Free transportation cost inequalities via random matrix approximation

Dénes Petz

Alfréd Rényi Institute of Mathematics

Hungarian Academy of Sciences

POB 127, H-1364 Budapest, Hungary

E-mail: petz@math.bme.hu **Home page:** http://www.math.bme.hu/˜petz

Abstract

The result presented here is a joint work with **Hiai** and **Ueda** and extends slightly the free transportation cost inequality obtained by **Biane** and **Voiculescu**. Their method is based on the complex Burger's equation parallelly to the work of Otto and Villani, while we benefit from large deviation theorems for certain random matrix ensembles.

Main references

- 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.
- F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*, Mathematical Surveys and Monographs, Vol. 77, Amer. Math. Soc., Providence, 2000.
- F. Hiai, D. Petz and Y. Ueda, Free transportation cost inequalities via random matrix approximation, to be published in Prob. Theory Rel. Fields.
- 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.

Contents

Distances for measures

The *relative entropy*

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

is useful in information theory but it is not a metric.

The *Wasserstein distance* between $\mu, \nu \in \mathcal{M}(\mathbb{R}^n)$ is defined by

$$W(\mu,\nu) := \inf_{\pi \in \Pi(\mu,\nu)} \sqrt{\iint \frac{1}{2} d(x,y)^2\, d\pi(x,y)},$$

where $d(x,y) = \|x-y\|_2$ and $\Pi(\mu,\nu)$ denotes the set of all probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals μ and ν, i.e., $\pi(\,\cdot \times \mathbb{R}^n) = \mu$ and $\pi(\mathbb{R}^n\, \cdot\,) = \nu$.

Talagrand's inequality

In 1996, M. Talagrand obtained an interesting inequality, called *the transportation cost inequality* (TCI), comparing the (quadratic) Wasserstein distance with the square root of the relative entropy:

$$W(\mu_0, \nu) \leq \sqrt{\frac{1}{\rho} S(\mu_0, \nu)},$$

where ν is a Gaussian measure on $\mathbb{R}^n$,

$$d\nu(x) = \frac{1}{Z} \exp\big(-\psi(x)\big)$$

with $\psi(x) = \rho \|x\|^2/2$. Actually, the inequality holds under the more general condition that $\psi(x) - \rho\|x\|^2/2$ is convex for a positive constant ρ.

Random matrices

A probability measure $P_n \in M_n(\mathbb{C})^{sa}$ is called a random self-adjoint $n \times n$ matrix. All the measures we consider here are absolutely continuous with respect to the Lebesgue measure

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

and unitarily invariant in the sense that A and VAV^* are identically distributed for any unitary V. Such measures are uniquely determined by the joint distribution of the eigenvalues.

Eigenvalue densities

A measure $P_n \in M_n(\mathbb{C})^{sa}$ induces the *empirical eigenvalue density* $\tilde{P}_n \in \mathcal{M}(\mathcal{M}(\mathbb{R}))$

$$\tilde{P}_n(G) := \text{Prob}\left(\frac{\delta_{\lambda_1(\omega)} + \cdots + \delta_{\lambda_n(\omega)}}{n} \in G\right)$$

for $G \subset \mathcal{M}(\mathbb{R})$, where $\lambda_1(\omega), \ldots, \lambda_n(\omega)$ are the (random) eigenvalues and *the mean eigenvalue density* $\hat{P}_n \in \mathcal{M}(\mathbb{R})$

$$\hat{P}_n := E(\tilde{P}_n).$$

The empirical eigenvalue density appears in large deviation theorems.

Large deviation result

If P_n is defined as

$$dP_n = \frac{1}{Z_n} \exp(-n \mathrm{Tr}\, Q(A)) dA$$

and $Q : \mathbb{R} \to \mathbb{R}$ satisfies a growth condition,

$$\varepsilon\psi(x) - \log|x| \to +\infty \quad \text{as} \quad x \to \pm\infty \quad \text{for all } \varepsilon > 0,$$

then the large deviation theorem holds and the rate function is

$$\tilde{\Sigma}_Q(\mu) := -\Sigma(\mu) + \int_{\mathbb{R}} Q(x)\, d\mu(x) + B$$

for $\mu \in \mathcal{M}(\mathbb{R})$. [G. Ben Arous and A. Guionnet, 1997]

The rate function

The double integral

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

is called *logarithmic energy*, or *free entropy*.

The rate function is weakly lower semi-continuous, strictly convex and has a unique minimizer of compact support. The minimizer has a good characterization and can be computed in a few concrete example. For instance, if $\psi(x) = x^2/2$, then the minimizer is the standard semi-circle law. This example is related to Wigner's theorem.

The free TCI inequality

Theorem 1 *Assume that* $\psi : \mathbb{R} \to \mathbb{R}$ *is continuous and* $\psi(x) - \rho x^2/2$ *is convex with certain* $\rho > 0$ *and let* Φ *be defined as*

$$\Phi(\mu) = -\Sigma(\mu) + \int \psi(x) d\mu(x),$$

Then

$$W(\mu_0, \nu) \leq \sqrt{\frac{1}{\rho}(\Phi(\mu_0) - \Phi_0)},$$

where Φ_0 *is the minimum of the functional* Φ, ν *is the minimizer and* $\mu_0 \in \mathcal{M}(\mathbb{R})$ *is an arbitrary compactly supported measure.*

Idea of the proof (1)

1. Note that the assumption on $\psi(x)$ implies that the growth condition holds and the measures P_n defined by

$$dP_n = \frac{1}{Z_n} \exp(-n \mathrm{Tr}\, \psi(A)) dA$$

satisfy the large deviation principle and the rate function is

$$\Phi(\mu) - \Phi_0 = -\Sigma(\mu) + \int \psi(x)\, d\mu(x) + B$$

which appears also in the right-hand-side of our inequality. The large deviation gives automatically an approximation of ν by the sequence $\hat{P}_n$ and $\hat{P}_n \to \nu$ weakly.

Idea of the proof (2)

2. Assume that μ_0 is supported in an interval $[-R, R]$ and set

$$Q^{\infty}_{\mu_0}(x) := \begin{cases} 2\int_{\mathbb{R}} \log|x-y|\, d\mu_0(y) & \text{if} \quad x \in [-R, R], \\ +\infty & \text{otherwise.} \end{cases}$$

If the random matrices P'_n are defined now by this potential, then P'_n is supported on the set $\{A \in M_n(\mathbb{C})^{sa} : \|A\| \le R\}$ and supp $\hat{P'}_n \subset [-R, R]$. The large deviation holds and the rate function is

$$-\Sigma(\mu) + \int Q_{\mu_0}(x)\, d\mu(x) + B'$$

for $\mu \in \mathcal{M}([-R, R])$. It follows from the large deviation that $\hat{P}'_n \to \mu_0$ weakly.

Idea of the proof (3)

3. From the lower semi-continuity we have

$$W(\mu_0, \nu) \leq \liminf_{n \to \infty} W(\hat{P}'_n, \hat{P}_n)$$

and the basic lemma gives

$$W(\hat{P}', \hat{P}) \leq \frac{1}{\sqrt{n}} W(P', P).$$

Next we want to use Talagrand's inequality in more variables to estimate $W(P', P)$.

Idea of the proof (4)

4. Set $\psi_n(A) := n\mathrm{Tr}\,(\psi(A))$ for $A \in M_n(\mathbb{C})^{sa}$; then

$$dP_n(A) = \frac{1}{Z_n} e^{-\psi_n(A)}\, dA.$$

Since $\psi(x) - \frac{\rho}{2}x^2$ is convex on $\mathbb{R}$, so is

$$\psi_n(A) - \frac{\rho n}{2}\|A\|_{HS}^2 = n\mathrm{Tr}\left(\psi(A) - \frac{\rho}{2}A^2\right) \quad \text{on} \quad M_n^{sa}.$$

Talagrand's inequality implies that

$$W(P_n', P_n) \leq \sqrt{\frac{1}{\rho n} S(P_n', P_n)}$$

and we have to investigate the limit of $n^{-2}S(P_n', P_n)$.

The basic lemma

Lemma 1 *Let* $P, P' \in \mathcal{M}(M_n^{sa})$ *and* $\hat{P}, \hat{P}'$ *be the mean eigenvalue density on* $\mathbb{R}$ *of* P, P'*, respectively. Then*

$$W(\hat{P}, \hat{P}') \leq \frac{1}{\sqrt{n}} W(P, P'),$$

where $W(P, P')$ *is the Wasserstein distance with respect to the distance induced by the Hilbert-Schmidt norm on* M_n^{sa}.

For $A \in M_n(\mathbb{C})^{sa}$ let $\lambda_1(A), \ldots, \lambda_n(A)$ be the eigenvalues of A in increasing order with multiplicities. The mean eigenvalue density $\hat{P}$ is written as

$$\hat{P} = \int \frac{1}{n} \left(\delta_{\lambda_1(A)} + \cdots + \delta_{\lambda_n(A)} \right) dP(A).$$

Proof of the basic lemma (1)

For each $\pi \in \Pi(P, P')$ define $\hat{\pi} \in \mathcal{M}(\mathbb{R} \times \mathbb{R})$ by

$$\hat{\pi}(G) := \iint \frac{1}{n} \#\{i : (\lambda_i(A), \lambda_i(B)) \in G\} \, d\pi(A, B)$$

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

Since

$$\hat{\pi}(F \times \mathbb{R}) = \int \frac{1}{n} \#\{i : \lambda_i(A) \in F\} \, d\pi(A) = \hat{P}(F)$$

and similarly $\hat{\pi}(\mathbb{R} \times F) = \hat{P}'(F)$ for $F \subset \mathbb{R}$. In this way, we have shown that $\hat{\pi} \in \Pi(\hat{P}, \hat{P}')$.

Proof of the basic lemma (2)

$$W(\hat{P}, \hat{P}')^2 \leq \iint \frac{1}{2}(x-y)^2 \, d\hat{\pi}(x,y)$$

$$= \iint \Big(\iint \frac{1}{2}(x-y)^2 \, d\Big(\frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(A)} \otimes \delta_{\lambda_i(B)}\Big)\Big) \, d\tilde{\pi}(A,B)$$

$$= \frac{1}{n} \iint \frac{1}{2} \sum_{i=1}^{n} (\lambda_i(A) - \lambda_i(B))^2 \, d\tilde{\pi}(A,B).$$

The *Lidskii-Wielandt majorization* for self-adjoint matrices implies that

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

for all $A, B \in M_n(\mathbb{C})^{sa}$.

End of the proof

$$W(\hat{P}, \hat{P}')^2 \leq \frac{1}{n} \iint \frac{1}{2} \|A - B\|_{HS}^2 \, d\pi(A, B),$$

and taking the infimum over $\pi \in \Pi(P, P')$ gives

$$W(\hat{P}, \hat{P}')^2 \leq \frac{1}{n} W(P, P')^2.$$

Moving to the circle

Next, we will present the free analog of transportation cost inequalities for measures on the circle $\mathbf{T}$.

The idea is to use of the random matrix ensemble of special unitary random matrices and the large deviation for the empirical eigenvalue density.

The program is essentially the same as before but the unitaries form a more complicated manifold compared with the flat space of self-adjoint matrices.

Free TCI on the circle

Theorem 2 *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*

$$W(\mu, \mu_Q) \leq \sqrt{\frac{2}{1+2\rho}\widetilde{\Sigma}_Q(\mu)}$$

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

$$\widetilde{\Sigma}_Q(\mu) = -\iint_{\mathbf{T}^2} \log|\zeta - \zeta'|\, d\mu(\zeta)\, d\mu(\zeta') + \int_{\mathbf{T}} Q(\zeta)\, d\mu(\zeta) + B(Q).$$

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

$$W\left(\mu, \frac{d\theta}{2\pi}\right)^2 \leq -2 \iint_{\mathbf{T}^2} \log|\zeta - \zeta'|\, d\mu(\zeta)\, d\mu(\zeta')$$

TCI on a manifold

In the next theorem [Bakry-Emery and Otto-Villani] let M be an m-dimensional smooth complete Riemannian manifold equipped with the geodesic distance $d(x, y)$ and the volume measure dx.

Theorem 3 *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* $\mathrm{Ric}(M) + \mathrm{Hess}(\Psi) \geq \rho I_m$ *holds with a constant* $\rho > 0$, *then*

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

Riemannian structure on SU(n)

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

and the Bakry and Emery criterion gets the form

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

where $\Psi_n(V) := \frac{n}{2}\mathrm{Tr}\left(Q(V)\right)$ for $V \in \mathrm{SO}(n)$ since one computes that

$$\mathrm{Hess}(\Psi_n) \geq \rho I_{n^2-1}$$

under our hyporhesis.

Comparison of two distances

We compare the quadratic Wasserstein distances on $\mathcal{M}(\mathrm{SU}(n))$ and $\mathcal{M}(\mathbf{T})$. The first distance is defined with respect to the geodesic distance on $\mathrm{SU}(n)$. The Wasserstein distances between probability measures $\mu, \nu \in \mathcal{M}(\mathbf{T})$ is defined by means of the angular (or geodesic) distance on $\mathbf{T}$.

Lemma 2 *Let $\tilde{\mu}, \tilde{\nu} \in \mathcal{M}(\mathrm{SU}(n))$ and $W(\tilde{\mu}, \tilde{\nu})$ be the Wasserstein distance between $\tilde{\mu}, \tilde{\nu}$. Let $\hat{\mu}, \hat{\nu}$ be the mean eigenvalue distributions on $\mathbf{T}$ of $\tilde{\mu}, \tilde{\nu}$, respectively. Then*

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

Large deviation on SU(n)

Let Q be a real-valued continuous function on $\mathbf{T}$. The special unitary random matrix associated with Q is defined by

$$\frac{1}{Z_n} \exp(-n\mathrm{Tr}\,_n(Q(U)))\, dU,$$

where dU is the Haar probability measure on $\mathrm{SU}(n)$ and Z_n is a normalization constant.

Theorem 4 *The empirical eigenvalue density of the above unitaries satisfies the large deviation principle in the scale* $1/n^2$ *with the rate function*

$$-\iint_{\mathbf{T}^2} \log|\zeta - \zeta'|\, d\mu(\zeta)\, d\mu(\zeta') + \int_{\mathbf{T}} Q(\zeta)\, d\mu(\zeta) + B(Q).$$