Local functional inequalities in one-dimensional free probability ☆

Ionel Popescu [a,b,∗]

[a] *School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA*
[b] *Institute of Mathematics of Romanian Academy, 21, Calea Grivitei Street, 010702-Bucharest, Sector 1, Romania*

Abstract

In this note we introduce and prove local and potential independent transportation, Log-Sobolev and HWI inequalities in one-dimensional free probability on compact intervals which are sharp. We recover using this approach a free transportation inequality on the whole real line which was put forward recently by Maïda and Maurel-Segala (2012) [10]. Our method is based on the operator theoretic approach developed by Ledoux and Popescu [7] to deal with the free Poincaré inequality.
© 2013 Elsevier Inc. All rights reserved.

Keywords: Free probability; Functional inequalities

1. Introduction

There is a large literature on functional inequalities in the classical case. Some of these inequalities apply to the case of random matrices and produce, as the dimension grows to infinity, interesting functional inequalities in the limit. The main connection between the random matrices and free probability is due to the main result of Voiculescu [18,17] which states that large

∗ Correspondence to: School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA.
E-mail address: ipopescu@math.gatech.edu.

random matrices become asymptotically free as the dimension grows to infinity. This is a very rich bridge from free probability to random matrices and back.

Among the counterparts of classical inequalities, we mention the transportation which was first discussed by Biane and Voiculescu in [3] for the quadratic potentials and was inspired by the work of Otto and Villani [11]. We describe now the statement of this inequality as it is an important point in the economy of this note.

For a given potential $V : \mathbb{R} \to \mathbb{R}$, the logarithmic energy with external field V of a probability measure on $\mathbb{R}$ is defined by

$$E_V(\mu) = \int V \, d\mu - \iint \log |x - y| \mu(dx)\mu(dy).$$

It is a standard result (cf. [15]) that under some mild growth of V, there is a unique probability measure μ_V, which minimizes the functional E_V. If we set $E_V(\mu|\mu_V) = E_V(\mu) - E_V(\mu_V)$ this plays the analog of the entropy in the classical case. The transportation inequality associated to V states that there is a positive ρ such that for any compactly supported probability measure μ

$$\rho W_2^2(\mu, \mu_V) \leqslant E_V(\mu|\mu_V) \tag{1.1}$$

where $W_2(\mu, \nu)$ is the Wasserstein distance based on quadratic cost function given by

$$W_2(\mu, \nu) = \left\{ \inf_{\pi \in \Pi(\mu,\nu)} \int |x - y|^2 \pi(dx\,dy) \right\}^{1/2}$$

for measures μ, ν of finite second moment with $\Pi(\mu, \nu)$ being the set of probability measures on $\mathbb{R} \times \mathbb{R}$ with marginals μ and ν. This was first proved for the case of $V(x) = x^2/2$ by Biane and Voiculescu in [3] using complex Burger's equation and then for the case of $V(x) - \rho x^2$ convex in [6] using random matrices. Yet, another direct approach is using tools from mass transportation and is given in [14,8].

Another classical inequality which found a natural analog in the free probability world is the Log-Sobolev which was introduced in a certain form by Voiculescu in [19] and then proved to be equivalent to the one which is most common now by Biane and Speicher in [2]. With the notation from above, it states that there is a positive ρ such that for any probability measure μ,

$$4\rho E_V(\mu|\mu_V) \leqslant I(\mu|\mu_V) \tag{1.2}$$

where

$$I(\mu|\mu_V) = \int \left(H\mu - V' \right)^2 d\mu \quad \text{with } H\mu(x) = \begin{cases} p.v. \int \frac{2}{x-y} \mu(dy) & \frac{d\mu}{dx} \in L^3(\mathbb{R}), \\ +\infty & \text{otherwise} \end{cases}$$

where the integral in the definition of the Hilbert transform $H\mu$ is in the principal value sense. This inequality has received a random matrix proof in [1] and in the case $V(x) - \rho x^2$ is convex. Another alternative proof using tools from the mass transportation is given in [8].

Notice that so far these inequalities require some convexity on the potential V. A natural question is to ask if there is a transportation or Log-Sobolev without the convexity assumption on V. For the transportation case there is a version put forward recently by Maïda and Maurel-Segala

in [10] in which the main condition on V is a quadratic growth at infinity and the base metric W_2 is replaced by W_1, a weaker metric. They use this to say something about the concentration of the empirical distribution of eigenvalues of random matrices with general potentials.

For the Log-Sobolev case, without convexity assumption, the statement from (1.2) cannot be true as it is. What is the natural replacement of (1.2) if we drop the convexity on V is not clear.

At one end is the case of (strong) convex potentials where both (1.1) and (1.2) are well understood but as soon as we loose the convexity property, the inequalities in discussion become problematic. At the other end of the spectrum is the case of these inequalities which are in fact potential independent. It is this topic which is under investigation here.

Now we describe a little bit the main results and how the paper is organized.

To formulate the question clearly, the first change is that in place of the entropy $E_V(\mu|\mu_V)$ we use a very closely related quantity which for any two probability measures μ, ν, is given by

$$\mathcal{H}(\mu, \nu) = -\iint \log|x - y|(\mu - \nu)(dx)(\mu - \nu)(dy).$$

As one can see, this is independent of the potential V, but it is not really very different from $E_V(\mu|\mu_V)$ (see for instance, (4.2) below).

One of the main results of this note is that for probability measures supported on $[-2, 2]$,

$$W_1^2(\mu, \nu) \leqslant 2\mathcal{H}(\mu, \nu) \tag{1.3}$$

where the inequality is actually sharp. This inequality can be seen as a local version of the transportation inequality which is at the same time potential independent. By scaling, this can be extended to probability measures on any compact interval, and thus it can be really interpreted as some form of universal transportation inequality. This is treated in Section 2.

In Section 3 we show that the metric W_1 is optimal in (1.3) and cannot be replaced by any other W_p with $p > 1$.

The interesting fact is that now if we take a potential V, with at least quadratic growth at infinity, then we can actually turn the local version of (1.3) into a global transportation inequality which states that for some $C > 0$ and any probability measure μ on $\mathbb{R}$,

$$CW_1^2(\mu, \mu_V) \leqslant E_V(\mu|\mu_V),$$

which is the result from [10]. This is the content of Section 4. The idea we use here is borrowed from the mass transportation techniques to combine the local transportation with the growth of V at infinity.

Section 5 is dedicated to a local version of the Log-Sobolev. The first thing we need to set up properly is the analog of the Fisher information, $I_V(\mu|\mu_V)$. When restricted to $[-2, 2]$, the version we propose is the following

$$\mathcal{J}(\mu, \nu) = \begin{cases} \int (H\mu - H\nu)^2 \, d\alpha & \text{if } H\mu, H\nu \in L^2(\alpha), \\ +\infty & \text{otherwise} \end{cases}$$

where α is the semicircle law on $[-2, 2]$ and $H\mu$ is the Hilbert transform of the measure μ. One of the main results in this section is that

$$\mathcal{J}(\mu, \nu) = \begin{cases} 2\int(\frac{d\mu}{d\beta} - \frac{d\nu}{d\beta})^2\, d\beta & \text{if } \frac{d\mu}{d\beta}, \frac{d\nu}{d\beta} \in L^2(\beta), \\ +\infty & \text{otherwise} \end{cases}$$

with β being the arcsine law on $[-2, 2]$. This last equality is nothing but the interesting property of the Hilbert transform which says that the map

$$H_\beta : L^2(\beta) \to L^2(\alpha), \qquad H_\beta \phi = H(\phi\, d\beta)$$

is an isometry up to a multiplication by a constant. Using these properties, we prove the following local version of the free Log-Sobolev. For any probability measures on $[-2, 2]$,

$$2\mathcal{H}(\mu, \nu) \leqslant \mathcal{J}(\mu, \nu). \tag{1.4}$$

As in the case of local transportation this turns out to be sharp.

We continue this discussion in Section 6 of the local Log-Sobolev in which the L^2 norm of $H\mu - H\nu$ from the definition of $\mathcal{J}$ above is replaced by the square of the L^p norm with $1 < p < 2$. It is shown that if such a Log-Sobolev holds true, then necessarily $3/2 \leqslant p$ but it is posted as an open problem if $p = 3/2$ is the optimal threshold for which the inequality is satisfied. At any rate, even though $p = 3/2$ does not produce an L^p version of the Log-Sobolev, it is still natural to look for the smallest which does produce such an inequality.

Finally in Section 7 we discuss a version of the celebrated Otto–Villani HWI inequality which links together W_1, $\mathcal{H}$ and $\mathcal{J}$. This is a refinement of the Log-Sobolev inequality.

On the technical side, the main tools we use here are borrowed from the operator theoretical approach to the free Poincaré inequality put forward in [8] and, as we already mentioned, for the global version of transportation inequality we employ some tools from the classical mass transportation.

2. Potential independent transportation inequality on $[-L, L]$

We will treat here essentially the case of measures on $[-2, 2]$. The case of measures on $[-L, L]$ following by simple scaling.

Given a $p \geqslant 1$ and two measures μ, ν on the real line such that $\int |x|^p \mu(dx)$ and $\int |x|^p \nu(dx)$ are both finite, we define

$$W_p(\mu, \nu) = \left\{ \inf_{\pi \in \Pi(\mu, \nu)} \int |x - y|^p \pi(dx\, dy) \right\}^{1/p} \tag{2.1}$$

where here $\Pi(\mu, \nu)$ denotes the set of probability measures on $\mathbb{R}^2$ with marginals μ, ν. W_p is a metric for the weak topology on the set of probability measures with pth finite moment.

We will be interested in W_1 which can also be characterized as

$$W_1(\mu, \nu) = \sup \left\{ \int g\, d(\mu - \nu), \big|g(x) - g(y)\big| \leqslant |x - y| \right\}. \tag{2.2}$$

W_1 is a distance for the topology of weak convergence of probability measures with finite first moment.

Another description of the distance W_p is given by the following (see for example [16, p. 75]). If μ, ν are two probability measures on $\mathbb{R}$ such that ν does not have atoms, then there is a unique non-decreasing map θ such that $\theta_\# \nu = \mu$ (i.e. $\mu(A) = \nu(\theta^{-1}(A))$). In addition,

$$W_1(\mu, \nu) = \int |\theta(x) - x| \nu(dx). \tag{2.3}$$

Next we define the *free reduced relative entropy* of two compactly supported measures $\mu, \nu \in \mathcal{P}(\mathbb{R})$ to be given by

$$\mathcal{H}(\mu, \nu) = -\iint \log |x - y| (\mu - \nu)(dx)(\mu - \nu)(dy). \tag{2.4}$$

It is well known [15, Lemma 1.8] or [5, Lemma 6.41] that $\mathcal{H}(\mu, \nu) \geqslant 0$ with equality if and only if $\mu = \nu$. For the integrability properties see a detailed discussion in [5, p. 142] the only thing we point out here being that $\mathcal{H}(\mu, \nu)$ is finite if and only if

$$\iint \big|\log |x - y|\big| \mu(dx)\mu(dy) < \infty \quad \text{and} \quad \iint \big|\log |x - y|\big| \nu(dx)\nu(dy) < \infty.$$

If either of these conditions fail, we set $\mathcal{H}(\mu, \nu) = +\infty$.

For measures on the interval $[-2, 2]$, the reduced relative entropy can be well understood in terms of the operator structure associated to the logarithmic potential of measures on $[-2, 2]$. To this matter we recall here some of the main results discussed in [7] which were put forward in order to deal with the free Poincaré inequality.

We will work with the following reference measures on $[-2, 2]$

$$\alpha(dx) = \frac{\sqrt{4 - x^2}}{2\pi} \, dx, \quad \text{and} \quad \beta(dx) = \frac{dx}{\pi \sqrt{4 - x^2}}.$$

Most of the action takes place around the arcsine measure β and we will use $\langle, \rangle$ to denote the inner product in $L^2(\beta)$. In the sequel we will use the following notation

$$\phi_n(x) = T_n\left(\frac{x}{2}\right) \quad \text{and} \quad \psi_n(x) = U_n\left(\frac{x}{2}\right) \quad \text{for } n \geqslant 0, \tag{2.5}$$

where $T_n(x)$, the *Chebyshev polynomials of the first kind*, are defined by $T_n(\cos\theta) = \cos(n\theta)$ and U_n, the *Chebyshev polynomials of second kind*, are described by $U_n(\cos\theta) = \frac{\sin(n+1)\theta}{\sin\theta}$. Adjusting a little bit the polynomials T_n as $\tilde{T}_0 = T_0$ and $\tilde{T}_n(x) = \sqrt{2}T_n(x)$, then it is easy to see that $\{\tilde{T}_n(x/2)\}_{n \geqslant 0}$ is an orthonormal basis for $L^2(\beta)$. Similarly, $\{U_n(x/2)\}_{n \geqslant 0}$ forms an orthonormal basis for $L^2(\alpha)$. Another relation which plays an important role here is

$$\phi_n' = \frac{n}{2} \psi_{n-1}. \tag{2.6}$$

First, we introduce the operators $\mathcal{E}, \mathcal{N}, \mathcal{L}$ on C^2 functions on $[-2, 2]$ as follows. Given a C^2 function $\phi : [-2, 2] \to \mathbb{R}$, set

$$(\mathcal{E}\phi)(x) = -\int \log|x - y|\phi(y)\beta(dy),$$

$$(\mathcal{N}\phi)(x) = \int y\phi'(y)\beta(dy) + x\int \phi'(y)\beta(dy) - (4 - x^2)\int \frac{\phi'(x) - \phi'(y)}{x - y}\beta(dy),$$

$$(\mathcal{L}\phi)(x) = -(4 - x^2)\phi''(x) + x\phi'(x). \tag{2.7}$$

For convenience in what follows we will use the space

$$K = \left\{ f \in L^2(\beta) : \int f\, d\beta = 0 \right\}$$

which is the orthogonal to constants in $L^2(\beta)$. The reason we single out this space is that the operators $\mathcal{N}$ and $\mathcal{E}$ (properly extended) are the inverse of each other.

Now we summarize the main properties presented in [7, Proposition 1] and needed in this note.

Proposition 1.

(1) $\mathcal{E}$ sends $C^2([-2, 2])$ into $C^2([-2, 2])$ and can be extended to a bounded selfadjoint operator from $L^2(\beta)$ into itself.

(2) For any C^2 function $\phi \in K$,

$$\mathcal{E}\mathcal{N}\phi = \phi, \qquad \mathcal{N}\mathcal{E}\phi = \phi. \tag{2.8}$$

(3) In addition $\mathcal{E}\phi_0 = 0$, while for $n \geqslant 1$, $\mathcal{E}\phi_n = \frac{1}{n}\phi_n$ and $\mathcal{N}\phi_n = n\phi_n$ for any $n \geqslant 0$. In other words, $\mathcal{N}$ is the counting number operator for the Chebyshev basis in $L^2(\beta)$.

(4) $\mathcal{N}$ can be canonically extended to a selfadjoint operator on $L^2(\beta)$ which restricted to K has inverse $\mathcal{E}$.

(5) $\mathcal{L} = \mathcal{N}^2$ and it satisfies for any C^2 functions $\phi, \psi \in L^2(\beta)$,

$$\langle \mathcal{L}\phi, \psi \rangle = 2\int \phi'\psi'\, d\alpha. \tag{2.9}$$

In what follows, a key role is played by the following corollary.

Corollary 1. *For any C^2 functions $\phi, \psi \in K$,*

$$\langle \phi, \psi \rangle = 2\int (\mathcal{E}^2\phi)'\psi'\, d\alpha = 2\int (\mathcal{E}\phi)'(\mathcal{E}\psi)'\, d\alpha. \tag{2.10}$$

Proof. This follows from (2.9) with ϕ replaced by $\mathcal{E}^2\phi$ and the fact that $\mathcal{E}$ and $\mathcal{N}$ are the inverse of each other. Indeed, the first equality follows by replacing ϕ by $\mathcal{E}^2\phi$ in (2.9), while the second equality follows by replacing in (2.9) ϕ by $\mathcal{E}\phi$ and ψ by $\mathcal{E}\psi$ and then using that $\mathcal{E}$ is selfadjoint on K and $\mathcal{E}\mathcal{L}\mathcal{E} = \mathcal{E}^2\mathcal{L} = (\mathcal{E}\mathcal{N})^2 = I$ on K. $\quad\square$

Another simple consequence of Proposition 1 is the following representation.

Corollary 2. *If $\mu(dx) = \phi(x)\beta(dx)$ and $\nu(dx) = \psi(x)\beta(dx)$ with $\phi, \psi \in L^2(\beta)$ on $[-2, 2]$, then*

$$\mathcal{H}(\mu, \nu) = \langle \mathcal{E}(\phi - \psi), \phi - \psi \rangle. \tag{2.11}$$

The first main result of this note is the following.

Theorem 1. *For any two probability measures μ, ν on $[-2, 2]$,*

$$W_1^2(\mu, \nu) \leqslant 2\mathcal{H}(\mu, \nu). \tag{2.12}$$

Equality is attained for any two probability measures μ, ν such that $\mu(dx) - \nu(dx) = cx\beta(dx)$ for some constant c.

Proof. We should notice first that the maximization from (2.2) can be taken over the set of smooth functions g with the property that $\int g \, d\beta = 0$. Hence,

$$W_1(\mu, \nu) = \sup\left\{ \int g \, d(\mu - \nu), g \in K \text{ and } |g'| \leqslant 1 \right\}. \tag{2.13}$$

Now we prove first (2.12) for the case of measures μ, ν which have smooth densities with respect to the reference measure β. Therefore, write $\mu - \nu = \psi d\beta$ and continue with

$$W_1(\mu, \nu) = \sup\left\{ \int g\psi \, d\beta, |g'| \leqslant 1 \right\},$$

where now the supremum is taken over all smooth functions g.

Furthermore, for any smooth function g with bounded derivative by 1, using (2.10) we have the following string of equalities and inequalities

$$\int g\psi \, d\beta = 2 \int g' (\mathcal{E}^2 \psi)' \, d\alpha$$

$$\leqslant 2 \int |(\mathcal{E}^2 \psi)'| \, d\alpha \leqslant 2 \left(\int ((\mathcal{E}^2 \psi)')^2 \, d\alpha \right)^{1/2} = \sqrt{2} \langle \mathcal{E}\psi, \mathcal{E}\psi \rangle^{1/2}$$

$$= \sqrt{2} \langle \mathcal{E}^2 \psi, \psi \rangle^{1/2}. \tag{2.14}$$

The inequality we want to prove now certainly follows from (2.11) and the fact that

$$\langle \mathcal{E}^2 \psi, \psi \rangle \leqslant \langle \mathcal{E}\psi, \psi \rangle, \tag{2.15}$$

which is a consequence of the fact that the spectrum of $\mathcal{E} : K \to K$ is $\{1, 1/2, 1/3, \ldots\}$ with eigenfunctions $\{\phi_1, \phi_2, \phi_3, \ldots\}$. More precisely, if we write $\psi = \sum_{n \geqslant 1} \gamma_n \phi_n$, then (2.15) is equivalent to

$$\sum_{n \geqslant 1} \frac{1}{n^2} \gamma_n^2 \leqslant \sum_{n \geqslant 1} \frac{1}{n} \gamma_n^2,$$

which is obvious. In fact this inequality is saturated when γ_n are all 0 with the exception of $n = 1$. Thus tracing back all the inequalities in between we get equality in (2.12) for any measures μ, ν for which $\mu - \nu = cx \, d\beta$ with some constant c.

In the second place, an approximation procedure shows that we can reduce the proof to the case of measures having smooth densities with respect to β.

To carry out this reduction, notice that if $\mathcal{H}(\mu, \nu)$ is infinite then there is nothing to prove here, so we will assume that $\mathcal{H}(\mu, \nu)$ is finite in which case, according to [5, Eq. 6.47],

$$\mathcal{H}(\mu, \nu) = \int_0^\infty \frac{|\hat{\mu}(t) - \hat{\nu}(t)|^2}{t} \, dt \tag{2.16}$$

with the standard notation of $\hat{\mu}$ being the Fourier transform (or the characteristic function) of the measures μ.

In order to use this equation we consider a smooth compactly supported function $\zeta : [0, 1] \to [0, \infty]$ such that $\int_0^1 \zeta(x) \, dx = 1$. Now, set $\zeta_\epsilon(x) = \frac{1}{\epsilon} \zeta(x/\epsilon)$ for small ϵ and consider the measure $\xi_\epsilon(dx) = \zeta_\epsilon(x) \, dx$. Based on the measure ξ_ϵ we construct $\mu_\epsilon = \xi_\epsilon \star \mu$ and similarly $\nu_\epsilon = \xi_\epsilon \star \nu$ where $\star$ is the standard classical convolution operation. It is now clear in the first place that μ_ϵ and ν_ϵ are probability measures with smooth compact support such that $\mu_\epsilon \xrightarrow[\epsilon \to 0]{} \mu$ and $\nu_\epsilon \xrightarrow[\epsilon \to 0]{} \nu$ in the weak topology. In particular this means that

$$W_1(\mu_\epsilon, \nu_\epsilon) \xrightarrow[\epsilon \to 0]{} W_1(\mu, \nu). \tag{2.17}$$

At the same time

$$\mathcal{H}(\mu_\epsilon, \nu_\epsilon) \xrightarrow[\epsilon \to 0]{} \mathcal{H}(\mu, \nu) \tag{2.18}$$

which follows from (2.16) and

$$\mathcal{H}(\mu_\epsilon, \nu_\epsilon) = \int_0^\infty \frac{|\hat{\mu}(t) - \hat{\nu}(t)|^2 |\hat{\zeta}_\epsilon|^2}{t} \, dt \xrightarrow[\epsilon \to 0]{} \int_0^\infty \frac{|\hat{\mu}(t) - \hat{\nu}(t)|^2}{t} \, dt$$

where in the last part we used $|\hat{\zeta}_\epsilon| \leqslant 1$ and $\lim_{\epsilon \to 0} \hat{\zeta}_\epsilon = 1$ combined with the dominated converge.

Hence, we have smooth compactly supported approximations of μ, ν, the only problem is that the approximations $\mu_\epsilon, \nu_\epsilon$ have support in $[-2 - \epsilon, 2 + \epsilon]$. Thus what we can do is to take $\ell_\epsilon(x) = x/(1 + \epsilon)$ which maps $[-2 - \epsilon, 2 + \epsilon]$ into $[-2 + \epsilon/(1 + \epsilon), 2 - \epsilon/(1 + \epsilon)]$ and therefore the push forward $\tilde{\mu}_\epsilon = (\ell_\epsilon)_\# \mu_\epsilon$ and $\tilde{\nu}_\epsilon = (\ell_\epsilon)_\#$ are probability measures with smooth densities supported on $[-2 + \epsilon/(1 + \epsilon), 2 - \epsilon/(1 + \epsilon)]$. On the other hand it is easy to see that

$$W_1(\mu_\epsilon, \nu_\epsilon) = (1 + \epsilon) W_1(\tilde{\mu}_\epsilon, \tilde{\nu}_\epsilon)$$

and

$$\mathcal{H}(\tilde{\mu}_\epsilon, \tilde{\nu}_\epsilon) = \mathcal{H}(\mu_\epsilon, \nu_\epsilon) + \log(1 + \epsilon).$$

The conclusion is that now we can apply the smooth case to the measures $\tilde{\mu}_\epsilon$ and $\tilde{\nu}_\epsilon$ to deduce that

$$W_1^2(\mu_\epsilon, \nu_\epsilon) \leqslant 2(1 + \epsilon)^2 \big(\mathcal{H}(\mu_\epsilon, \nu_\epsilon) + \log(1 + \epsilon)\big)$$

which combined with (2.17) and (2.18) leads to (2.12) for any probability measures μ, ν on $[-2, 2]$. □

A corollary of the proof of Theorem 1 is the following nice representation of W_1 in terms of the operator $\mathcal{E}$ which we want to record and use later as a separate result.

Corollary 3. *If μ, ν are two probability measures on $[-2, 2]$ such that $d\mu - d\nu = \psi \, d\beta$, where ψ is in $L^2(\beta)$, then*

$$W_1^2(\mu, \nu) = 2\langle \mathcal{E}^2 \psi, \psi \rangle. \tag{2.19}$$

Proof. In the case of smooth ψ, the proof is nothing but the content of the key sequence of inequalities from (2.14).

For the general case, we need to approximate $\psi \in L^2(\beta)$ by smooth functions. This requires a little care but it is straightforward and we point only the main steps.

First, notice that by simple approximations, it is sufficient to prove the statement for measures μ, ν which are supported inside $(-2, 2)$.

Second, take the mollifier ζ_ϵ as in the proof of Theorem 1 and consider the standard mollification of ψ as $\psi_\epsilon = \zeta_\epsilon \star \psi$. It is clear that ψ_ϵ converges a.s. to ψ and in $L^2(\beta)$. To see the last part, the convergence in $L^2(\beta)$, one needs to observe that due to Cauchy's inequality, $|\psi_\epsilon(x)|^2 \leqslant 2\pi \|\psi\|_{L^2(\beta)}^2$, from which ψ_ϵ is certainly in $L^2(\beta)$ and its norm is controlled by the norm of ψ. This is sufficient to conclude that ψ_ϵ converges in $L^2(\beta)$ toward ψ. □

By simple scaling we have the following consequence of Theorem 1.

Corollary 4. *If μ, ν are two probability measures supported on $[-L, L]$, then*

$$W_1^2(\mu, \nu) \leqslant 2L^2 \mathcal{H}(\mu, \nu). \tag{2.20}$$

The constant $2L^2$ in front of $\mathcal{H}$ is sharp.

Remark 1. In the classical case, the Csiszár–Kullback–Pinsker inequality states that

$$\|\mu - \nu\|_\nu^2 \leqslant 2\mathcal{H}(\mu, \nu)$$

for any two measures on the real line where $\mathcal{H}(\mu|\nu)$ is the classical relative entropy given by $\mathcal{H}(\mu|\nu) = \int \frac{d\mu}{d\nu} \log \frac{d\mu}{d\nu} \, d\nu$ in the case μ is absolutely continuous with respect to ν and is $+\infty$ otherwise. Also the distance $\|\mu - \nu\|_\nu$ is the total variation distance.

Édouard Maurel-Segala and Mylène Maïda asked if there is such an inequality in the free case. The answer turns out to be negative and is relatively straightforward in the setup put up here.

To see this choose some probability measures μ, ν on $[-2, 2]$ such that $\mu - \nu = c\phi_n d\beta$ with $n \geqslant 1$. With this choice, using (2.11),

$$\|\mu - \nu\|_v = |c| \int |\phi_n| \, d\beta \quad \text{while } \mathcal{H}(\mu, \nu) = c^2 \langle \mathcal{E}\phi_n, \phi_n \rangle = \frac{c^2}{2n}.$$

Since it is not hard to check that for any $n \geqslant 1$

$$\int |\phi_n| \, d\beta = \frac{1}{\pi} \int_0^\pi \left| \cos(nt) \right| dt = \frac{2}{\pi},$$

the conclusion is that there is no constant $C > 0$ such that

$$C\|\mu - \nu\|_v^2 \leqslant \mathcal{H}(\mu|\nu)$$

for all probability measures μ, ν on $[-2, 2]$.

In the classical case Csiszár–Kullback–Pinsker on compact intervals implies the transportation with W_1 metric and this is also the reason why the latter is eclipsed by the former. In the free context, since the Csiszár–Kullback–Pinsker fails, it makes the W_1 transportation more interesting.

3. No local transportation with respect to the W_p-metric, $p > 1$

Given the above local transportation from Theorem 1, one natural question in this framework is whether one can extend it to the case of W_p metric instead of W_1 with $p > 1$.

In other words, is there a constant $C > 0$ such that for any measures μ, ν supported in $[-2, 2]$, the following holds true

$$CW_p^2(\mu, \nu) \leqslant \mathcal{H}(\mu, \nu)? \tag{3.1}$$

As we will see, the answer is no. To see why this is the case, we start with the following [16, p. 75]

$$W_p^p(\mu, \nu) = \int_0^1 \left| F_\mu^{-1}(t) - F_\nu^{-1}(t) \right|^p dt,$$

where F_μ, F_ν are the cumulative functions of μ, ν and F_μ^{-1}, F_ν^{-1} are their generalized inverses. Now, take $d\mu = \phi \, d\beta$, $\phi > 0$ on $[-2, 2]$ with $\phi \in C^\infty([-2, 2])$. Then choose $\nu_\epsilon = (\phi + \epsilon h) \, d\beta$, where h is a smooth compactly supported function on $(-2, 2)$ with $\int h \, d\beta = 0$ and ϵ is small enough. With this choice and (2.11) it is obvious that

$$\mathcal{H}(\mu, \nu_\epsilon) = \epsilon^2 \langle \mathcal{E}h, h \rangle. \tag{3.2}$$

On the other hand, it is not hard to show that the cumulative functions of μ, ν are smooth and they actually depend smoothly also on ϵ. Denoting now for simplicity $F(x) = F_\mu(x)$, $F_\epsilon(x) = F_{\nu_\epsilon}(x)$, and $G(x) = \int_{-2}^{x} h(y)\beta(dy)$, then

$$F_\epsilon(x) = F(x) + \epsilon G(x).$$

From this, it is not hard to see that

$$F_\epsilon^{-1}(t) = F^{-1}(t) - \epsilon \frac{G(F^{-1}(t))}{f(F^{-1}(t))} + O(\epsilon^2)$$

where $f(x) = F'(x)$. Consequently, it follows that

$$W_p^2(\mu, \nu_\epsilon) = \epsilon^2 \int_0^1 \frac{|G(F^{-1}(t))|^p}{|f(F^{-1}(t))|^p} \, dt + O(\epsilon^4) = \int \frac{G^2(x)}{f(x)} \, dx$$

$$= \frac{\epsilon^2}{\pi} \int_{-2}^2 \frac{|\int_{-2}^x h(y)\beta(dy)|^p}{\phi^{p-1}(x)} \left(4 - x^2\right)^{(p-1)/2} dx + O(\epsilon^4).$$

If we assume (3.1) is true for some constant $C > 0$, then letting ϵ go to 0, the above considerations lead to the following conclusion. For any $\phi > 0$ such that $\int \phi \, d\beta = 1$ and any compactly supported smooth function h on $(-2, 2)$, we get that

$$\frac{C}{\pi} \int_{-2}^2 \frac{|\int_{-2}^x h(y)\beta(dy)|^p}{\phi^{p-1}(x)} \left(4 - x^2\right)^{(p-1)/2} dx \leqslant \langle \mathcal{E}h, h \rangle.$$

Then this yields that this must be true for any smooth function h on $[-2, 2]$ and any smooth positive ϕ on $[-2, 2]$ such that $\int \phi \, d\beta = 1$. However, if we fix a function h, say $h(x) = x$, then the above inequality implies that there is a positive constant $C' > 0$ such that

$$\int_{-2}^2 \frac{(4 - x^2)^{(2p-1)/2}}{\phi^{p-1}(x)} \, dx \leqslant C'$$

for any smooth positive function ϕ such that $\int \phi \, d\beta = 1$. But we can choose a function ϕ which is arbitrary close to 0 on the interval $[-1/3, 1/3]$ and the above inequality leads to a contradiction. This argument shows that (3.1) cannot be true, so there is no local transportation with W_p-metric, $p > 1$.

From a certain perspective, this argument is just a linearization of the transportation inequality and this in the classical case corresponds to a Poincaré inequality for the reference measure μ. In a similar fashion, the argument outlined above says that not every measure μ on $[-2, 2]$ satisfies a free Poincaré inequality. The right version of such a Poincaré inequality in this framework is actually the one discussed in details in [7], but we do not enter into further discussion here.

4. A global version of transportation inequality on the whole real line

Now we want to prove a transportation inequality on the whole real line which in this case is potential dependent. To setup the scene, we will take a potential $V : \mathbb{R} \to \mathbb{R}$ such that

Assumption 1. *V is bounded below, measurable and satisfies*

$$\liminf_{|x| \to \infty} \frac{V(x)}{x^2} > 0.$$

For any probability measure $\mu \in \mathcal{P}(\mathbb{R})$,

$$E_V(\mu) = \int V \, d\mu - \iint \log |x - y| \mu(dx) \mu(dy).$$

It is known, see for example [15, Theorem 1.3] that there is a unique measure (also with compact support) μ_V which minimizes $E_V(\mu)$ over all probabilities $\mu \in \mathcal{P}(\mathbb{R})$.

The equilibrium measure μ_V (cf. [15, Theorem I.1.3]) satisfies

$$V(x) \geqslant 2 \int \log |x - y| \mu_V(dy) + K_V \quad \text{quasi-everywhere on } \mathbb{R}$$

$$V(x) = 2 \int \log |x - y| \mu_V(dy) + K_V \quad \text{quasi-everywhere on } \operatorname{supp} \mu_V, \qquad (4.1)$$

where K_V is known as Robin constant.

For the definition of the notion of quasi-everywhere see for instance [15, p. 25]. In what follows, to simplify a little bit the exposition, we will denote

$$U(x) = 2 \int \log |x - y| \mu_V(dy) + K_V.$$

For simplicity in what follows, we denote $E_V(\mu_V) = E_V$ and we set *the relative free entropy*

$$E_V(\mu | \mu_V) = E_V(\mu) - E_V(\mu_V).$$

For any measure μ, using the above equalities we can write now,

$$E_V(\mu | \mu_V) = \int \big(V(x) - U(x) \big) \mu(dx) + \mathcal{H}(\mu, \mu_V). \qquad (4.2)$$

Because of Assumption 1 and (4.1), there are constants $A, B > 0$ such that

$$V(x) - U(x) \geqslant A \mathbb{1}_{[-B, B]^c}(x) x^2. \qquad (4.3)$$

For a given measure $\mu \in \mathcal{P}(\mathbb{R})$ with compact support, say in $[-L, L]$, one thing we can try to do is to use (2.20) and (4.1) to estimate from below $\mathcal{H}(\mu, \mu_V)$. Hence we get at first that
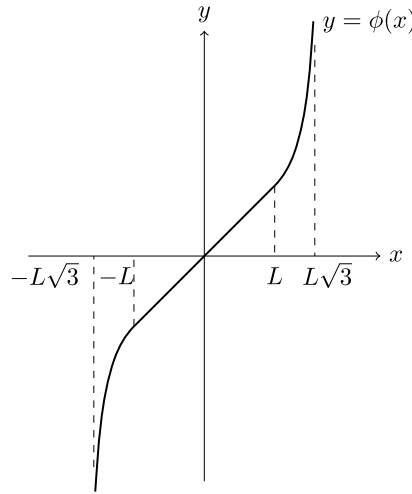
Fig. 1. The graph of ϕ.

$$E_V(\mu|\mu_V) \geqslant \int (V - U)\,d\mu + \frac{1}{2L^2} W_1^2(\mu, \mu_V).$$

The problem here is that as L goes to infinity the W_1 term simply disappears. On the other hand, for large values of L, the potential V is bounded from below by a quadratic, while U grows at most logarithmically, thus for large values of L, $V - U$ grows at least quadratically. With this in mind, the idea is to refine the above scheme so that it takes better advantage of the quadratic growth.

To carry this idea through, we do the following. Take a large $L \geqslant B$ such that μ_V is supported on $[-L/2, L/2]$. From (4.3), we learn that $V(x) \geqslant Ax^2$ for $|x| \geqslant L$. Consider the function $\phi : [-\sqrt{3}L, \sqrt{3}L] \to \mathbb{R}$ (see Fig. 1) given by

$$\phi(x) = \begin{cases} -\frac{2L^3}{3L^2 - x^2}, & -\sqrt{3}L < x \leqslant -L, \\ x, & -L < x \leqslant L, \\ \frac{2L^3}{3L^2 - x^2}, & L < x \leqslant \sqrt{3}L. \end{cases}$$

There are several elementary properties of this function we will use in the sequel. For $L \geqslant 3^{1/4}$,

   1) ϕ is odd, increasing and C^1 on $(-\sqrt{3}L, \sqrt{3}L)$

   2) $\phi'(x)$ is decreasing on $(-\sqrt{3}L, 0]$ and is increasing on $[0, \sqrt{3}L)$

   3) $\lim\limits_{x \to -\sqrt{3}L} \phi(x) = -\infty$, $\lim\limits_{x \to \sqrt{3}L} \phi(x) = \infty$

   4) $\phi(x) = x$ for $-L \leqslant x \leqslant L$

   5) $\log\left|\dfrac{\phi(x) - \phi(y)}{x - y}\right| \leqslant 2\log|\phi(x)| + 2\log|\phi(y)|$ if $|x| \vee |y| \geqslant L$.   (4.4)

Next, define $\nu = (\phi^{-1})_{\#}\mu$ and notice that ν is supported on $[-\sqrt{3}L, \sqrt{3}L]$. Now we start using (4.1) to justify that

$$E_V(\mu|\mu_V) = \int \left(V\big(\phi(x)\big) - U(x)\right)v(dx) - \iint \log\left|\frac{\phi(x) - \phi(y)}{x - y}\right| v(dx)v(dy) + \mathcal{H}(v, \mu_V).$$

$$(4.5)$$

The point of this is that v is supported on $[-\sqrt{3}L, \sqrt{3}L]$ and from (2.20),

$$\mathcal{H}(v, \mu_V) \geqslant \frac{1}{6L^2} W_1^2(v, \mu_V).$$

$$(4.6)$$

Now let's turn our attention to the first two terms of (4.5) and notice that because $\phi(x) = x$ for $x \in [-L, L]$ and $V(x) \geqslant U(x)$, combined with the last line of (4.4), yields

$$\int \left(V\big(\phi(x)\big) - U(x)\right)v(dx) - \iint \log\left|\frac{\phi(x) - \phi(y)}{x - y}\right| v(dx)v(dy)$$

$$\geqslant \int_{|x|\geqslant L} \left(A\big|\phi(x)\big|^2 - 2\log(2\sqrt{3}L) - C\right)v(dx) - 4\int_{|x|\geqslant L} \log\big|\phi(x)\big|v(dx)$$

$$\geqslant \frac{A}{2} \int_{|x|\geqslant L} \phi^2(x)v(dx)$$

for large L and a certain constant $C > 0$. Putting together these findings, we conclude that for large L,

$$E_V(\mu|v) \geqslant \frac{A}{2} \int_{|x|\geqslant L} \phi^2(x)v(dx) + \frac{1}{6L^2} W_1^2(v, \mu_V).$$

At this point we use the characterization of the Wasserstein distance given by (2.3). Namely, if θ is the transportation map of μ_V into v, then $\phi \circ \theta$ is the transportation map of μ_V into μ. Thus, the right-hand side of the above equation can be continued as

$$\frac{A}{2} \int_{|x|\geqslant L} \phi^2(x)v(dx) + \frac{1}{6L^2}\left(\int \big|\theta(x) - x\big|\mu_V(dx)\right)^2$$

$$= \frac{A}{2} \int_{|\theta(x)|\geqslant L} \phi^2\big(\theta(x)\big)\mu_V(dx) + \frac{1}{6L^2}\left(\int_{|\theta(x)|\leqslant L} \big|\theta(x) - x\big|\mu_V(dx)\right)^2$$

$$= \frac{A}{2} \int_{|\theta(x)|\geqslant L} \phi^2\big(\theta(x)\big)\mu_V(dx) + \frac{1}{6L^2}\left(\int_{|\theta(x)|\leqslant L} \big|\phi\big(\theta(x)\big) - x\big|\mu_V(dx)\right)^2$$

$$\geqslant \frac{2A}{9} \int_{|\theta(x)|\geqslant L} \big(\phi\big(\theta(x)\big) - x\big)^2\mu_V(dx) + \frac{1}{6L^2}\left(\int_{|\theta(x)|\leqslant L} \big|\phi\big(\theta(x)\big) - x\big|\mu_V(dx)\right)^2$$

$$\geqslant \frac{2A}{9}\left(\int_{|\theta(x)|\geqslant L} \big|\phi\big(\theta(x)\big) - x\big|\mu_V(dx)\right)^2 + \frac{1}{6L^2}\left(\int_{|\theta(x)|\leqslant L} \big|\phi\big(\theta(x)\big) - x\big|\mu_V(dx)\right)^2$$

$$\geqslant \frac{2A}{12AL^2 + 9} W_1^2(\mu, \mu_V).$$

To clarify this long equation we make the following points. The second equality in the above follows from the fourth property of (4.4). The third inequality follows from the fact that $3|\phi(y)| \geqslant 2|\phi(y) - x|$ for any $|x| \leqslant L/2$ and $|y| \geqslant L$ which (because ϕ is increasing and odd) it is easy to see it is equivalent to $3\phi(L) \geqslant 2(\phi(L) + L/2)$ which is obviously clear from $\phi(L) = L$. The second inequality is just Cauchy's inequality and the last one is easy to see from $am^2 + bn^2 \geqslant (m + n)^2/(1/a + 1/b)$ for $a, b, m, n \geqslant 0$.

What we just proved is the following global version of the transportation inequality.

Theorem 2. *Under Assumption* 1, *there is a constant* $C > 0$ *depending on* V *such that for any probability measure* μ *on* $\mathbb{R}$,

$$C W_1^2(\mu, \mu_V) \leqslant E_V(\mu|\mu_V). \tag{4.7}$$

Remark 2. What is worth pointing here is that the constant C depends on the choice of L, which in turn is determined by the constants A, B from (4.3), the constant K_V from (4.1) and the support of μ_V.

5. Local versions of free Log-Sobolev inequality

The main feature of the transportation inequality in Theorem 1 is that its formulation does not depend on any potential on the interval $[-2, 2]$. Perhaps better said, the quantities $\mathcal{H}(\mu, \nu)$ and $W_1(\mu, \nu)$ are defined independently of the potential V defining the relative entropy in (4.1). However the potential independent result of Theorem 1 combined with a growth at infinity, provides the necessary ingredients for a transportation inequality with a potential V as is presented in Theorem 2.

On the other hand in [2] the free Log-Sobolev is introduced and deduced from an inequality of [19]. The main statement is as follows. We say that the free Log-Sobolev associated to a potential $V : \mathbb{R} \to \mathbb{R}$ holds true if there is a constant $\rho > 0$ such that for any probability measure $\mu \in \mathcal{P}(\mathbb{R})$,

$$\rho E_V(\mu|\mu_V) \leqslant I(\mu|\mu_V) \tag{5.1}$$

where

$$I(\mu|\mu_V) = \begin{cases} \int (H\mu - V')^2 \, d\mu & \text{if } \frac{d\mu}{dx} \in L^3(\mathbb{R}), \\ +\infty & \text{otherwise} \end{cases} \tag{5.2}$$

and

$$H\mu(x) = p.v. \int \frac{2}{x - y} \mu(dy) \tag{5.3}$$

with the integral taken in the principal value sense.

This inequality was then reproved by Biane in [1] for the case of strongly convex potentials V using random matrix approximations and also in [8] using tools form mass transportation.

One immediate consequence of the above inequality and the uniqueness of the equilibrium measure is that if $V' = H\mu$ holds almost surely on the support of μ, then μ must be equal to μ_V.

It was pointed out in [1] by Biane that in the case of nonconvex potentials, the free logarithmic Sobolev inequality cannot hold true due essentially to the fact that one can construct different measures whose Hilbert transforms agree to V' on their supports and this violates (5.1) and the uniqueness of the minimizer of $E_V(\mu)$. The example given there is the one in which V is a double well potential with $V(x) = (x - a_1)^2/2$ on $J_1 = [a_1 - 2, a_1 + 2]$ and $V(x) = (x - a_2)^2/2$ on $J_2 = [a_2 - 2, a_2 + 2]$ with $|a_1 - a_2| > 4$. With this choice it is clear that V' is equal to the Hilbert transform on J_1 of the semicircular measure on J_1. Similarly, V' is equal to the Hilbert transform on J_2 of the semicircular measure on J_2.

Therefore Log-Sobolev inequality cannot be true without extra assumptions on the potential V. In order to get a version which is potential independent, one reasonable thought is to try replacing the derivative of V' by something which is in terms of the measure μ_V. A natural candidate to that is the Hilbert transform of μ_V. Furthermore the roles played by μ and μ_V should be symmetric and thus the integral with respect to μ in (5.2) should be replaced by an integral with respect to a measure which is symmetric with respect to both, μ and μ_V. One way of fixing this is to integrate with respect to a fixed measure rather than a measure depending on μ and μ_V.

To summarize the preceding paragraph, we introduce a potential independent Log-Sobolev in the form of

$$\rho \mathcal{H}(\mu, v) \leqslant \int (H\mu - Hv)^2 \, d\zeta$$

where here we take a reference measure ζ supported on a certain set K and the measures μ, v to be supported by K.

We do not investigate here the general situation alluded above but focus on the following version of the free information for measures supported on $[-2, 2]$ with the reference measure ζ being the semicircular measure α:

$$\mathcal{J}(\mu, v) = \begin{cases} \int (H\mu - Hv)^2 \, d\alpha & \text{if } H\mu, Hv \in L^2(\alpha), \\ +\infty & \text{otherwise.} \end{cases} \tag{5.4}$$

If α_L is the semicircular law on $[-2L, 2L]$, then the extension of the above is

$$\mathcal{J}_L(\mu, v) = \begin{cases} \int (H\mu - Hv)^2 d\alpha_L & \text{if } H\mu, Hv \in L^2(\alpha_L), \\ +\infty & \text{otherwise.} \end{cases} \tag{5.5}$$

Assume for the moment that $d\mu = \phi \, d\beta$ with ϕ a C^2 function on $[-2, 2]$. In this case, since $\mathcal{E}\phi$ is a C^2 function, the principal value integral defining $H\mu$ for $x \in (-2, 2)$ can be shown to be equal to

$$\frac{1}{2}(H\mu)(x) = \frac{d}{dx}(\mathcal{E}\phi)(x) = (\mathcal{U}\phi)(x) := \int \frac{\phi(x) - \phi(y)}{x - y} \beta(dy). \tag{5.6}$$

Given a function ϕ, we will use the notation $H_\beta \phi$ for $H(\phi\beta)$.

The operator $\mathcal{U}$ appears in [7] and the main property it satisfies is that

$$\|\mathcal{U}f\|_\alpha^2 = \frac{1}{2}\,\mathrm{Var}_\beta(f) \tag{5.7}$$

for any $f \in L^2(\beta)$ where $\mathrm{Var}_\beta(f)$ is the variance of f with respect to β. In particular, what this says is that the operator $\mathcal{U}$ is an isometry between the spaces orthogonal to constants in $L^2(\beta)$ and $L^2(\alpha)$.

To see this property, it suffices to check that

$$\mathcal{U}\phi_n = \frac{1}{2}\psi_{n-1} \tag{5.8}$$

which actually follows from $\frac{d}{dx}(\mathcal{E}\phi)(x) = (\mathcal{U}\phi)(x)$ and (2.6).

The upshot of the above consideration is that if $d\mu = \phi\,d\beta$ and $d\nu = \psi\,d\beta$ are two probability measures with ϕ, ψ being C^2 on $[-2, 2]$, then

$$\int (H\mu - H\nu)^2\,d\alpha = 4\int \big(\mathcal{U}(\phi - \psi)\big)^2\,d\alpha = 2\int (\phi - \psi)^2\,d\beta.$$

It is this formula which inspires the following definition

$$\mathcal{I}(\mu, \nu) = \begin{cases} \int(\frac{d\mu}{d\beta} - \frac{d\nu}{d\beta})^2\,d\beta & \text{if } \frac{d\mu}{d\beta}, \frac{d\nu}{d\beta} \in L^2(\beta), \\ +\infty & \text{otherwise.} \end{cases} \tag{5.9}$$

For a given $L > 0$, we define on measures supported on $[-2L, 2L]$,

$$\mathcal{I}_L(\mu, \nu) = \begin{cases} \int(\frac{d\mu}{d\beta_L} - \frac{d\nu}{d\beta_L})^2\,d\beta_L & \text{if } \frac{d\mu}{d\beta_L}, \frac{d\nu}{d\beta_L} \in L^2(\beta_L), \\ +\infty & \text{otherwise.} \end{cases} \tag{5.10}$$

Theorem 3.

(1) *For a measure μ supported on $[-2, 2]$, $\frac{d\mu}{d\beta} \in L^2(\beta)$ if and only if $H\mu \in L^2(\alpha)$.*
(2) *For any probability measures, μ, ν on $[-2, 2]$,*

$$\mathcal{J}(\mu, \nu) = 2\mathcal{I}(\mu, \nu). \tag{5.11}$$

(3) *For any probability measures μ, ν on $[-2, 2]$,*

$$2\mathcal{H}(\mu, \nu) \leqslant \mathcal{J}(\mu, \nu) = 2\mathcal{I}(\mu, \nu), \tag{5.12}$$

with equality if $\mu(dx) - \nu(dx) = Cx\beta(dx)$.
(4) *In particular, by simple rescaling, for any probability measures μ, ν on $[-L, L]$,*

$$2\mathcal{H}(\mu, \nu) \leqslant \mathcal{J}(\mu, \nu) = 2L^2\mathcal{I}_L(\mu, \nu).$$

Proof.

(1) To prove the assertion, we recall two facts from the theory of Hilbert transform. The definition of the Hilbert transform is given by (notice that this is twice the one appearing in the literature)

$$H\mu(x) = p.v. \int \frac{2}{x-y} \mu(dx) = \lim_{\epsilon \to 0} \int \frac{2(x-y)}{(x-y)^2 + \epsilon^2} \mu(dy).$$

It is a standard result in the theory of Hilbert transform (see for instance [4] and reproved in [9]) that this is well defined λ-a.s., for all points x, with λ being the Lebesgue measure on $\mathbb{R}$. The first result we will use states that there is a constant $C > 0$ such that for any finite positive measure μ

$$t\lambda\big(\big\{x\colon \big|H\mu(x)\big| \geqslant t\big\}\big) \leqslant C\|\mu\|_v \quad \text{for all } t > 0. \tag{5.13}$$

Here $\|\mu\|_v$ is the variation of μ (i.e. in this case is just $\mu(\mathbb{R})$). It is a straightforward fact that this can be extended to any finite measure μ not necessarily positive, with the change that $\|\mu\|_v = \mu^+(\mathbb{R}) + \mu^-(\mathbb{R})$ where $\mu = \mu^+ - \mu^-$ is the standard decomposition of μ into the non-negative and non-positive parts. This fact can be found for instance in [9] but it is eventually attributed to Kolmogorov.

The second result we will use is that if μ is a finite measure and we take $d\mu = f d\lambda + \mu_s$, where μ_s is the singular part of the measure μ, then

$$\lim_{t \to \infty} \frac{\pi t}{2} \mathbb{1}_{\{x\colon |H\mu(x)| \geqslant t\}}\, d\lambda = d\mu_s \tag{5.14}$$

where the convergence is in the sense of weak convergence. This can be found for instance in [12, Eq. 5.4] and [13, Theorem 1].

Now, we proceed to the proof of the main statement. Assume first that $\phi = \frac{d\mu}{d\beta} \in L^2(\beta)$. Then we want to prove that $H\mu \in L^2(\alpha)$. To do this, observe that since $\phi \in L^2(\beta)$, we can use (5.7) to show that there is an extension of $H\mu$ to $L^2(\alpha)$. More precisely, approximate ϕ with a sequence of smooth positive functions ζ_n in $[-2, 2]$ with $\int \zeta_n d\beta = 1$ and this in turn, using (5.7), shows that $H\zeta_n$ converges in $L^2(\alpha)$. So what it remains to show is that $H\mu = \lim_{n \to \infty} H\zeta_n$. For this last part, observe that because ζ_n converges to ϕ in $L^2(\beta)$, it is relatively easy to show that

$$\int \frac{|\zeta_n(x) - \phi(x)|}{\pi\sqrt{4-x^2}} \lambda(dx) \xrightarrow[n \to \infty]{} 0.$$

Combining this with (5.13), we obtain that (here $\nu_n = \zeta_n\, d\beta$)

$$\lim_{n \to \infty} \lambda\big(\big\{x\colon \big|H\mu(x) - H\nu_n(x)\big| \geqslant t\big\}\big) = 0$$

which yields the convergence in measure of $H\nu_n$ toward $H\mu$ as $n \to \infty$. Since, $H\nu_n$ converges also in $L^2(\alpha)$, it means that in fact $H\nu_n$ converges in $L^2(\alpha)$ and $H\mu \in L^2(\alpha)$ which is what we wanted.

For the reverse implication, assume now we have $H\mu \in L^2(\alpha)$. The first step is to show that μ is absolutely continuous with respect to β or alternatively with respect to λ. To this end, we will exploit (5.14). Indeed for any continuous function f supported on $[-2, 2]$,

$$\lim_{t \to \infty} \frac{\pi t}{2} \int \mathbb{1}_{\{x: \, |H\mu(x)| \geqslant t\}} f(x) \, dx = \int f \, d\mu_s.$$

On the other hand, since f is supported on $[-2, 2]$, we can apply Hölder's inequality followed by Chebyshev's to obtain

$$\int \mathbb{1}_{\{x: \, |H\mu(x)| \geqslant t\}} f(x) \, dx \leqslant \lambda\big(\big\{x \in [-2, 2]: \, \big|H\mu(x)\big| \geqslant t\big\}\big)^{1/p} \|f\|_q$$

$$\leqslant \frac{\|f\|_q}{t^{r/p}} \left(\int_{-2}^{2} |H\mu|^r \, d\lambda \right)^{1/p}$$

for any $p, q \geqslant 1$ with $1/p + 1/q = 1$ and $r > p$. In fact, we choose p, q, r such that $1 < p < r < 4/3$ and using again Hölder's inequality, we continue writing

$$\int_{-2}^{2} |H\mu|^r \, d\lambda \leqslant \left(\int_{-2}^{2} \big|H\mu(x)\big|^2 \alpha(dx) \right)^{r/2} \left(\int_{-2}^{2} \left(\frac{4\pi^2}{4 - x^2} \right)^{\frac{r}{2(2-r)}} \right)^{(2-r)/2} \leqslant C_r \|H\mu\|^r_{L^2(\alpha)}.$$

The conclusion we draw from these estimates is that

$$\frac{\pi t}{2} \left| \int \mathbb{1}_{\{x: \, |H\mu(x)| \geqslant t\}} f(x) \, dx \right| \leqslant \frac{C}{t^{r/p-1}} \|H\mu\|^{r/p}_{L^2(\alpha)} \|f\|_q$$

and thus, letting $t \to \infty$,

$$\int f \, d\mu_s = 0,$$

which is nothing but the fact that μ is absolutely continuous with respect to λ and thus with respect to β.

To go forward, we need to show that $\phi = \frac{d\mu}{d\beta}$ is in $L^2(\beta)$, or alternatively that $\phi - 1 \in L^2(\beta)$. To achieve this, consider $L^2_0(\beta)$ as the set of L^2 functions of mean zero with respect to β and let $L^2_0(\alpha)$ be the set of L^2 functions with mean zero with respect to α. We now show that H_β extends from the smooth functions in $L^2_0(\beta)$ into $L^2_0(\alpha)$ such that

$$\|H_\beta \psi\|_{L^2(\alpha)} = \frac{1}{\sqrt{2}} \|\psi\|_{L^2(\beta)} \tag{5.15}$$

for all $\psi \in L^2_0(\beta)$. From (5.7), it is clear that there is an operator $L : L^2_0(\beta) \to L^2_0(\alpha)$ such that L coincides with H_β on smooth functions. The point is to show that $L\phi = H_\beta \phi$ for any function $\phi \in L^2_0(\beta)$. This can be done as follows. Take a $\phi \in L^2_0(\beta)$ and approximate it in $L^2(\beta)$ with some smooth functions $\xi_n \in L^2_0(\beta)$. Then, from the equality (5.7) on smooth functions, it follows

that $H_\beta \xi_n$ is a Cauchy sequence in $L_0^2(\alpha)$. On the other hand, using that ξ_n converges in $L^1(\beta)$ combined with (5.13), we conclude that $H_\beta \xi_n$ forms a Cauchy sequence in measure and thus, its limit in measure must be the same as its limit in $L^2(\alpha)$, from which we deduce that $L\phi = H_\beta \phi$. Once the isometry property above is established, the fact that $\phi \in L^2(\beta)$ follows now easily.

(2) We just proved this above as in Eq. (5.15).

(3) It is enough to consider the case $\mathcal{I}(\mu, \nu)$ finite, in which case we certainly have that both μ, ν have densities $\frac{d\mu}{d\beta}, \frac{d\nu}{d\beta} \in L^2(\beta)$. Writing, $\mu = \phi d\beta$ and $\nu = \psi d\beta$ with $\phi, \psi \in L^2(\beta)$, using (2.11) the inequality to be proved becomes equivalent to

$$\langle \mathcal{E}(\phi - \psi), (\phi - \psi) \rangle \leqslant \langle \phi - \psi, \phi - \psi \rangle$$

which follows from the fact that the spectrum of $\mathcal{E}$ (which is a bounded selfadjoint operator on $L^2(\beta)$) restricted to $L_0^2(\beta)$ is $\{1/n; n \geqslant 1\}$. Clearly, the equality is attained if $\phi(x) - \psi(x) = Cx$, the same thing as $\mu(dx) - \nu(dx) = Cx\beta(dx)$.

(4) Follows by scaling. $\quad\square$

Another consistent argument for the choice of the Fisher information is given by the following analogy with the classical case.

In the classical case of the Ornstein–Uhlenbeck operator L, the connection between the entropy and the Fisher information is given by the fact that the Fisher information appears naturally as the derivative of the entropy along the semigroup generated by L.

We want to draw a similar picture in the case of measures on the interval $[-2, 2]$ with the role of the Ornstein–Uhlenbeck operator L being played by the counting number operator $\mathcal{N}$ on functions on $[-2, 2]$.

What we have in mind here is the following. The semigroup generated by the counting number operator $\mathcal{N}$ is $\mathcal{P}_t$ and can be shown to be computed as

$$\mathcal{P}_t f(x) = \int k_t(x, y) f(y) \beta(dy)$$

with

$$k_t(x, y) = 1 + 2 \sum_{n \geqslant 1} e^{-tn} \phi_n(x) \phi_n(y)$$

where the factor 2 in front of the summation comes from the fact that $\int \phi_n^2 \beta(dy) = 1/2$.

In the classical case, the derivative of the entropy along the semigroup $\mathcal{P}_t$ is exactly the Fisher information. The same phenomena holds true in this local versions of the entropy and Fisher information.

More precisely, assume that $\mathcal{I}(\mu, \nu)$ is finite. This means that both measures μ, ν have densities in $L^2(\beta)$ and taking $\phi = \frac{d\mu}{d\beta} - \frac{d\nu}{d\beta}$, from (2.11),

$$\mathcal{H}(\mu, \nu) = \langle \mathcal{E}\phi, \phi \rangle.$$

If we consider now the measure $\mu_t = \mathcal{P}_t(\frac{d\mu}{d\beta}) d\beta$ and similarly $\nu_t = \mathcal{P}_t(\frac{d\nu}{d\beta}) d\beta$, then

$$\frac{d}{dt} \mathcal{H}(\mu_t, \nu_t) = \frac{d}{dt} \langle \mathcal{E}\mathcal{P}_t \phi, \mathcal{P}_t \phi \rangle = \langle \mathcal{E}\mathcal{N}\mathcal{P}_t \phi, \mathcal{P}_t \phi \rangle = \langle \mathcal{P}_t \phi, \mathcal{P}_t \phi \rangle = \mathcal{I}(\mu_t, \nu_t).$$

It is the content of Theorem 3 which actually unifies the two versions of the Fisher information, the one in terms of the Hilbert transform and the other one in terms of the densities of the measures involved.

Thus the two points of view outlined above converge toward the same thing and gives a consistent notion of Fisher information in this framework.

6. L^p considerations around local Log-Sobolev and an open problem

From the discussion in the previous section we learn that the local Log-Sobolev hold in the form

$$2\mathcal{H}(\mu, \nu) \leqslant \int |H\mu - H\nu|^2 \, d\alpha.$$

It is natural to ask if this still remains true if we replace the L^2 norm on the right-hand side by another L^p norm. Phrased differently, is it true that there is a $1 \leqslant p < 2$ and a constant $C_p > 0$ such that for all probability measures μ, ν on $[-2, 2]$,

$$C_p \mathcal{H}(\mu, \nu) \leqslant \left(\int |H\mu - H\nu|^p \, d\alpha \right)^{2/p}. \tag{6.1}$$

We do not know the answer to this question, but want to show that for $p < 3/2$ this is not possible. This is based on the example given by measures μ, ν such that for $0 < r < 1$ and a small constant η

$$\mu(dx) - \nu(dx) = \eta \left(\sum_{n \geqslant 1} r^{n-1} T_n(x/2) \right) \beta(dx).$$

From (5.6), we get that

$$H(\mu - \nu)(x) = \eta \left(\sum_{n \geqslant 1} r^{n-1} U_{n-1}(x/2) \right) = \frac{\eta}{1 - rx + r^2}.$$

With this choice and Proposition 1, we obtain that for $0 < r < 1$,

$$\mathcal{H}(\mu, \nu) = 2\eta^2 \sum_{n \geqslant 1} \frac{r^{2(n-1)}}{n} = -2\eta^2 \log(1 - r^2).$$

Now for $p < 3/2$,

$$\int |H\mu - H\nu|^p \, d\alpha = \frac{\eta^p}{2\pi} \int_{-2}^{2} \frac{1}{(1 - rx + r^2)^p} \sqrt{4 - x^2} \, dx = 16\eta^2 \int_0^1 \frac{\sqrt{u(1-u)}}{(4ru + (1-r)^2)^p} \, du$$

$$\leqslant \frac{16\eta^p}{(4r)^p} \int_0^1 u^{1/2 - p} \, du < \infty$$

where in the middle we made the change of variable $x = 2 - 4u$, $0 \leqslant u \leqslant 1$. The moral of this calculation is that (6.1) cannot hold true with $1 \leqslant p < 3/2$ because the right-hand side is bounded in $r \in (0, 1)$ and the left-hand side blows up as r approaches 1 from below.

For $r = 3/2$, using Mathematica, we obtain that

$$\int |H\mu - H\nu|^{3/2} \, d\alpha = 16\eta^{3/2} \int\limits_0^1 \frac{\sqrt{u(1-u)}}{(4ru + (1-r)^2)^{3/2}} \, du$$

$$= \frac{\eta^{3/2}}{8} \left(-4 + 6\log(2) - 2\log(1-r) \right) + O\left(\left(\log(1-r) \right)^2 \right)$$

which does not rule out (6.1).

For $p = 3/2$ we do not have a counterexample to (6.1) nor a proof of validity of it. We post this as an open problem here.

Open Problem 1. *There is a constant $C > 0$ such that for any probability measures μ, ν on $[-2, 2]$,*

$$C\mathcal{H}(\mu, \nu) \leqslant \left(\int |H\mu - H\nu|^{3/2} \, d\alpha \right)^{4/3}.$$

A positive answer to this question would give the optimal p for which (6.1) is true. A negative answer would continue with the following.

Open Problem 2. *Is there a $3/2 < p < 2$ such that for some constant $C_p > 0$, (6.1) holds true for any probability measures μ, ν? And if so, what is the smallest such p?*

A reformulation of these open problems in terms of trigonometric series can be done based on (5.6), (5.8) and the definition of the Chebyshev polynomials of second kind. Using an approximation of the measures μ, ν with measures of the form $\phi \, d\beta$ and $\psi \, d\beta$, an equivalent form of (6.1) is the following. What is the smallest $1 < p$ such that for any $n \geqslant 1$ and $a_1, a_2, \ldots, a_n \in \mathbb{R}$,

$$C_p \sum_{k=1}^n \frac{a_k^2}{k} \leqslant \left(\int\limits_0^\pi \left| \sum_{k=1}^n a_k \sin(kt) \right|^p \sin^{2-p}(t) \, dt \right)^{2/p} ? \tag{6.2}$$

The conclusion of this section is that definitely $3/2 \leqslant p$, but it is not clear that the smallest p is exactly $3/2$.

7. HWI inequality

This section is dedicated to a version of the celebrated HWI from [11]. The statement is the following.

Theorem 4. *For any probability measures μ, ν on $[-2, 2]$,*

$$\mathcal{H}(\mu, \nu) \leqslant \sqrt{2\mathcal{I}(\mu, \nu)} W_1(\mu, \nu) - \frac{1}{2} W_1^2(\mu, \nu) \tag{7.1}$$

with equality if $\mu(dx) - \nu(dx) = Cx\beta(dx)$.

Proof. It is sufficient to prove this in the case $\mathcal{I}(\mu, \nu)$ is finite. Thus, let $\eta = \frac{d\mu}{d\beta}$ and $\zeta = \frac{d\nu}{d\beta}$ with $\eta, \zeta \in L^2(\beta)$. Let $\psi = \eta - \zeta$. Clearly, $\psi \in L^2(\beta)$.

Here, the main observation is that conform Corollary 3, the inequality we need to prove writes equivalently as

$$\langle \mathcal{E}\psi, \psi \rangle \leqslant 2\sqrt{\langle \psi, \psi \rangle \langle \mathcal{E}^2 \psi, \psi \rangle} - \langle \mathcal{E}^2 \psi, \psi \rangle.$$

Writing $\psi = \sum_{n \geqslant 1} \alpha_n \phi_n$, this can be reinterpreted as

$$\left(\sum_{n \geqslant 1} \alpha_n^2 (1/n + 1/n^2) \right)^2 \leqslant 4 \sum_{n \geqslant 1} \alpha_n^2 \sum_{n \geqslant 1} \alpha_n^2 / n^2. \tag{$*$}$$

To justify this, apply Cauchy–Schwartz as

$$\left(\sum_{n \geqslant 1} \alpha_n^2 (1/n + 1/n^2) \right)^2 \leqslant \sum_{n \geqslant 1} \alpha_n^2 \sum_{n \geqslant 1} \alpha_n^2 (1/n + 1/n^2)^2$$

and then $(*)$ follows from the fact that $(1/n + 1/n^2)^2 \leqslant 4/n^2$. Tracing back all inequalities, we see that equality follows for the case $\mu(dx) - \nu(dx) = Cx\beta(dx)$. $\quad\square$

8. Remarks

(1) It is interesting to point out that these local versions of the functional inequalities are essentially on intervals. Taking an arbitrary set, say K, for instance a finite union of intervals, the local transportation still holds for all measures supported on K. This can be easily seen by simply considering the set K as a subset of an interval. Perhaps an interesting thing to investigate here is the significance of the best constant in the inequality and the measures for which this is achieved.

(2) It is interesting to figure out a similar global version of the Log-Sobolev inequality. This should have some connection with the global transportation inequality.

(3) For the local Log-Sobolev, if we take this on an arbitrary set K, then it would be nice to see a similar picture as in the case of the interval. It is not clear what the natural replacement of the measure α from the definition of $\mathcal{J}$ should be. Even in the interval case, it is somehow an interesting play between the semicircle and the arcsine laws, whose replacement is not obvious for an arbitrary set K.

Acknowledgments

I want to thank Florent Benaych-Georges for pointing to the author the work of [10]. I also want to thank Édouard Maurel-Segala and Mylène Maïda for their comments and particularly for the question from Remark 1.

Many thanks to the referee of this paper for the comments and suggested corrections which led to a better paper.

References

[1] Philippe Biane, Logarithmic Sobolev inequalities, matrix models and free entropy, in: Linfen, 2001, Acta Math. Sin. (Engl. Ser.) 19 (3) (2003) 497–506.

[2] Philippe Biane, Roland Speicher, Free diffusions, free entropy and free Fisher information, Ann. Inst. H. Poincare Probab. Statist. 37 (5) (2001) 581–606.

[3] Philippe Biane, Dan Voiculescu, A free probability analogue of the Wasserstein metric on a trace-state space, Geom. Funct. Anal. 11 (2001) 1125–1138.

[4] George Boole, On the comparison of transcendents, with certain applications to the theory of definite integrals, Philos. Trans. R. Soc. Lond. 147 (1857) 748–803.

[5] Percy A. Deift, Orthogonal Polynomials and Random Matrices: A Riemann–Hilbert Approach, Courant Lect. Notes Math., vol. 3, New York University Courant Institute of Mathematical Sciences, New York, 1999.

[6] Fumio Hiai, Denes Petz, Yoshimichi Ueda, Free transportation cost inequalities via random matrix approximation, Probab. Theory Related Fields 130 (2004) 199–221.

[7] Michel Ledoux, Ionel Popescu, The one dimensional free Poincaré inequality, Trans. Amer. Math. Soc., in press.

[8] Michel Ledoux, Ionel Popescu, Mass transportation proofs of free functional inequalities, and free Poincaré inequalities, J. Funct. Anal. 257 (4) (2009) 1175–1221.

[9] Lynn H. Loomis, A note on the Hilbert transform, Bull. Amer. Math. Soc. 52 (1946) 1082–1086.

[10] Mylène Maïda, Édouard Maurel-Segala, Free transport-entropy inequalities for non-convex potentials and application to concentration of random matrices, arXiv:1204.3208v1, 2012.

[11] Felix Otto, Cédric Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, J. Funct. Anal. 173 (2) (2000) 361–400.

[12] Alexei Poltoratski, Barry Simon, Maxim Zinchenko, The Hilbert transform of a measure, J. Anal. Math. 111 (2010) 247–265.

[13] Alexei G. Poltoratski, On the distributions of boundary values of Cauchy integrals, Proc. Amer. Math. Soc. 124 (8) (1996) 2455–2463.

[14] Ionel Popescu, Talagrand inequality for the semicircular law and energy of the eigenvalues of beta ensembles, Math. Res. Lett. 14 (6) (2007) 1023–1032.

[15] Edward B. Saff, Vilmos Totik, Logarithmic Potentials With External Fields, Grundlehren Math. Wiss. (Fund. Principles Math. Sci.), vol. 316, Springer-Verlag, Berlin, 1997.

[16] Cédric Villani, Topics in Optimal Transportation, Grad. Stud. Math., vol. 58, American Mathematical Society, Providence, RI, 2003.

[17] Dan V. Voiculescu, Ken J. Dykema, Alexandru Nica, A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups, in: Free Random Variables, in: CRM Monogr. Ser., vol. 1, American Mathematical Society, Providence, RI, 1992.

[18] Dan Voiculescu, Limit laws for random matrices and free products, Invent. Math. 104 (1) (1991) 201–220.

[19] Dan Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory. V. Noncommutative Hilbert transforms, Invent. Math. 132 (1) (1998) 189–227.