambda_n^{\mathrm{SU}}(Q_\mu), \lambda_n^{\mathrm{SU}}(Q)\bigr). \tag{2.13}$$

Set $\Psi_n(U) := n\mathrm{Tr}_n\left(Q(U)\right)$ for $U \in \mathrm{SU}(n)$. Lemma 1.3 (ii) and (1.11) verify the Bakry and Emery criterion:

$$\mathrm{Ric}(\mathrm{SU}(n)) + \mathrm{Hess}(\Psi_n) \ge \left(\frac{n}{2} + n\rho\right) I_{n^2-1}. \tag{2.14}$$

Hence, Theorem 2.3 implies that

$$W\big(\lambda_n^{\mathrm{SU}}(Q_\mu), \lambda_n^{\mathrm{SU}}(Q)\big) \le \sqrt{\frac{2}{n+2n\rho} S\big(\lambda_n^{\mathrm{SU}}(Q_\mu), \lambda_n^{\mathrm{SU}}(Q)\big)}. \tag{2.15}$$

Similarly to the real line case, since

$$\frac{d\lambda_n^{\mathrm{SU}}(Q_\mu)}{d\lambda_n^{\mathrm{SU}}(Q)}(U) = \frac{\widetilde{Z}_n^{\mathrm{SU}}(Q)}{\widetilde{Z}_n^{\mathrm{SU}}(Q_\mu)} \exp\big(-n\mathrm{Tr}_n(Q_\mu(U)) + n\mathrm{Tr}_n(Q(U))\big), \qquad U \in \mathrm{SU}(n),$$

with the normalization constants $\widetilde{Z}_n^{\mathrm{SU}}(Q)$ and $\widetilde{Z}_n^{\mathrm{SU}}(Q_\mu)$ of the joint eigenvalue distributions (see §§1.5), we have

$$\begin{aligned}
&\frac{1}{n^2} S\big(\lambda_n^{\mathrm{SU}}(Q_\mu), \lambda_n^{\mathrm{SU}}(Q)\big) \\
&\quad = \frac{1}{n^2}\log \widetilde{Z}_n^{\mathrm{SU}}(Q) - \frac{1}{n^2}\log \widetilde{Z}_n^{\mathrm{SU}}(Q_\mu) \\
&\qquad - \int_{\mathrm{SU}(n)} \frac{1}{n}\mathrm{Tr}_n(Q_\mu(U))\, d\lambda_n^{\mathrm{SU}}(Q_\mu)(U) + \int_{\mathrm{SU}(n)} \frac{1}{n}\mathrm{Tr}_n(Q(U))\, d\lambda_n^{\mathrm{SU}}(Q_\mu)(U) \\
&\quad = \frac{1}{n^2}\log \widetilde{Z}_n^{\mathrm{SU}}(Q) - \frac{1}{n^2}\log \widetilde{Z}_n^{\mathrm{SU}}(Q_\mu) \\
&\qquad - \int_{\mathbf{T}} Q_\mu(\zeta)\, d\hat{\lambda}_n^{\mathrm{SU}}(Q_\mu)(\zeta) + \int_{\mathbf{T}} Q(\zeta)\, d\hat{\lambda}_n^{\mathrm{SU}}(Q_\mu)(\zeta) \\
&\quad \longrightarrow B(Q) - B(Q_\mu) - \int_{\mathbf{T}} Q_\mu(\zeta)\, d\mu(\zeta) + \int_{\mathbf{T}} Q(\zeta)\, d\mu(\zeta) = \widetilde{\Sigma}_Q(\mu),
\end{aligned} \tag{2.16}$$

where the last equality comes from that μ is the minimizer with $\widetilde{\Sigma}_{Q_\mu}(\mu) = 0$. The above (2.12), (2.13), (2.15) and (2.16) altogether prove the second inequality of (2.9) under assumptions (a) and (b).

Next, let Q be as stated in the theorem (hence Q is continuous on $\mathbf{T}$) and $\mu \in \mathcal{M}(\mathbf{T})$ be general. For $0 < r < 1$, we consider the Poisson integrals Q_r and p_r of Q and p, respectively; that is,

$$\begin{aligned}
Q_r\Big(e^{\mathrm{i}\theta}\Big) &:= \frac{1}{2\pi}\int_0^{2\pi} P_r(\theta - t) Q\Big(e^{\mathrm{i}t}\Big)\, dt, \\
p_r\Big(e^{\mathrm{i}\theta}\Big) &:= \frac{1}{2\pi}\int_0^{2\pi} P_r(\theta - t) p\Big(e^{\mathrm{i}t}\Big)\, dt
\end{aligned}$$

with the Poisson kernel $P_r(\theta) := (1-r^2)/(1-2r\cos\theta + r^2)$. Define $\mu_r \in \mathcal{M}(\mathbf{T})$ by $d\mu_r(\zeta) := p_r(\zeta)d\zeta$. Then it is plain to see that Q_r satisfies the assumption (a) and that μ_r does (b). The convexity assumption of Q in the theorem means that

$$\lambda Q\Big(e^{\mathrm{i}s}\Big) + (1-\lambda) Q\Big(e^{\mathrm{i}t}\Big) - Q\Big(e^{\mathrm{i}(\lambda s + (1-\lambda)t)}\Big) \ge \frac{\rho}{2}\lambda(1-\lambda)(t-s)^2$$

for all $s, t \in \mathbf{R}$ and $0 < \lambda < 1$. It is easy to check that each Q_r, $0 < r < 1$, satisfies the same convexity assumption so that

$$W(\mu_r, \mu_{Q_r}) \le \sqrt{\frac{2}{1+2\rho}\widetilde{\Sigma}_{Q_r}(\mu_r)}. \tag{2.17}$$

by what we have already shown.

It is known (see [13] and also [14, p. 224]) that $\mu_r \to \mu$ weakly and $\Sigma(\mu_r) \to \Sigma(\mu)$ as $r \nearrow 1$. Moreover, it is known (see [17, 5.3.2]) that $\|Q_r - Q\|_\infty \to 0$ as $r \nearrow 1$, where $\|\cdot\|_\infty$ means the uniform norm on $C(\mathbf{T})$. Since it is easily seen that

$$\left|\frac{1}{n^2}\log \tilde{Z}_n(Q_r) - \frac{1}{n^2}\log \tilde{Z}_n(Q)\right| \le \|Q_r - Q\|_\infty ,$$

we have $B(Q_r) \to B(Q)$ as $r \nearrow 1$. Therefore, we get

$$\lim_{r\nearrow 1} \widetilde{\Sigma}_{Q_r}(\mu) = \widetilde{\Sigma}_Q(\mu).$$

Choose any sequence $0 < r(k) < 1$ with $r(k) \to 1$ such that $\mu_{Q_{r(k)}} \to \mu_0 \in \mathcal{M}(\mathbf{T})$ weakly. By the upper semicontinuity of $\Sigma(\mu)$, we get

$$0 \le \widetilde{\Sigma}_Q(\mu_0) \le \liminf_{k\to\infty} \widetilde{\Sigma}_{Q_{r(k)}}(\mu_{Q_{r(k)}}) = 0$$

so that $\mu_0 = \mu_Q$. This shows that $\mu_{Q_r} \to \mu_Q$ weakly as $r \nearrow 1$ and

$$W(\mu, \mu_Q) \le \liminf_{r\nearrow 1} W(\mu_r, \mu_{Q_r})$$

thanks to Lemma 2.1. Hence, the desired inequality finally follows by taking the limit of (2.17). □

3. CONCLUDING REMARKS

In this section we collect some remarks and examples of computations.

3.1. **Use of special orthogonal random matrices.** For a real-valued continuous function Q, an $n \times n$ special orthogonal random matrix $\lambda_n^{\mathrm{SO}}(Q)$ is defined by

$$d\lambda_n^{\mathrm{SO}}(Q)(V) := \frac{1}{Z_n^{\mathrm{SO}}(Q)} \exp\Bigl(-\frac{n}{2}\mathrm{Tr}_n(Q(V))\Bigr)\, dV,$$

where dV is the Haar probability measure on the special orthogonal group $\mathrm{SO}(n)$. The joint eigenvalue distribution on $\mathbf{T}^{n-1}$ of $\lambda_n^{\mathrm{SO}}(Q)$ is

$$d\tilde{\lambda}_n^{\mathrm{SO}}(Q)(\zeta_1, \dots, \zeta_{n-1}) = \frac{1}{\widetilde{Z}_n^{\mathrm{SO}}(Q)} \exp\left(-\frac{n}{2}\sum_{i=1}^n Q(\zeta_i)\right) \prod_{1\le i<j\le n} |\zeta_i - \zeta_j| \prod_{i=1}^n d\zeta_i \qquad \text{with} \quad \zeta_n = (\zeta_1 \cdots \zeta_{n-1})^{-1}.$$

The large deviation is analogous to Theorem 1.1; the rate function is just $\frac{1}{2}\widetilde{\Sigma}_Q(\mu)$ and its minimizer is the same μ_Q. On the other hand, note that the Ricci curvature tensor of $\mathrm{SO}(n)$ is

$$\mathrm{Ric}(\mathrm{SO}(n)) = \frac{n-2}{4} I_{n(n-1)/2},$$

and the Bakry and Emery criterion in place of (2.14) is

$$\mathrm{Ric}(\mathrm{SO}(n)) + \mathrm{Hess}(\Psi_n) \ge \left(\frac{n-2}{4} + \frac{n}{2}\rho\right) I_{n(n-1)/2},$$

where $\Psi_n(V) := \frac{n}{2}\mathrm{Tr}_n(Q(V))$ for $V \in \mathrm{SO}(n)$. In this way, a special orthogonal random matrix model can be used as well to obtain the free TCI in Theorem 2.7. Similarly, the free TCI in Theorem 2.5 can be shown by using a real symmetric random matrix model

$$d\lambda_n^{\mathrm{real}}(Q)(T) := \frac{1}{Z_n^{\mathrm{real}}(Q)} \exp\Bigl(-\frac{n}{2}\mathrm{Tr}_n(Q(T))\Bigr)\, dT,$$

where $dT := \prod_{i\le j} dT_{ij}$ on $M_n(\mathbf{R})^{sa} \cong \mathbf{R}^{n(n+1)/2}$.

3.2. **Some computations: Best possibility of free TCI's.** A formula of the Wasserstein distance $W(\mu,\nu)$ for $\mu,\nu\in\mathcal{M}(\mathbf{R})$ is found in [27, 2.18] as follows:

$$W(\mu,\nu)^2=\frac{1}{2}\int_0^1\bigl(F^{-1}(t)-G^{-1}(t)\bigr)^2\,dt, \tag{3.1}$$

where F and G are the respective distribution functions of μ and ν. This can be conveniently used to compute $W(\gamma_{0,r_1},\gamma_{0,r_2})$ between semicircular measures:

$$W(\gamma_{0,r_1},\gamma_{0,r_2})=\frac{|r_1-r_2|}{2\sqrt{2}}. \tag{3.2}$$

When $Q(x):=\rho x^2/2$ on $\mathbf{R}$ with $\rho>0$, the equilibrium measure associated with Q is $\gamma_{0,2/\sqrt{\rho}}$, and for $\alpha>0$ we compute

$$\widetilde{\Sigma}_Q(\gamma_{0,2/\sqrt{\alpha}})=\frac{1}{2}\log\alpha+\frac{\rho}{2\alpha}-\frac{1}{2}\log\rho-\frac{1}{2}.$$

Therefore, we get

$$\lim_{\alpha\to 0}\frac{W(\gamma_{0,2/\sqrt{\alpha}},\gamma_{0,2/\sqrt{\rho}})^2}{\widetilde{\Sigma}_Q(\gamma_{0,2/\sqrt{\alpha}})}=\frac{1}{\rho}.$$

This shows that the bound $1/\rho$ in the free TCI (2.3) is the best possible.

For $2\le\lambda\le\infty$ the equilibrium measure associated with $Q(\zeta):=-(2/\lambda)\mathrm{Re}\,\zeta$ on $\mathbf{T}$ is

$$\nu_\lambda:=\left(1+\frac{2}{\lambda}\cos\theta\right)\frac{d\theta}{2\pi}\qquad\left(\text{with }\nu_\infty=\frac{d\theta}{2\pi}\right) \tag{3.3}$$

and $\Sigma(\nu_\lambda)=-1/\lambda^2$ (see [14, 5.3.10]). Since the density of ν_λ is symmetric with respect to the real-axis, it is obvious that $W(\nu_\lambda,d\theta/2\pi)$ is equal to the Wasserstein distance between ν_λ and $d\theta/2\pi$ regarded as measures on $\mathbf{R}$ supported in $[-\pi,\pi]$. Hence, by applying (3.1) we can easily compute $W(\nu_\lambda,d\theta/2\pi)^2=1/\lambda^2$ for $2\le\lambda\le\infty$. Hence, we notice that the bound 2 in the free TCI (2.10) cannot be smaller than 1; however it is unknown whether 2 is the best possible bound or not.

3.3. **Classical TCI vs. free TCI.** Both classical and free TCI's are formulated in terms of the same (quadratic) Wasserstein distance for measures, and thus it seems interesting to compare these two. However, in the case of measures on $\mathbf{R}$, the natural reference measures are Gaussian (not being compactly supported) in the classical case, while semicircular (being compactly supported) in the free case, and hence the question is irrelevant in this case. In the case of the uniform probability measure $d\theta/2\pi$ on $\mathbf{T}$, our free TCI is

$$W\left(\mu,\frac{d\theta}{2\pi}\right)\le\sqrt{-2\Sigma(\mu)},\qquad\mu\in\mathcal{M}(\mathbf{T}),$$

while to the authors' best knowledge the sharpest classical TCI is

$$W\left(\mu,\frac{d\theta}{2\pi}\right)\le\sqrt{S\left(\mu,\frac{d\theta}{2\pi}\right)},\qquad\mu\in\mathcal{M}(\mathbf{T}).$$

(The latter inequality is seen as follows. It is known (see [19, p. 94]) that the "spectral gap" and "logarithmic Sobolev constant" are the same number 1, and [23, Theorem 1] implies the desired inequality.) Now, if the relative free entropy happens to dominate the (usual) relative entropy up to a positive constant, then a free TCI would immediately follow from the classical one. However, this is not, and we indeed have the following examples:

(1) For an arbitrary $k\in\mathbf{N}$ and for large $n\in\mathbf{N}$, let us choose k disjoint intervals $[a_j(n),b_j(n)]$, $1\le j\le k$, in $\mathbf{T}=[0,2\pi)$ whose lengths are all $2\pi/kn$ and whose center

points are fixed independently of the choice n. Consider $\mu_k(n) \in \mathcal{M}(\mathbf{T})$ whose density is $\sum_{j=1}^k n\chi_{[a_j(n),b_j(n)]}$. Then we have

$$S\biggl(\mu_k(n), \frac{d\theta}{2\pi}\biggr) = \log n.$$

On the other hand, by a straightforward computation we see that, for a sufficiently large $n_0 \in \mathbf{N}$, there are constants $c_k < C_k$ depending only on k such that

$$c_k + \frac{\log n}{k} \le -\Sigma\left(\mu_k(n)\right) \le C_k + \frac{\log n}{k} \quad \text{for } n \ge n_0,$$

and thus

$$\frac{-\Sigma\left(\mu_k(n)\right)}{S\bigl(\mu_k(n), \frac{d\theta}{2\pi}\bigr)} \to \frac{1}{k} \quad \text{as } n \to \infty.$$

The computation is somewhat similar to a free entropy dimension computation for single variables; see [25, Proposition 6.1] for example.

(2) For the measure ν_λ $(2 < \lambda < \infty)$ in (3.3), with the help of a table on integration formulas, we can compute

$$S\biggl(\nu_\lambda, \frac{d\theta}{2\pi}\biggr) = \log\Biggl(\frac{1}{2}\Biggl(1 + \sqrt{1 - \frac{4}{\lambda^2}}\Biggr)\Biggr) + 1 + \frac{4}{\lambda\sqrt{\lambda^2 - 4}} - \frac{1}{\sqrt{1 - \frac{4}{\lambda^2}}},$$

and hence we get

$$\frac{S\bigl(\nu_\lambda, \frac{d\theta}{2\pi}\bigr)}{-\Sigma\bigl(\nu_\lambda\bigr)} \longrightarrow 0 \quad \text{as } \lambda \to \infty.$$

These examples tell us that the minus free entropy $-\Sigma(\mu)$ cannot be compared with the relative entropy $S(\mu, d\theta/2\pi)$.

Acknowledgments. We are grateful to Professor Ph. Biane who pointed out the formula (3.2) to us.

References

[1] D. Bakry and M. Emery, Diffusion hypercontractives, in *Séminaire Probabilités XIX*, Lecture Notes in Math., Vol. 1123, Springer-Verlag, 1985, pp. 177–206.

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

[3] R. Bhatia, *Matrix Analysis*, Springer-Verlag, New York, 1996.

[4] Ph. Biane, Logarithmic Sobolev inequalities, matrix models and free entropy, *Acta Math. Sinica* **19**, No. 3 (2003), 1–11.

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

[6] Ph. Biane and D. Voiculescu, A free probabilistic analogue of the Wasserstein metric on the trace-state space, *Geom. Funct. Anal.* **11** (2001), 1125–1138.

[7] P. A. Deift, *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach*, Courant Lecture Notes in Math., Vol. 3, New York Univ., Courant Inst. Math. Sci., New York, 1999.

[8] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Second edition. Applications of Mathematics, Vol. 38, Springer-Verlag, New York, 1998.

[9] J. D. Deuschel and D. W. Stroock, *Large deviations*, Pure and Applied Mathematics, Vol. 137. Academic Press, Inc., Boston, MA, 1989.

[10] W. F. Donoghue, Jr., *Monotone Matrix Functions and Analytic Continuation*, Springer-Verlag, Berlin-Heidelberg-New York, 1974.

[11] F. Hiai, M. Mizuo and D. Petz, Free relative entropy for measures and a corresponding perturbation theory, *J. Math. Soc. Japan* **54** (2002), 679–718.

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

[13] 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.

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

[15] F. Hiai, D. Petz and Y. Ueda, in preparation.

[16] A. W. Knapp, *Representation Theory of Semisimple Groups, An Overview Based on Examples*, Princeton University Press, Princeton, 1986.

[17] P. Koosis, *Introduction to H_p Spaces*, Second edition, Cambridge Tracts in Math., Vol. 115, Cambridge Univ. Press, Cambridge, 1998.

[18] M. Ledoux, Concentration of measures and logarithmic Sobolev inequalities, In *Séminaire de Probabilités XXXIII*, Lecture Notes in Math., Vol. 1709, Springer-Verlag, 1999, pp. 120–216.

[19] M. Ledoux, *The Concentration of Measure Phenomenon*, Mathematical Surveys and Monographs, Vol. 89, Amer. Math. Soc., Providence, 2001.

[20] M. L. Mehta, *Random Matrices*, Second edition, Academic Press, Boston, 1991.

[21] J. Milnor, Curvature of left invariant metrics on Lie groups, *Adv. Math.*, **21** (1976), 293–329.

[22] M. Ohya and D. Petz, *Quantum Entropy and Its Use*, Springer-Verlag, Berlin, 1993.

[23] F. Otto and C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, *J. Funct. Anal.* **173** (2000), 361–400.

[24] E. B. Saff and V. Totik, *Logarithmic Potentials with External Fields*, Springer-Verlag, Berlin-Heidelberg-New York, 1997.

[25] D. Shlyakhtenko, Free fisher information with respect to a complete positive map and cost of equivalence relations, *Comm. Math. Phys.* **218** (2001), 133–152.

[26] M. Talagrand, Transportation cost for Gaussian and other product measures, *Geom. Funct. Anal.* **6** (1996), 587–600.

[27] C. Villani, *Topics in Optimal Transportation*, Grad. Studies in Math., Vol. 58, Amer. Math. Soc., Providence, 2003.

[28] D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory, I, *Comm. Math. Phys.* **155** (1993), 71–92.

[29] D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory, II, *Invent. Math.* **118** (1994), 411–440.

[30] D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory, V, Noncommutative Hilbert transforms, *Invent. Math.* **132** (1998), 189–227.

[31] D. Voiculescu, The analogue of entropy and of Fisher's information measure in free probability theory VI: Liberation and mutual free information, *Adv. Math.* **146** (1999), 101–166.

Graduate School of Information Sciences, Tohoku University, Aoba-ku, Sendai 980-8579, Japan

Department for Mathematical Analysis, Budapest University of Technology and Economics, H-1521 Budapest XI., Hungary

Graduate School of Mathematics, Kyushu University, Fukuoka 810-8560, Japan