TRANSPORT INEQUALITIES. A SURVEY

NATHAEL GOZLAN, CHRISTIAN LÉONARD

Abstract. This is a survey of recent developments in the area of transport inequalities. We investigate their consequences in terms of concentration and deviation inequalities and sketch their links with other functional inequalities and also large deviation theory.

Introduction

In the whole paper, $\mathcal{X}$ is a polish (complete metric and separable) space equipped with its Borel σ-field and we denote $\mathrm{P}(\mathcal{X})$ the set of all Borel probability measures on $\mathcal{X}$.

Transport inequalities relate a cost $\mathcal{T}(\nu,\mu)$ of transporting a generic probability measure $\nu\in\mathrm{P}(\mathcal{X})$ onto a reference probability measure $\mu\in\mathrm{P}(\mathcal{X})$ with another functional $J(\nu|\mu)$. A typical transport inequality is written:

$$\alpha(\mathcal{T}(\nu,\mu))\le J(\nu|\mu),\qquad \text{for all } \nu\in\mathrm{P}(\mathcal{X}),$$

where $\alpha:[0,\infty)\to[0,\infty)$ is an increasing function with $\alpha(0)=0$. In this case, it is said that the reference probability measure μ satisfies $\alpha(\mathcal{T})\le J$.

Typical transport inequalities are built with $\mathcal{T}=W_p^p$ where W_p is the Wasserstein metric of order p, and $J(\cdot|\mu)=H(\cdot|\mu)$ is the relative entropy with respect to μ. The left-hand side of

$$\alpha(W_p^p)\le H$$

contains W which is built with some metric d on $\mathcal{X}$, while its right-hand side is the relative entropy H which, as Sanov's theorem indicates, is a measurement of the difficulty for a large sample of independent particles with common law μ to deviate from the prediction of the law of large numbers. On the left-hand side: a cost for displacing mass in terms of the ambient metric d; on the right-hand side: a cost for displacing mass in terms of fluctuations. Therefore, it is not a surprise that this interplay between displacement and fluctuations gives rise to a quantification of how fast $\mu(A^r)$ tends to 1 as $r\ge 0$ increases, where $A^r:=\{x\in\mathcal{X};d(x,y)\le r \text{ for some } y\in A\}$ is the enlargement of size r with respect to the metric d of the subset $A\subset\mathcal{X}$. Indeed, we shall see that such transport-entropy inequalities are intimately related to the concentration of measure phenomenon and to deviation inequalities for average observables of samples.

Date: February 21, 2016.

2000 *Mathematics Subject Classification.* 26D10, 60E15.

Key words and phrases. Transport inequalities, optimal transport, relative entropy, Fisher information, concentration of measure, deviation inequalities, logarithmic Sobolev inequalities, inf-convolution inequalities, large deviations.

Other transport inequalities are built with the Fisher information $I(\cdot|\mu)$ instead of the relative entropy on the right-hand side. It is known since Donsker and Varadhan, see [39, 35], that I is a measurement of the fluctuations of the occupation measure of a very long trajectory of a time-continuous Markov process with invariant ergodic law μ. Again, the transport-information inequality $\alpha(W_p^p) \leq I$ allows to quantify concentration and deviation properties of μ.

Finally, there exist also free transport inequalities. They compare a transport cost with a free relative entropy which is the large deviation rate function of the spectral empirical measures of large random matrices, as was proved by Ben Arous and Guionnet [13].

This is a survey paper about transport inequalities: a research topic which flied off in 1996 with the publications of several papers on the subject by Dembo, Marton, Talagrand and Zeitouni [32, 33, 77, 78, 102]. It was known from the end of the sixties that the total variation norm of the difference of two probability measures is controlled by their relative entropy. This is expressed by the Csiszár-Kullback-Pinsker inequality [90, 31, 64] which is a transport inequality from which deviation inequalities have been derived. But the keystone of the edifice was the discovery in 1986 by Marton [76] of the link between transport inequalities and the concentration of measure. This result was motivated by information theoretic problems; it remained unknown to the analysts and probabilists during ten years. Meanwhile, during the second part of the nineties, important progresses about the understanding of optimal transport have been achieved, opening the way to new unified proofs of several related functional inequalities, including a certain class of transport inequalities.

Concentration of measure inequalities can be obtained by means of other functional inequalities such as isoperimetric and logarithmic Sobolev inequalities, see the textbook by Ledoux [68] for an excellent account on the subject. Consequently, one expects that there are deep connections between these various inequalities. Indeed, during the recent years, these links have been explored and some of them have been clarified.

These recent developments will be sketched in the following pages.

No doubt that our treatment of this vast subject fails to be exhaustive. We apologize in advance for all kind of omissions. All comments, suggestions and reports of omissions are welcome.

Contents

The authors are grateful to the organizers of the conference on Inhomogeneous Random Systems for this opportunity to present transport inequalities to a broad community of physicists and mathematicians.

1. An overview

In order to present as soon as possible a couple of important transport inequalities and their consequences in terms of concentration of measure and deviation inequalities, let us recall precise definitions of the optimal transport cost and the relative entropy.

Optimal transport cost. Let c be a $[0,\infty)$-valued lower semicontinuous function on the polish product space $\mathcal{X}^2$ and fix $\mu,\nu \in \mathrm{P}(\mathcal{X})$. The Monge-Kantorovich optimal transport problem is

$$\text{(MK)} \qquad \textit{Minimize } \pi \in \mathrm{P}(\mathcal{X}^2) \mapsto \int_{\mathcal{X}^2} c(x,y)\, d\pi(x,y) \in [0,\infty] \qquad \textit{subject to } \pi_0 = \nu, \pi_1 = \mu$$

where $\pi_0, \pi_1 \in P(\mathcal{X})$ are the first and second marginals of $\pi \in \mathrm{P}(\mathcal{X}^2)$. Any $\pi \in \mathrm{P}(\mathcal{X}^2)$ such that $\pi_0 = \nu$ and $\pi_1 = \mu$ is called a *coupling* of ν and μ. The value of this convex minimization problem is

$$\mathcal{T}_c(\nu,\mu) := \inf\left\{ \int_{\mathcal{X}^2} c(x,y)\, d\pi(x,y); \pi \in \mathrm{P}(\mathcal{X}^2); \pi_0 = \nu, \pi_1 = \mu \right\} \in [0,\infty]. \tag{1}$$

It is called the *optimal cost* for transporting ν onto μ. Under the natural assumption that $c(x,x) = 0$, for all $x \in \mathcal{X}$, we have: $\mathcal{T}_c(\mu,\mu) = 0$, and $\mathcal{T}_c(\nu,\mu)$ can be interpreted as a cost for coupling ν and μ.
A popular cost function is $c = d^p$ with d a metric on $\mathcal{X}$ and $p \geq 1$. One can prove that under some conditions

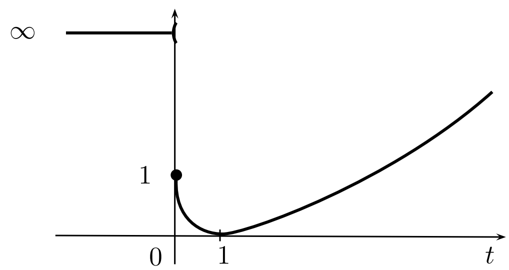
$$W_p(\nu,\mu) := \mathcal{T}_{d^p}(\nu,\mu)^{1/p}$$

defines a metric on a subset of $\mathrm{P}(\mathcal{X})$. This is the *Wasserstein metric* of order p (see e.g [104, Chp 6]). A deeper investigation of optimal transport is presented at Section 2. It will be necessary for a better understanding of transport inequalities.

Relative entropy. The relative entropy with respect to $\mu \in \mathrm{P}(\mathcal{X})$ is defined by

$$H(\nu|\mu) = \begin{cases} \int_{\mathcal{X}} \log\left(\frac{d\nu}{d\mu}\right) d\nu & \text{if } \nu \ll \mu \\ +\infty & \text{otherwise} \end{cases}, \quad \nu \in \mathrm{P}(\mathcal{X}).$$

For any probability measures $\nu \ll \mu$, one can rewrite $H(\nu|\mu) = \int h(d\nu/d\mu)\, d\mu$ with $h(t) = t\log t - t + 1$ which is a strictly convex nonnegative function such that $h(t) = 0 \Leftrightarrow t = 1$.



Graphic representation of $h(t) = t\log t - t + 1$.

Hence, $\nu \mapsto H(\nu|\mu) \in [0,\infty]$ is a convex function and $H(\nu|\mu) = 0$ if and only if $\nu = \mu$.

Transport inequalities. We can now define a general class of inequalities involving transport costs.

Definition 1.1 (Transport inequalities)**.** *Besides the cost function c, consider also two functions $J(\cdot|\mu) : \mathrm{P}(\mathcal{X}) \to [0,\infty]$ and $\alpha : [0,\infty) \to [0,\infty)$ an increasing function such that $\alpha(0) = 0$. One says that $\mu \in \mathrm{P}(\mathcal{X})$ satisfies the* transport inequality $\alpha(\mathcal{T}_c) \leq J$ *if*

$$(\alpha(\mathcal{T}_c) \leq J) \qquad \alpha(\mathcal{T}_c(\nu,\mu)) \leq J(\nu|\mu), \quad \text{for all } \nu \in \mathrm{P}(\mathcal{X}).$$

When $J(\cdot) = H(\cdot|\mu)$, one talks about transport-entropy inequalities*.*

For the moment, we focus on transport-entropy inequalities, but in Section 10, we shall encounter the class of transport-information inequalities, where the functional J is the Fisher information.

Note that, because of $H(\mu|\mu) = 0$, for the transport-entropy inequality to hold true, it is necessary that $\alpha(\mathcal{T}_c(\mu,\mu)) = 0$. A sufficient condition for the latter equality is

- $c(x,x) = 0$, for all $x \in \mathcal{X}$ and
- $\alpha(0) = 0$.

This will always be assumed in the remainder of this article.

Among this general family of inequalities, let us isolate the classical $\mathbf{T}_1$ and $\mathbf{T}_2$ inequalities. For $p = 1$ or $p = 2$, one says that $\mu \in \mathrm{P}_p := \{\nu \in \mathrm{P}(\mathcal{X}); \int d(x_o,\cdot)^p\, d\nu < \infty\}$ satisfies the inequality $\mathbf{T}_p(C)$, with $C > 0$ if

$$(\mathbf{T}_p(C)) \qquad W_p^2(\nu,\mu) \leq CH(\nu|\mu),$$

for all $\nu \in \mathrm{P}(\mathcal{X})$.

Remark 1.2. Note that this inequality implies that μ is such that $H(\nu|\mu) = \infty$ whenever $\nu \notin \mathrm{P}_p$.

With the previous notation, $\mathbf{T}_1(C)$ stands for the inequality $C^{-1}\mathcal{T}_d^2 \le H$ and $\mathbf{T}_2(C)$ for the inequality $C^{-1}\mathcal{T}_{d^2} \le H$. Applying Jensen inequality, we get immediately that

$$\mathcal{T}_d^2(\nu,\mu) \le \mathcal{T}_{d^2}(\nu,\mu). \tag{2}$$

As a consequence, for a given metric d on $\mathcal{X}$, the inequality $\mathbf{T}_1$ is always weaker than the inequality $\mathbf{T}_2$.

We now present two important examples of transport-entropy inequalities: the Csiszár-Kullback-Pinsker inequality, which is a $\mathbf{T}_1$ inequality and Talagrand's $\mathbf{T}_2$ inequality for the Gaussian measure.

Csiszár-Kullback-Pinsker inequality. The total variation distance between two probability measures ν and μ on $\mathcal{X}$ is defined by

$$\|\nu-\mu\|_{TV} = \sup|\nu(A)-\mu(A)|,$$

where the supremum runs over all measurable $A \subset \mathcal{X}$. It appears that the total variation distance is an optimal transport-cost. Namely, consider the so-called Hamming metric

$$d_H(x,y) = \mathbf{1}_{x\neq y},\ x,y\in\mathcal{X},$$

which assigns the value 1 if x is different from y and the value 0 otherwise. Then we have the following result whose proof can be found in e.g [81, Lemma 2.20].

Proposition 1.3. *For all $\nu,\mu \in \mathrm{P}(\mathcal{X})$, $\mathcal{T}_{d_H}(\nu,\mu) = \|\nu-\mu\|_{TV}$.*

The following theorem gives the celebrated Csiszár-Kullback-Pinsker inequality (see [90, 31, 64]).

Theorem 1.4. *The inequality*

$$\|\nu-\mu\|_{TV}^2 \le \frac{1}{2}H(\nu|\mu),$$

holds for all probability measures μ,ν on $\mathcal{X}$.

In other words, any probability μ on $\mathcal{X}$ enjoy the inequality $\mathbf{T}_1(1/2)$ with respect to the Hamming distance d_H on $\mathcal{X}$.

Proof. The following proof is taken from [104, Remark 22.12] and is attributed to Talagrand. Suppose that $H(\nu|\mu) < +\infty$ (otherwise there is nothing to prove) and let $f = \frac{d\nu}{d\mu}$ and $u = f-1$. By definition and since $\int u\,d\mu = 0$,

$$H(\nu|\mu) = \int_{\mathcal{X}} f\log f\,d\mu = \int_{\mathcal{X}} (1+u)\log(1+u) - u\,d\mu.$$

The function $\varphi(t) = (1+t)\log(1+t)-t$, verifies $\varphi'(t) = \log(1+t)$ and $\varphi''(t) = \frac{1}{1+t}$, $t>-1$. So, using a Taylor expansion,

$$\varphi(t) = \int_0^t (t-x)\varphi''(x)\,dx = t^2\int_0^1 \frac{1-s}{1+st}\,ds,\ t>-1.$$

So,

$$H(\nu|\mu) = \int_{\mathcal{X}\times[0,1]} \frac{u^2(x)(1-s)}{1+su(x)}\,ds\,d\mu(x).$$

According to Cauchy-Schwarz inequality,

$$\begin{aligned}\Big(\int_{\mathcal{X}\times[0,1]} |u|(x)(1-s)\, d\mu(x)ds\Big)^2 \\ &\le \int_{\mathcal{X}\times[0,1]} \frac{u(x)^2(1-s)}{1+su(x)}\, d\mu(x)ds \cdot \int_{\mathcal{X}\times[0,1]} (1-s)(1+su(x))\, d\mu(x)ds \\ &= \frac{H(\nu|\mu)}{2}.\end{aligned}$$

Since $\|\nu-\mu\|_{TV} = \frac{1}{2}\int |1-f|\, d\mu$, the left-hand side equals $\|\nu-\mu\|_{TV}^2$ and this completes the proof. $\square$

Talagrand's transport inequality for the Gaussian measure. In [102], Talagrand proved the following transport inequality $\mathbf{T}_2$ for the standard Gaussian measure γ on $\mathbb{R}$ equipped with the standard distance $d(x,y) = |x-y|$.

Theorem 1.5. *The standard Gaussian measure γ on $\mathbb{R}$ verifies*

$$W_2^2(\nu,\gamma) \le 2H(\nu|\gamma), \tag{3}$$

for all $\nu \in \mathrm{P}(\mathbb{R})$.

This inequality is sharp. Indeed, taking ν to be a translation of γ, that is a normal law with unit variance, we easily check that equality holds true.

Proof. In the following lines, we present the short and elegant proof of (3), as it appeared in [102]. Let us consider a reference measure

$$d\mu(x) = e^{-V(x)}\, dx.$$

We shall specify later to the Gaussian case, where the potential V is given by $V(x) = x^2/2 + \log(2\pi)/2$, $x \in \mathbb{R}$. Let ν be another probability measure on $\mathbb{R}$. It is known since Fréchet that any measurable map $y = T(x)$ which verifies the equation

$$\nu((-\infty, T(x)]) = \mu((-\infty, x]), \quad x \in \mathbb{R} \tag{4}$$

is a coupling of ν and μ, i.e. such that $\nu = T_\#\mu$, which minimizes the average squared distance (or equivalently: which maximizes the correlation), see (22) below for a proof of this statement. Such a transport map is called a monotone rearrangement. Clearly T is increasing, and assuming from now on that $\nu = f\mu$ is absolutely continuous with respect to μ, one sees that T is Lebesgue almost everywhere differentiable with $T' > 0$. Equation (4) becomes for all real x, $\int_{-\infty}^{T(x)} f(z)e^{-V(z)}\, dz = \int_{-\infty}^{x} e^{-V(z)}\, dz$. Differentiating, one obtains

$$T'(x)f(T(x))e^{-V(T(x))} = e^{-V(x)}, \quad x \in \mathbb{R}. \tag{5}$$

The relative entropy writes: $H(\nu|\mu) = \int \log(f)\, d\nu = \int \log(f(T(x))\, d\mu$ since $\nu = T_\#\mu$. Extracting $f(T(x))$ from (5) and plugging it into this identity, we obtain

$$H(\nu|\mu) = \int [V(T(x)) - V(x) - \log T'(x)]\, e^{-V(x)}\, dx.$$

On the other hand, we have $\int (T(x)-x)V'(x)e^{-V(x)}\,dx = \int (T'(x)-1)e^{-V(x)}\,dx$ as a result of an integration by parts. Therefore,

$$\begin{aligned} H(\nu|\mu) &= \int \Big(V(T(x)) - V(x) - V'(x)[T(x)-x]\Big)\,d\mu(x) \\ &\qquad\qquad + \int (T'(x) - 1 - \log T'(x))\,d\mu(x). \\ &\geq \int \Big(V(T(x)) - V(x) - V'(x)[T(x)-x]\Big)\,d\mu(x) \end{aligned} \tag{6}$$

where we took advantage of $b-1-\log b \geq 0$ for all $b>0$, at the last inequality. Of course, the last integral is nonnegative if V is assumed to be convex.
Considering the Gaussian potential $V(x) = x^2/2 + \log(2\pi)/2$, $x \in \mathbb{R}$, we have shown that

$$H(\nu|\gamma) \geq \int_{\mathbb{R}} (T(x)-x)^2/2\ d\gamma(x) \geq W_2^2(\nu,\gamma)/2$$

for all $\nu \in \mathrm{P}(\mathbb{R})$, which is (3). $\square$

Concentration of measure. If d is a metric on $\mathcal{X}$, for any $r \geq 0$, one defines the r-neighborhood of the set $A \subset \mathcal{X}$ by

$$A^r := \{x \in \mathcal{X}; d(x,A) \leq r\},\ r \geq 0,$$

where $d(x,A) := \inf_{y\in A} d(x,y)$ is the distance of x from A.

Let $\beta : [0,\infty) \to \mathbb{R}^+$ such that $\beta(r) \to 0$ when $r \to \infty$; it is said that the probability measure μ verifies the *concentration inequality* with profile β if

$$\mu(A^r) \geq 1 - \beta(r),\ r \geq 0,$$

for all measurable $A \subset \mathcal{X}$, with $\mu(A) \geq 1/2$.

According to the following classical proposition, the concentration of measure (with respect to metric enlargement) can be alternatively described in terms of deviations of Lipschitz functions from their median.

Proposition 1.6. *Let $(\mathcal{X}, d)$ be a metric space, $\mu \in \mathrm{P}(\mathcal{X})$ and $\beta : [0,\infty) \to [0,1]$; the following propositions are equivalent*

(1) The probability μ verifies the concentration inequality

$$\mu(A^r) \geq 1 - \beta(r),\ r \geq 0,$$

for all $A \subset \mathcal{X}$, with $\mu(A) \geq 1/2$.

(2) For all 1-Lipschitz function $f : \mathcal{X} \to \mathbb{R}$,

$$\mu(f > m_f + r) \leq \beta(r),\ r \geq 0,$$

where m_f denotes a median of f.

Proof. (1) $\Rightarrow$ (2). Let f be a 1-Lipschitz function and define $A = \{f \leq m_f\}$. Then it is easy to check that $A^r \subset \{f \leq m_f + r\}$. Since $\mu(A) \geq 1/2$, one has $\mu(f \leq m_f + r) \geq \mu(A^r) \geq 1 - \beta(r)$, for all $r \geq 0$.

(2) ⇒ (1). For all $A \subset \mathcal{X}$, the function $f_A : x \mapsto d(x,A)$ is 1-Lipschitz. If $\mu(A) \geq 1/2$, then 0 is median of f_A. Since $A^r = \{f_A \leq r\}$, one has $\mu(A^r) \geq 1 - \mu\{f_A > r\} \geq 1 - \beta(r)$, $r \geq 0$. □

Applying the deviation inequality to $\pm f$, we arrive at

$$\mu(|f - m_f| < r) \leq 2\beta(r),\ r \geq 0.$$

In other words, Lipschitz functions are, with a high probability, concentrated around their median, when the concentration profile β decreases rapidly to zero. In the above proposition, the median can be replaced by the mean $\mu(f)$ of f (see e.g. [68]):

$$\mu(f > \mu(f) + r) \leq \beta(r),\ r \geq 0. \tag{7}$$

The following theorem explains how to derive concentration inequalities (with profiles decreasing exponentially fast) from transport-entropy inequalities of the form $\alpha(\mathcal{T}_d) \leq H$, where the cost function c is the metric d. The argument used in the proof is due to Marton [76] and is referred as "Marton's argument" in the literature.

Theorem 1.7. *Let $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ be a bijection and suppose that $\mu \in \mathrm{P}(\mathcal{X})$ verifies the transport-entropy inequality $\alpha(\mathcal{T}_d) \leq H$. Then, for all measurable $A \subset \mathcal{X}$ with $\mu(A) \geq 1/2$, the following concentration inequality holds*

$$\mu(A^r) \geq 1 - e^{-\alpha(r - r_o)}, \quad \forall r \geq r_o := \alpha^{-1}(\log 2),$$

where A^r is the enlargement of A for the metric d which is defined above.

Equivalently, for all 1-Lipschitz $f : \mathcal{X} \to \mathbb{R}$, the following inequality holds

$$\mu(f > m_f + r + r_o) \leq e^{-\alpha(r)},\ r \geq 0.$$

Proof. Take $A \subset \mathcal{X}$, with $\mu(A) \geq 1/2$ and set $B = \mathcal{X} \setminus A^r$. Consider the probability measures $d\mu_A(x) = \frac{1}{\mu(A)}\mathbf{1}_A(x)\,d\mu(x)$ and $d\mu_B(x) = \frac{1}{\mu(B)}\mathbf{1}_B(x)\,d\mu(x)$. Obviously, if $x \in A$ and $y \in B$, then $d(x,y) \geq r$. Consequently, if π is a coupling between μ_A and μ_B, then $\int d(x,y)\,d\pi(x,y) \geq r$ and so $\mathcal{T}_d(\mu_A, \mu_B) \geq r$. Now, using the triangle inequality and the transport-entropy inequality we get

$$r \leq \mathcal{T}_d(\mu_A, \mu_B) \leq \mathcal{T}_d(\mu_A, \mu) + \mathcal{T}_d(\mu_B, \mu) \leq \alpha^{-1}\left(H(\mu_A|\mu)\right) + \alpha^{-1}\left(H(\mu_B|\mu)\right).$$

It is easy to check that $H(\mu_A|\mu) = -\log \mu(A) \leq \log 2$ and $H(\mu_B|\mu) = -\log(1 - \mu(A^r))$. It follows immediately that $\mu(A^r) \geq 1 - e^{-\alpha(r - r_o)}$, for all $r \geq r_o := \alpha^{-1}(\log 2)$. □

If μ verifies $\mathbf{T}_2(C)$, by (2) it also verifies $\mathbf{T}_1(C)$ and one can apply Theorem 1.7. Therefore, it appears that if μ verifies $\mathbf{T}_1(C)$ or $\mathbf{T}_2(C)$, then it concentrates like a Gaussian measure:

$$\mu(A^r) \geq 1 - e^{-(r - r_o)^2/C},\ r \geq r_o = \sqrt{C \log(2)}.$$

At this stage, the difference between $\mathbf{T}_1$ and $\mathbf{T}_2$ is invisible. It will appear clearly in the next paragraph devoted to tensorization of transport-entropy inequalities.

Tensorization. A central question in the field of concentration of measure is to obtain concentration estimates not only for μ but for the entire family $\{\mu^n; n \geq 1\}$ where μ^n denotes the product probability measure $\mu \otimes \cdots \otimes \mu$ on $\mathcal{X}^n$. To exploit transport-entropy inequalities, one has to know how they tensorize. This will be investigated in details in Section 4. Let us give in this introductory section, an insight on this important question.

It is enough to understand what happens with the product $\mathcal{X}_1 \times \mathcal{X}_2$ of two spaces. Indeed, it will be clear in a moment that the extension to the product of n spaces will follow by induction.
We are interested in transport costs from $\mathcal{X}_1$ to $\mathcal{Y}_1$, from $\mathcal{X}_2$ to $\mathcal{Y}_2$ and from $\mathcal{X}_1 \times \mathcal{X}_2$ to $\mathcal{Y}_1 \times \mathcal{Y}_2$. Let μ_1, μ_2 be two probability measures on two polish spaces $\mathcal{X}_1$, $\mathcal{X}_2$, respectively. The cost functions $c_1(x_1, y_1)$ and $c_2(x_2, y_2)$ on $\mathcal{X}_1 \times \mathcal{X}_1$ and $\mathcal{X}_2 \times \mathcal{X}_2$ give rise to the optimal transport cost functions $\mathcal{T}_{c_1}(\nu_1, \mu_1)$, $\nu_1 \in \mathrm{P}(\mathcal{X}_1)$ and $\mathcal{T}_{c_2}(\nu_2, \mu_2)$, $\nu_2 \in \mathrm{P}(\mathcal{X}_2)$.
On the product space $\mathcal{X}_1 \times \mathcal{X}_2$, we now consider the product measure $\mu_1 \otimes \mu_2$ and the cost function

$$c_1 \oplus c_2\big((x_1, y_1), (x_2, y_2)\big) := c_1(x_1, y_1) + c_2(x_2, y_2), \quad x_1, y_1 \in \mathcal{X}_1, x_2, y_2 \in \mathcal{X}_2$$

which give rise to the tensorized optimal transport cost function

$$\mathcal{T}_{c_1 \oplus c_2}(\nu, \mu_1 \otimes \mu_2), \quad \nu \in \mathrm{P}(\mathcal{X}_1 \times \mathcal{X}_2).$$

A fundamental example is $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y}_1 = \mathcal{Y}_2 = \mathbb{R}^k$ with $c_1(x, y) = c_2(x, y) = |y - x|_2^2$: the Euclidean metric on $\mathbb{R}^k$ tensorizes as the squared Euclidean metric on $\mathbb{R}^{2k}$.

For any probability measure ν on the product space $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$, let us write the disintegration of ν (conditional expectation) as follows:

$$d\nu(y_1, y_2) = d\nu_1(y_1) d\nu_2^{y_1}(y_2). \tag{8}$$

As was suggested by Marton [77] and Talagrand [102], it is possible to prove the intuitively clear following assertion:

$$\mathcal{T}_{c_1 \oplus c_2}(\nu, \mu_1 \otimes \mu_2) \leq \mathcal{T}_{c_1}(\nu_1, \mu_1) + \int_{\mathcal{Y}_1} \mathcal{T}_{c_2}(\nu_2^{y_1}, \mu_2)\, d\nu_1(y_1). \tag{9}$$

We give a detailed proof of this claim at the Appendix, Proposition A.1.
On the other hand, it is well-known that the fundamental property of the logarithm together with the product form of the disintegration formula (8) yield the analogous tensorization property of the relative entropy:

$$H(\nu \mid \mu_1 \otimes \mu_2) = H(\nu_1 \mid \mu_1) + \int_{\mathcal{Y}_1} H(\nu_2^{y_1} \mid \mu_2)\, d\nu_1(y_1). \tag{10}$$

Recall that the inf-convolution of two functions α_1 and α_2 on $[0, \infty)$ is defined by

$$\alpha_1 \Box \alpha_2(t) := \inf\{\alpha_1(t_1) + \alpha_2(t_2); t_1, t_2 \geq 0 : t_1 + t_2 = t\}, \quad t \geq 0.$$

Proposition 1.8. *Suppose that the transport-entropy inequalities*

$$\begin{aligned} \alpha_1(\mathcal{T}_{c_1}(\nu_1, \mu_1)) &\leq H(\nu_1 \mid \mu_1), \ \forall \nu_1 \in \mathrm{P}(\mathcal{X}_1) \\ \alpha_2(\mathcal{T}_{c_2}(\nu_2, \mu_2)) &\leq H(\nu_2 \mid \mu_2), \ \forall \nu_2 \in \mathrm{P}(\mathcal{X}_2) \end{aligned}$$

hold with $\alpha_1, \alpha_2 : [0, \infty) \to [0, \infty)$ convex increasing functions. Then, on the product space $\mathcal{X}_1 \times \mathcal{X}_2$, we have

$$\alpha_1 \Box \alpha_2\big(\mathcal{T}_{c_1 \oplus c_2}(\nu, \mu_1 \otimes \mu_2)\big) \leq H(\nu \mid \mu_1 \otimes \mu_2),$$

for all $\nu \in \mathrm{P}(\mathcal{X}_1 \times \mathcal{X}_2)$.

Proof. Take $\mathcal{Y}_1 = \mathcal{X}_1$ and $\mathcal{Y}_2 = \mathcal{X}_2$. For all $\nu \in \mathrm{P}(\mathcal{X}_1 \times \mathcal{X}_2)$,

$$\begin{aligned}
\alpha_1 \square \alpha_2 (\mathcal{T}_{c_1 \oplus c_2}(\nu, \mu_1 \otimes \mu_2)) &\overset{(a)}{\leq} \alpha_1 \square \alpha_2 \left(\mathcal{T}_{c_1}(\nu_1, \mu_1) + \int_{\mathcal{Y}_1} \mathcal{T}_{c_2}(\nu_2^{y_1} \mu_2)\, d\nu_1(y_1) \right) \\
&\overset{(b)}{\leq} \alpha_1(\mathcal{T}_{c_1}(\nu_1, \mu_1)) + \alpha_2 \left(\int_{\mathcal{Y}_1} \mathcal{T}_{c_2}(\nu_2^{y_1}, \mu_2)\, d\nu_1(y_1) \right) \\
&\overset{(c)}{\leq} \alpha_1(\mathcal{T}_{c_1}(\nu_1, \mu_1)) + \int_{\mathcal{Y}_1} \alpha_2 \big(\mathcal{T}_{c_2}(\nu_2^{y_1}, \mu_2) \big)\, d\nu_1(y_1) \\
&\overset{(d)}{\leq} H(\nu_1 \mid \mu_1) + \int_{\mathcal{Y}_1} H(\nu_2^{y_1} \mid \mu_2)\, d\nu_1(y_1) \\
&= H(\nu \mid \mu_1 \otimes \mu_2).
\end{aligned}$$

Inequality (a) is verified thanks to (9) since $\alpha_1 \square \alpha_2$ is increasing, (b) follows from the very definition of the inf-convolution, (c) follows from Jensen inequality since α_2 is convex, (d) follows from the assumptions $\alpha_1(\mathcal{T}_1(\nu_1)) \leq H(\nu_1 \mid \mu_1)$ for all ν_1 and $\alpha_2(\mathcal{T}_2(\nu_2)) \leq H(\nu_2 \mid \mu_2)$ for all ν_2 (with obvious notations) and the last equality is (10). $\square$

Obviously, it follows by an induction argument on the dimension n that, if μ verifies $\alpha(\mathcal{T}_c) \leq H$, then μ^n verifies $\alpha^{\square n}(\mathcal{T}_{c^{\oplus n}}) \leq H$ where as a definition

$$c^{\oplus n}\Big((x_1, y_1), \ldots, (x_n, y_n)\Big) := \sum_{i=1}^{n} c(x_i, y_i).$$

Since $\alpha^{\square n}(t) = n\alpha(t/n)$ for all $t \geq 0$, we have proved the next proposition.

Proposition 1.9. *Suppose that $\mu \in \mathrm{P}(\mathcal{X})$ verifies the transport-entropy inequality $\alpha(\mathcal{T}_c) \leq H$ with $\alpha : [0, \infty) \to [0, \infty)$ a convex increasing function. Then, $\mu^n \in \mathrm{P}(\mathcal{X}^n)$ verifies the transport-entropy inequality*

$$n\alpha\Big(\mathcal{T}_{c^{\oplus n}}(\nu, \mu^n)/n\Big) \leq H(\nu|\mu^n), \quad \forall \nu \in \mathrm{P}(\mathcal{X}^n).$$

We also give at the end of Section 3 an alternative proof of this result which is based on a duality argument. The general statements of Propositions 1.8 and 1.9 appeared in the authors' paper [53].

In particular, when α is linear, one observes that the inequality $\alpha(\mathcal{T}_c) \leq H$ tensorizes independently of the dimension. This is for example the case for the inequality $\mathbf{T}_2$. So, using the one dimensional $\mathbf{T}_2$ verified by the standard Gaussian measure γ together with the above tensorization property, we conclude that for all positive integer n, the standard Gaussian measure γ^n on $\mathbb{R}^n$ verifies the inequality $\mathbf{T}_2(2)$.

Now let us compare the concentration properties of product measures derived from $\mathbf{T}_1$ or $\mathbf{T}_2$. Let d be a metric on $\mathcal{X}$, and let us consider the ℓ_1 and ℓ_2 product metrics associated to d:

$$d_1(x, y) = \sum_{i=1}^{n} d(x_i, y_i) \quad \text{and} \quad d_2(x, y) = \left(\sum_{i=1}^{n} d^2(x_i, y_i) \right)^{1/2}, \; x, y \in \mathcal{X}^n.$$

The distance d_1 and d_2 are related by the following obvious inequality

$$\frac{1}{\sqrt{n}}d_1(x,y) \leq d_2(x,y) \leq d_1(x,y),\ x,y \in \mathcal{X}^n.$$

If μ verifies $\mathbf{T}_1$ on $\mathcal{X}$, then according to Proposition 1.9, μ^n verifies the inequality $\mathbf{T}_1(nC)$ on the space $\mathcal{X}^n$ equipped with the metric d_1. It follows from Marton's concentration Theorem 1.7, that

$$\mu^n(f > m_f + r + r_o) \leq e^{-\frac{r^2}{nC}},\ r \geq r_o = \sqrt{nC\log(2)}, \tag{11}$$

for all function f which is 1-Lipschitz with respect to d_1. So the constants appearing in the concentration inequality are getting worse and worse when the dimension increases.

On the other hand, if μ verifies $\mathbf{T}_2(C)$, then according to Proposition 1.9, μ^n verifies the inequality $\mathbf{T}_2(C)$ on the space $\mathcal{X}^n$ equipped with d_2. Thanks to Jensen inequality μ^n also verifies the inequality $\mathbf{T}_1(C)$ on $(\mathcal{X}^n, d_2)$, and so

$$\mu^n(g > m_g + r + r_o) \leq e^{-\frac{r^2}{C}},\ r \geq r_o = \sqrt{C\log(2)}, \tag{12}$$

for all function g which is 1-Lipschitz with respect to d_2. This time, one observes that the concentration profile does not depend on the dimension n. This phenomenon is called (Gaussian) *dimension-free concentration of measure*. For instance, if $\mu = \gamma$ is the standard Gaussian measure, we thus obtain

$$\gamma^n(f > m_f + r + r_o) \geq 1 - e^{-r^2/2},\ r \geq r_o := \sqrt{2\log 2} \tag{13}$$

for all function f which is 1-Lipschitz for the Euclidean distance on $\mathbb{R}^n$. This result is very near the optimal concentration profile obtained by an isoperimetric method, see [68]. In fact the Gaussian dimension-free property (12) is intrinsically related to the inequality $\mathbf{T}_2$. Indeed, a recent result of Gozlan [52] presented in Section 5 shows that Gaussian dimension concentration holds if and only if the reference measure μ verifies $\mathbf{T}_2$ (see Theorem 5.4 and Corollary 5.5).

Since a 1-Lipschitz function f for d_1 is $\sqrt{n}$-Lipschitz for d_2, it is clear that (12) gives back (11), when applied to $g = f/\sqrt{n}$. On the other hand, a 1-Lipschitz function g for d_2 is also 1-Lipschitz for d_1, and its is clear that for such a function g the inequality (12) is much better than (11) applied to $f = g$. So, we see from this considerations that $\mathbf{T}_2$ is a much stronger property than $\mathbf{T}_1$. We refer to [68] or [99, 101], for examples of applications where the independence on n in concentration inequalities plays a decisive role.

Nevertheless, dependence on n in concentration is not always something to fight against, as shown in the following example of deviation inequalities. Indeed, suppose that μ verifies the inequality $\alpha(\mathcal{T}_d) \leq H$, then for all positive integer n,

$$\mu^n\left(f \geq \int f\,d\mu^n + t\right) \leq e^{-n\alpha(t/n)},\ t \geq 0,$$

for all f 1-Lipschitz for d_1 (see Corollary 5.3). In particular, choose $f(x) = u(x_1) + \cdots u(x_n)$, with u a 1-Lipschitz function for d; then f is 1-Lipschitz for d_1, and so if X_i is an i.i.d sequence of law μ, we easily arrive at the following deviation inequality

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} u(X_i) \geq \mathbb{E}[u(X_1)] + t\right) \leq e^{-n\alpha(t)},\ t \geq 0.$$

This inequality presents the right dependence on n. Namely, according to Cramér theorem (see [34]) this probability behaves like $e^{-n\Lambda_u^*(t)}$ when n is large, where Λ_u^* is the Cramér transform of $u(X_1)$. The reader can look at [53] for more information on this subject. Let us mention that this family of deviation inequalities characterize the inequality $\alpha(\mathcal{T}_d) \leq H$ (see Theorem 5.2 and Corollary 5.3).

2. Optimal transport

Optimal transport is an active field of research. The recent textbooks by Villani [103, 104] make a very good account on the subject. Here, we recall basic results which will be necessary to understand transport inequalities. But the interplay between optimal transport and functional inequalities in general is wider than what will be exposed below, see [103, 104] for instance.

Let us make our underlying assumptions precise. The cost function c is assumed to be a lower semicontinuous $[0, \infty)$-valued function on the product $\mathcal{X}^2$ of the polish space $\mathcal{X}$. The Monge-Kantorovich problem with cost function c and marginals ν, μ in $\mathrm{P}(\mathcal{X})$, as well as its optimal value $\mathcal{T}_c(\nu, \mu)$ were stated at (MK) and (1) in Section 1.

Proposition 2.1. *The Monge-Kantorovich problem* (MK) *admits a solution if and only if* $\mathcal{T}_c(\nu, \mu) < \infty$.

Outline of the proof. The main ingredients of the proof of this proposition are

- the compactness with respect to the narrow topology of $\{\pi \in \mathrm{P}(\mathcal{X}^2); \pi_0 = \nu, \pi_1 = \mu\}$ which is inherited from the tightness of ν and μ and
- the lower semicontinuity of $\pi \mapsto \int_{\mathcal{X}^2} c\, d\pi$ which is inherited from the lower semicontinuity of c.

The polish assumption on $\mathcal{X}$ is invoked at the first item. □

The minimizers of (MK) are called optimal transport plans, they are not unique in general since (MK) is not a strictly convex problem: it is an infinite dimensional linear programming problem.

If d is a lower semicontinuous metric on $\mathcal{X}$ (possibly different from the metric which turns $\mathcal{X}$ into a polish space), one can consider the cost function $c = d^p$ with $p \geq 1$. One can prove that $W_p(\nu, \mu) := \mathcal{T}_{d^p}(\nu, \mu)^{1/p}$ defines a metric on the set $P_{d^p}(\mathcal{X})$ (or P_p for short) of all probability measures which integrate $d^p(x_o, \cdot)$: it is the so-called Wasserstein metric of order p. Since $\mathcal{T}_{d^p}(\nu, \mu) < \infty$ for all ν, μ in P_p, Proposition 2.1 tells us that the corresponding problem (MK) is attained in P_p.

Kantorovich dual equality. In the perspective of transport inequalities, the keystone is the following result. Let $\mathcal{C}_b(\mathcal{X})$ be the space of all continuous bounded functions on $\mathcal{X}$ and denote $u \oplus v(x,y) = u(x) + v(y)$, $x, y \in \mathcal{X}$.

Theorem 2.2 (Kantorovich dual equality)**.** *For all μ and ν in $\mathrm{P}(\mathcal{X})$, we have*

$$\begin{aligned}(14)\quad \mathcal{T}_c(\nu,\mu) &= \sup\left\{\int_{\mathcal{X}} u(x)\,d\nu(x) + \int_{\mathcal{X}} v(y)\,d\mu(y); u,v \in \mathcal{C}_b(\mathcal{X}), u\oplus v \le c\right\}\\ (15)\quad &= \sup\left\{\int_{\mathcal{X}} u(x)\,d\nu(x) + \int_{\mathcal{X}} v(y)\,d\mu(y); u \in L^1(\nu), v \in L^1(\mu), u\oplus v \le c\right\}.\end{aligned}$$

Note that for all π such that $\pi_0 = \nu, \pi_1 = \mu$ and (u,v) such that $u \oplus v \le c$, we have $\int_{\mathcal{X}} u\,d\nu + \int_{\mathcal{X}} v\,d\mu = \int_{\mathcal{X}^2} u \oplus v\,d\pi \le \int_{\mathcal{X}^2} c\,d\pi$. Optimizing both sides of this inequality leads us to

$$\begin{aligned}(16)\quad &\sup\left\{\int_{\mathcal{X}} u\,d\nu + \int_{\mathcal{X}} v\,d\mu;\ u \in L^1(\nu), v \in L^1(\mu), u\oplus v \le c\right\}\\ &\qquad\le \inf\left\{\int_{\mathcal{X}^2} c\,d\pi; \pi \in \mathrm{P}(\mathcal{X}^2); \pi_0 = \nu, \pi_1 = \mu\right\}\end{aligned}$$

and Theorem 2.2 appears to be a *no dual gap* result.

The following is a sketch of proof which is borrowed from Léonard's paper [72].

Outline of the proof of Theorem 2.2. For a detailed proof, see [72, Thm 2.1]. Denote $\mathrm{M}(\mathcal{X}^2)$ the space of all signed measures on $\mathcal{X}^2$ and $\iota_{\{x\in A\}} = \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{otherwise}\end{cases}$. Consider the $(-\infty, +\infty]$-valued function

$$K(\pi,(u,v)) = \int_{\mathcal{X}} u\,d\nu + \int_{\mathcal{X}} v\,d\mu - \int_{\mathcal{X}^2} u\oplus v\,d\pi + \int_{\mathcal{X}^2} c\,d\pi + \iota_{\{\pi\ge 0\}}, \quad \pi \in \mathrm{M}(\mathcal{X}^2), u,v \in \mathcal{C}_b(\mathcal{X}).$$

For each fixed (u,v), it is a convex function of π and for each fixed π, it is a concave function of (u,v). In other words, K is a convex-concave function and one can expect that it admits a saddle value, i.e.

$$(17)\quad \inf_{\pi\in\mathrm{M}(\mathcal{X}^2)} \sup_{u,v\in\mathcal{C}_b(\mathcal{X})} K(\pi,(u,v)) = \sup_{u,v\in\mathcal{C}_b(\mathcal{X})} \inf_{\pi\in\mathrm{M}(\mathcal{X}^2)} K(\pi,(u,v)).$$

The detailed proof amounts to check that standard assumptions for this min-max result hold true for (u,v) as in (14). We are going to show that (17) is the desired equality (14). Indeed, for fixed π,

$$\begin{aligned}\sup_{(u,v)} K(\pi,(u,v)) &= \int_{\mathcal{X}^2} c\,d\pi + \iota_{\{\pi\ge0\}} + \sup_{(u,v)}\left\{\int_{\mathcal{X}} u\,d\nu + \int_{\mathcal{X}} v\,d\mu - \int_{\mathcal{X}^2} u\oplus v\,d\pi\right\}\\ &= \int_{\mathcal{X}^2} c\,d\pi + \iota_{\{\pi\ge0\}} + \sup_{(u,v)}\left\{\int_{\mathcal{X}} u\,d(\nu-\pi_0) + \int_{\mathcal{X}} v\,d(\mu-\pi_1)\right\}\\ &= \int_{\mathcal{X}^2} c\,d\pi + \iota_{\{\pi_0=\nu,\pi_1=\mu\}}\end{aligned}$$

and for fixed (u,v),

$$\inf_\pi K(\pi,(u,v)) = \int_{\mathcal{X}} u\,d\nu + \int_{\mathcal{X}} v\,d\mu + \inf_{\pi\geq 0}\int_{\mathcal{X}^2}(c-u\oplus v)\,d\pi = \int_{\mathcal{X}} u\,d\nu + \int_{\mathcal{X}} v\,d\mu - \iota_{\{u\oplus v\leq c\}}.$$

Once (14) is obtained, (15) follows immediately from (16) and the following obvious inequality: $\sup_{u,v\in\mathcal{C}_b(\mathcal{X}),u\oplus v\leq c} \leq \sup_{u\in L^1(\nu),v\in L^1(\mu),u\oplus v\leq c}$. □

Let u and v be measurable functions on $\mathcal{X}$ such that $u\oplus v\leq c$. The family of inequalities $v(y)\leq c(x,y)-u(x)$, for all x,y is equivalent to $v(y)\leq \inf_x\{c(x,y)-u(x)\}$ for all y. Therefore, the function

$$u^c(y) := \inf_{x\in\mathcal{X}}\{c(x,y)-u(x)\},\quad y\in\mathcal{X}$$

satisfies $u^c\geq v$ and $u\oplus u^c\leq c$. As $J(u,v) := \int_{\mathcal{X}} u\,d\nu + \int_{\mathcal{X}} v\,d\mu$ is an increasing function of its arguments u and v, in view of maximizing J on the set $\{(u,v)\in L_1(\nu)\times L_1(\mu) : u\oplus v\leq c\}$, the couple (u,u^c) is better than (u,v). Performing this trick once again, we see that with $v^c(x) := \inf_{y\in\mathcal{X}}\{c(x,y)-v(y)\},\ x\in\mathcal{X}$, the couple (u^{cc},u^c) is better than (u,u^c) and (u,v). We have obtained the following result.

Lemma 2.3. *Let u and v be functions on $\mathcal{X}$ such that $u(x)+v(y)\leq c(x,y)$ for all x,y. Then, u^c and u^{cc} also satisfy $u^{cc}\geq u, u^c\geq v$ and $u^{cc}(x)+u^c(y)\leq c(x,y)$ for all x,y.*

Iterating the trick of Lemma 2.3 doesn't improve anything.

Remark 2.4. (Measurability of u^c). This issue is often neglected in the literature. The aim of this remark is to indicate a general result which solves this difficult problem. If c is continuous, u^c and u^{cc} are upper semicontinuous, and therefore they are Borel measurable. In the general case where c is lower semicontinuous, it can be shown that some measurable version of u^c exists. More precisely, Beiglböck and Schachermayer have proved recently in [12, Lemmas 3.7, 3.8] that, even if c is only supposed to be Borel measurable, for each probability measure $\mu\in \mathrm{P}(\mathcal{X})$, there exists a $[-\infty,\infty)$-valued Borel measurable function $\tilde u^c$ such that $\tilde u^c\leq u^c$ everywhere and $\tilde u^c = u^c$, μ-almost everywhere. This is precisely what is needed for the purpose of defining the integral $\int u^c\,d\mu$.

Recall that whenever A and B are two vector spaces linked by the duality bracket $\langle a,b\rangle$, the convex conjugate of the function $f: A\to(-\infty,\infty]$ is defined by

$$f^*(b) := \sup_{a\in A}\{\langle a,b\rangle - f(a)\}\in(-\infty,\infty],\quad b\in B.$$

Clearly, the definition of u^c is reminiscent of that of f^*. Indeed, with the quadratic cost function $c_2(x,y)=|y-x|^2/2 = \frac{|x|^2}{2}+\frac{|y|^2}{2} - x\cdot y$ on $\mathbb{R}^k$, one obtains

$$\frac{|\cdot|^2}{2} - u^{c_2} = \left(\frac{|\cdot|^2}{2} - u\right)^*. \tag{18}$$

It is worth recalling basic facts about convex conjugates for we shall use them several times later. Being the supremum of a family of affine continuous functions, f^* is convex and $\sigma(B,A)$-lower semicontinuous. Defining $f^{**}(a) = \sup_{b\in B}\{\langle a,b\rangle - f^*(b)\}\in(-\infty,\infty]$, $a\in A$, one knows that $f^{**}=f$ if and only if f is a lower semicontinuous convex function. It is a trivial remark that

$$\langle a,b\rangle \leq f(a)+f^*(b),\quad \forall a,b. \tag{19}$$

The case of equality (Fenchel's identity) is of special interest, we have

$$\langle a, b\rangle = f(a) + f^*(b) \Leftrightarrow b \in \partial f(a) \Leftrightarrow a \in \partial f^*(b) \tag{20}$$

whenever f is convex and $\sigma(A, B)$-lower semicontinuous. Here, $\partial f(a) := \{b \in B; f(a+h) \geq f(a) + \langle h, b\rangle, \forall h \in A\}$ stands for the subdifferential of f at a.

Metric cost. The cost function to be considered is $c(x, y) = d(x, y)$: a lower semicontinuous metric on $\mathcal{X}$ which might be *different* from the original polish metric on $\mathcal{X}$.

Remark 2.5. In the sequel, the Lipschitz functions are to be considered with respect to the metric cost d and not with respect to the underlying metric on the polish space $\mathcal{X}$ which is here to generate the Borel σ-field, specify the continuous, lower semicontinuous or Borel functions. Indeed, we have in mind to work sometimes with trivial metric costs (weighted Hamming's metrics) which are lower semicontinuous with respect to any reasonable non-trivial metric but generate a too rich Borel σ-field. As a consequence a d-Lipschitz function might not be Borel measurable.

One writes that u is d-Lipschitz(1) to specify that $|u(x) - u(y)| \leq d(x, y)$ for all $x, y \in \mathcal{X}$. Denote $\mathrm{P}_1 := \{\nu \in \mathrm{P}(\mathcal{X}); \int_{\mathcal{X}} d(x_o, x)\, d\nu(x)\}$ where x_o is any fixed element in $\mathcal{X}$. With the triangle inequality, one sees that P_1 doesn't depend on the choice of x_o.
Let us denote the Lipschitz seminorm $\|u\|_{\mathrm{Lip}} := \sup_{x \neq y} \frac{|u(y)-u(x)|}{d(x,y)}$. Its dual norm is for all μ, ν in P_1, $\|\nu - \mu\|^*_{\mathrm{Lip}} = \sup \left\{\int_{\mathcal{X}} u(x)\, [\nu - \mu](dx); u \text{ measurable}, \|u\|_{\mathrm{Lip}} \leq 1\right\}$. As it is assumed that $\mu, \nu \in \mathrm{P}_1$, note that any measurable d-Lipschitz function is integrable with respect to μ and ν.

Theorem 2.6 (Kantorovich-Rubinstein)**.** *For all* $\mu, \nu \in \mathrm{P}_1$, $W_1(\nu, \mu) = \|\nu - \mu\|^*_{\mathrm{Lip}}$.

Proof. For all measurable d-Lipschitz(1) function u and all π such that $\pi_0 = \nu$ and $\pi_1 = \mu$, $\int_{\mathcal{X}} u(x)\, [\nu - \mu](dx) = \int_{\mathcal{X}^2} (u(x) - u(y))\, d\pi(x, y) \leq \int_{\mathcal{X}^2} d(x, y)\, d\pi(x, y)$. Optimizing in u and π one obtains $\|\nu - \mu\|^*_{\mathrm{Lip}} \leq W_1(\nu, \mu)$.
Let us look at the reverse inequality.
Claim. For any function u on $\mathcal{X}$, (i) u^d is d-Lipschitz(1) and (ii) $u^{dd} = -u^d$.
Let us prove (i). Since $y \mapsto d(x, y)$ is d-Lipschitz(1), $y \mapsto u^d(y) = \inf_x\{d(x, y) - u(x)\}$ is also d-Lipschitz(1) as an infinum of d-Lipschitz(1) functions.
Let us prove (ii). Hence for all x, y, $u^d(y) - u^d(x) \leq d(x, y)$. But this implies that for all y, $-u^d(x) \leq d(x, y) - u^d(y)$. Optimizing in y leads to $-u^d(x) \leq u^{dd}(x)$. On the other hand, $u^{dd}(x) = \inf_y\{d(x, y) - u^d(y)\} \leq -u^d(x)$ where the last inequality is obtained by taking $y = x$.

With Theorem 2.2, Lemma 2.3 and the above claim, we obtain that

$$\begin{aligned} W_1(\nu, \mu) &= \sup_{(u,v)}\{\int_{\mathcal{X}} u\, d\nu + \int_{\mathcal{X}} v\, d\mu\} = \sup_u\{\int_{\mathcal{X}} u^{dd}\, d\nu + \int_{\mathcal{X}} u^d\, d\mu\} \\ &\leq \sup\left\{\int_{\mathcal{X}} u\, d[\nu - \mu]; u : \|u\|_{\mathrm{Lip}} \leq 1\right\} = \|\nu - \mu\|^*_{\mathrm{Lip}}. \end{aligned}$$

which completes the proof of the theorem. □

For interesting consequences in probability theory, one can look at Dudley's textbook [40, Chp. 11].

Optimal plans. What about the optimal plans? If one applies *formally* the Karush-Kuhn-Tucker characterization of the saddle point of the Lagrangian function K in the proof of Theorem 2.2, one obtains that $\hat\pi$ is an optimal transport plan if and only if $0 \in \partial_\pi K(\hat\pi,(\hat u,\hat v))$ for some couple of functions $(\hat u,\hat v)$ such that $0 \in \widehat{\partial}_{(u,v)} K(\hat\pi,(\hat u,\hat v))$ where $\partial_\pi K$ stands for the subdifferential of the convex function $\pi \mapsto K(\pi,(\hat u,\hat v))$ and $\widehat{\partial}_{(u,v)} K$ for the superdifferential of the concave function $(u,v) \mapsto K(\hat\pi,(u,v))$. This gives us the system of equations

$$\begin{cases} \hat u \oplus \hat v - c & \in & \partial(\iota_{M_+})(\hat\pi) \\ (\hat\pi_0,\hat\pi_1) & = & (\nu,\mu) \end{cases}$$

where M_+ is the cone of all positive measures on $\mathcal{X}^2$. Such a couple $(\hat u,\hat v)$ is called a dual optimizer. The second equation expresses the marginal constraints of (MK) while by (20) one can recast the first one as the Fenchel identity $\langle \hat u \oplus \hat v - c, \hat\pi\rangle = \iota^*_{M_+}(\hat u \oplus \hat v - c) + \iota_{M_+}(\hat\pi)$. Since for any function h, $\iota^*_{M_+}(h) = \sup_{\pi\in M_+}\langle h,\pi\rangle = \iota_{\{h\le 0\}}$, one sees that $\langle \hat u \oplus \hat v - c, \hat\pi\rangle = 0$ with $\hat u \oplus \hat v - c \le 0$ and $\hat\pi \ge 0$ which is equivalent to $\hat\pi \ge 0$, $\hat u \oplus \hat v \le c$ everywhere and $\hat u \oplus \hat v = c$, $\hat\pi$-almost everywhere. As $\hat\pi_0 = \nu$ has a unit mass, so has the positive measure $\hat\pi$: it is a probability measure. These formal considerations should prepare the reader to trust the subsequent rigorous statement.

Theorem 2.7. *Assume that $\mathcal{T}_c(\nu,\mu) < \infty$. Any $\pi \in \mathrm{P}(\mathcal{X}^2)$ with the prescribed marginals $\pi_0 = \nu$ and $\pi_1 = \mu$ is an optimal plan if and only if there exist two measurable functions $u,v : \mathcal{X} \to [-\infty,\infty)$ such that*

$$\begin{cases} u \oplus v & \le & c, & \textit{everywhere} \\ u \oplus v & = & c, & \pi\textit{-almost everywhere.} \end{cases}$$

This theorem can be found in [104, Thm. 5.10] with a proof which has almost nothing in common with the saddle-point strategy that has been described above.

An important instance of this result is the special case of the quadratic cost.

Corollary 2.8. *Let us consider the quadratic cost $c_2(x,y) = |y-x|^2/2$ on $\mathcal{X} = \mathbb{R}^k$ and take two probability measures ν and μ in $\mathrm{P}_2(\mathcal{X})$.*

(a) There exists an optimal transport plan.
(b) Any $\pi \in \mathrm{P}(\mathcal{X}^2)$ is optimal if and only if there exists a convex lower semicontinuous function $\phi : \mathcal{X} \to (-\infty,\infty]$ such that the Fenchel identity $\phi(x) + \phi^*(y) = x\cdot y$ holds true π-almost everywhere.

Proof. Proof of (a). Since $|y-x|^2/2 \le |x|^2 + |y|^2$ and $\nu,\mu \in \mathrm{P}_2$, any $\pi \in \mathrm{P}(\mathcal{X}^2)$ such that $\pi_0 = \nu$ and $\pi_1 = \mu$ satisfies $\int_{\mathcal{X}^2} c\,d\pi \le \int_{\mathcal{X}} |x|^2\,d\nu(x) + \int_{\mathcal{X}} |y|^2\,d\mu(y) < \infty$. Therefore $\mathcal{T}(\nu,\mu) = W_2^2(\nu,\mu) < \infty$, and one concludes with Proposition 2.1.

Proof of (b). In view of Lemma 2.3, one sees that an optimal dual optimizer is necessarily of the form (u^{cc},u^c). Theorem 2.7 tells us that π is optimal if and only there exists some function u such that $u^{cc} \oplus u^c = c$, π-almost everywhere. With (18), by considering the functions $\phi(x) = |x|^2/2 - u^{c_2c_2}(x)$ and $\psi(y) = |y|^2/2 - u^{c_2}(y)$, one also obtains $\phi = \psi^*$ and

$\psi = \phi^*$, which means that ϕ and ψ are convex conjugate to each other and in particular that ϕ is convex and lower semicontinuous. □

By (20), another equivalent statement for the Fenchel identity $\phi(x) + \phi^*(y) = x{\cdot}y$ is

$$y \in \partial\phi(x). \tag{21}$$

In the special case of the real line $\mathcal{X} = \mathbb{R}$, a popular coupling of ν and $\mu = T_\#\nu$ is given by the so-called *monotone rearrangement.* It is defined by

$$y = T(x) := F_\mu^{-1} \circ F_\nu(x), \quad x \in \mathbb{R}, \tag{22}$$

where $F_\nu(x) = \nu((-\infty, x])$, $F_\mu(y) = \mu((-\infty, y])$ are the distribution functions of ν and μ, and $F_\mu^{-1}(u) = \inf\{y \in \mathbb{R}; F_\mu(y) > u\}$, $u \in [0, 1]$ is the generalized inverse of F_μ. When $\mathcal{X} = \mathbb{R}$, the identity (21) simply states that (x, y) belongs to the graph of an increasing function. Of course, this is the case of (22). Hence Corollary 2.8 tells us that the monotone rearrangement is an optimal transport map for the quadratic cost.

Let us go back to $\mathcal{X} = \mathbb{R}^k$. If ϕ is Gâteaux differentiable at x, then $\partial\phi(x)$ is restricted to a single element: the gradient $\nabla\phi(x)$, and (21) simply becomes $y = \nabla\phi(x)$. Hence, if ϕ were differentiable everywhere, condition (b) of Corollary 2.8 would be $y = \nabla\phi(x)$, π-almost everywhere. But this is too much demanding. Nevertheless, Rademacher's theorem states that a convex function on $\mathbb{R}^k$ is differentiable Lebesgue almost everywhere on its effective domain. This allows to derive the following improvement.

Theorem 2.9 (Quadratic cost on $\mathcal{X} = \mathbb{R}^k$)**.** *Let us consider the quadratic cost $c_2(x, y) = |y-x|^2/2$ on $\mathcal{X} = \mathbb{R}^k$ and take two probability measures ν and μ in $\mathrm{P}_2(\mathcal{X})$ which are absolutely continuous. Then, there exists a unique optimal plan. Moreover, $\pi \in \mathrm{P}(\mathcal{X}^2)$ is optimal if and only if $\pi_0 = \nu, \pi_1 = \mu$ and there exists a convex function ϕ such that*

$$\left\{ \begin{array}{rcl} y &=& \nabla\phi(x) \\ x &=& \nabla\phi^*(y) \end{array} \right. , \quad \pi\textit{-almost everywhere.}$$

Proof. It follows from Rademacher's theorem that, if ν is an absolutely continuous measure, ϕ is differentiable ν-almost everywhere. This and Corollary 2.8 prove the statement about the characterization of the optimal plans. Note that the quadratic transport is symmetric with respect to x an y, so that one obtains the same conclusion if μ is absolutely continuous; namely $x = \nabla\phi^*(y)$, π-almost everywhere, see (20).

We have just proved that, under our assumptions, an optimal plan is concentrated on a functional graph. The uniqueness of the optimal plan follows directly from this. Indeed, if one has two optimal plans π^0 and π^1, by convexity their half sum $\pi^{1/2}$ is still optimal. But for $\pi^{1/2}$ to be concentrated on a *functional* graph, it is necessary that π^0 and π^1 share the same graph. □

For more details, one can have a look at [103, Thm. 2.12]. The existence result has been obtained by Knott and Smith [63], while the uniqueness is due to Brenier [23] and McCann [84]. The application $y = \nabla\phi(x)$ is often called the *Brenier map* pushing forward ν to μ.

3. Dual equalities and inequalities

In this section, we present Bobkov and Götze dual approach of transport-entropy inequalities [17]. More precisely, we are going to take advantage of variational formulas for the optimal transport cost and for the relative entropy to give another formulation of transport-entropy inequalities. The relevant variational formula for the transport cost is given by the Kantorovich dual equality at Theorem 2.2

$$\mathcal{T}_c(\nu,\mu) = \sup\left\{\int u\,d\nu + \int v\,d\mu; u,v\in\mathcal{C}_b(\mathcal{X}), u\oplus v\le c\right\}. \tag{23}$$

On the other hand, the relative entropy admits the following variational representations. For all $\nu\in \mathrm{P}(\mathcal{X})$,

$$\begin{aligned} H(\nu|\mu) &= \sup\left\{\int u\,d\nu - \log\int e^u\,d\mu; u\in\mathcal{C}_b(\mathcal{X})\right\}. \\ &= \sup\left\{\int u\,d\nu - \log\int e^u\,d\mu; u\in\mathcal{B}_b(\mathcal{X})\right\} \end{aligned} \tag{24}$$

and for all $\nu\in\mathrm{P}(\mathcal{X})$ such that $\nu\ll\mu$,

$$H(\nu|\mu) = \sup\left\{\int u\,d\nu - \log\int e^u\,d\mu; u:\text{measurable}, \int e^u\,d\mu<\infty, \int u_-\,d\nu<\infty\right\} \tag{25}$$

where $u_- = (-u)\vee 0$ and $\int u\,d\nu\in(-\infty,\infty]$ is well-defined for all u such that $\int u_-\,d\nu<\infty$. The identities (24) are well-known, but the proof of (25) is more confidential. This is the reason why we give their detailed proofs at the Appendix, Proposition B.1.

As regards Remark 1.2, a sufficient condition for μ to satisfy $H(\nu|\mu)=\infty$ whenever $\nu\notin\mathrm{P}_p$ is

$$\int e^{s_o d^p(x_o,x)}\,d\mu(x)<\infty \tag{26}$$

for some $x_o\in\mathcal{X}$ and $s_o>0$. Indeed, by (25), for all $\nu\in\mathrm{P}(\mathcal{X})$, $s_o\int d^p(x_o,x)\,d\nu(x)\le H(\nu|\mu)+\log\int e^{s_o d^p(x_o,x)}\,d\mu(x)$. On the other hand, Proposition 6.1 below tells us that (26) is also a necessary condition.

Since $u\mapsto\Lambda(u):=\log\int e^u\,d\mu$ is convex (use Hölder inequality to show it) and lower semicontinuous on $\mathcal{C}_b(\mathcal{X})$ (resp. $\mathcal{B}_b(\mathcal{X})$) (use Fatou's lemma), one observes that $H(\,\cdot\,|\mu)$ (more precisely its extension to the vector space of signed bounded measures which achieves the value $+\infty$ outside $\mathrm{P}(\mathcal{X})$) and Λ are convex conjugate to each other:

$$\begin{cases} H(\,\cdot\,|\mu)=\Lambda^*,\\ \Lambda = H(\,\cdot\,|\mu)^*.\end{cases}$$

It appears that $\mathcal{T}_c(\,\cdot\,,\mu)$ and $H(\,\cdot\,|\mu)$ both can be written as convex conjugates of functions on a class of functions on $\mathcal{X}$. This structure will be exploited in a moment to give a dual formulation of inequalities $\alpha(\mathcal{T}_c)\le H$, for α belonging to the following class.

Definition 3.1 (of $\mathcal{A}$). *The class $\mathcal{A}$ consists of all the functions α on $[0,\infty)$ which are convex, increasing with $\alpha(0)=0$.*

The convex conjugate of a function $\alpha \in \mathcal{A}$ is replaced by the monotone conjugate $\alpha^{\circledast}$ defined by

$$\alpha^{\circledast}(s) = \sup_{r\geq 0}\{sr - \alpha(r)\}, s \geq 0$$

where the supremum is taken on $r \geq 0$ instead of $r \in \mathbb{R}$.

Theorem 3.2. *Let c be a lower semicontinuous cost function, $\alpha \in \mathcal{A}$ and $\mu \in \mathrm{P}(\mathcal{X})$; the following propositions are equivalent.*

(1) The probability measure μ verifies the inequality $\alpha(\mathcal{T}_c) \leq H$.
(2) For all $u, v \in \mathcal{C}_b(\mathcal{X})$, such that $u \oplus v \leq c$,

$$\int e^{su}\, d\mu \leq e^{-s\int v\, d\mu + \alpha^{\circledast}(s)},\ s \geq 0.$$

Moreover, the same result holds with $\mathcal{B}_b(\mathcal{X})$ instead of $\mathcal{C}_b(\mathcal{X})$.

A variant of this result can be found in the authors' paper [53] and in Villani's textbook [104, Thm. 5.26]. It extends the dual characterization of transport inequalities $\mathbf{T}_1$ and $\mathbf{T}_2$ obtained by Bobkov and Götze in [17].

Proof. First we extend α to the whole real line by defining $\alpha(r) = 0$, for all $r \leq 0$. Using Kantorovich dual equality and the fact that α is continuous and increasing on $\mathbb{R}$, we see that the inequality $\alpha(\mathcal{T}_c) \leq H$ holds if and only if for all $u, v \in \mathcal{C}_b(\mathcal{X})$, such that $u \oplus v \leq c$, one has

$$\alpha\left(\int u\, d\nu + \int v\, d\mu\right) \leq H(\nu|\mu),\ \nu \in \mathrm{P}(\mathcal{X}).$$

Since α is convex and continuous on $\mathbb{R}$, it satisfies $\alpha(r) = \sup_s\{sr - \alpha^*(s)\}$. So the preceding condition is equivalent to the following one

$$s\int u\, d\nu - H(\nu|\mu) \leq -s\int v\, d\mu + \alpha^*(s),\ \nu \in \mathrm{P}(\mathcal{X}), s \in \mathbb{R}, u \oplus v \leq c.$$

Since $H(\,\cdot\,|\mu)^* = \Lambda$, optimizing over $\nu \in \mathrm{P}(\mathcal{X})$, we arrive at

$$\log\int e^{su}\, d\mu \leq -s\int v\, d\mu + \alpha^*(s),\ s \in \mathbb{R}, u \oplus v \leq c.$$

Since $\alpha^*(s) = +\infty$ when $s < 0$ and $\alpha^*(s) = \alpha^{\circledast}(s)$ when $s \geq 0$, this completes the proof. $\square$

Let us define, for all $f, g \in \mathcal{B}_b(\mathcal{X})$,

$$P_c f(y) = \sup_{x\in\mathcal{X}}\{f(x) - c(x,y)\},\ y \in \mathcal{X},$$

and

$$Q_c g(x) = \inf_{y\in\mathcal{X}}\{g(y) + c(x,y)\},\ x \in \mathcal{X}.$$

For a given function $f : \mathcal{X} \to \mathbb{R}$, $P_c f$ is the best function $g : \mathcal{X} \to \mathbb{R}$ (the smallest) such that $f(x) - g(y) \leq c(x,y)$, for all $x, y \in \mathcal{X}$. And for a given function $g : \mathcal{X} \to \mathbb{R}$, $Q_c g$ is the best function $f : \mathcal{X} \to \mathbb{R}$ (the biggest) such that $f(x) - g(y) \leq c(x,y)$, for all $x, y \in \mathcal{X}$.

The following immediate corollary gives optimized forms of the dual condition (2) stated in Theorem 3.2.

Corollary 3.3. *Let c be a lower semicontinuous cost function, $\alpha \in \mathcal{A}$ and $\mu \in \mathrm{P}(\mathcal{X})$. The following propositions are equivalent.*

1. *The probability measure μ verifies the inequality $\alpha(\mathcal{T}_c) \leq H$.*
2. *For all $f \in \mathcal{C}_b(\mathcal{X})$,*
$$\int e^{sf}\,d\mu \leq e^{s\int P_c f\,d\mu + \alpha^{\circledast}(s)},\ s \geq 0.$$
3. *For all $g \in \mathcal{C}_b(\mathcal{X})$,*
$$\int e^{sQ_c g}\,d\mu \leq e^{s\int g\,d\mu + \alpha^{\circledast}(s)},\ s \geq 0.$$

Moreover, the same result holds true with $\mathcal{B}_b(\mathcal{X})$ instead of $\mathcal{C}_b(\mathcal{X})$.

When the cost function is a lower semicontinuous distance, we have the following.

Corollary 3.4. *Let d be a lower semicontinuous distance, $\alpha \in \mathcal{A}$ and $\mu \in \mathrm{P}(\mathcal{X})$. The following propositions are equivalent.*

1. *The probability measure μ verifies the inequality $\alpha(\mathcal{T}_d) \leq H$.*
2. *For all 1-Lipschitz function f,*
$$\int e^{sf}\,d\mu \leq e^{s\int f\,d\mu + \alpha^{\circledast}(s)},\ s \geq 0.$$

Corollary 3.3 enables us to give an alternative proof of the tensorization property given at Proposition 1.9.

Proof of Proposition 1.9. For the sake of simplicity, let us explain the proof for $n = 2$. The general case is done by induction (see for instance [53, Theorem 5]). Let us consider, for all $f \in \mathcal{B}_b(\mathcal{X})$
$$Q_c f(x) = \inf_{y \in \mathcal{X}} \{f(y) + c(x, y)\},\ x \in \mathcal{X}$$
and for all $f \in \mathcal{B}_b(\mathcal{X} \times \mathcal{X})$,
$$Q_c^{(2)} f(x) = \inf_{y \in \mathcal{X} \times \mathcal{X}} \{f(y_1, y_2) + c(x_1, y_1) + c(x_2, y_2)\},\ x \in \mathcal{X} \times \mathcal{X}.$$
According to the dual formulation of transport-entropy inequalities (Corollary 3.3 (2)), μ verifies the inequality $\alpha(\mathcal{T}_c) \leq H$ if and only if
$$\int e^{sQ_c f}\,d\mu \leq e^{s\int f\,d\mu + \alpha^*(s)},\ s \geq 0 \tag{27}$$
for all $f \in \mathcal{B}_b(\mathcal{X})$. On the other hand, μ^2 verifies the inequality $2\alpha\left(\frac{\mathcal{T}_{c^{\oplus 2}}}{2}\right) \leq H$ if and only if
$$\int e^{sQ_c^{(2)} f}\,d\mu^2 \leq e^{s\int f\,d\mu^2 + 2\alpha^*(s)},\ s \geq 0,$$

holds for all $f \in \mathcal{B}_b(\mathcal{X}\times\mathcal{X})$. Let $f \in \mathcal{B}_b(\mathcal{X}\times\mathcal{X})$,

$$\begin{aligned} Q_c^{(2)} f(x_1,x_2) &= \inf_{y_1,y_2\in\mathcal{X}} \{f(y_1,y_2)+c(x_1,y_1)+c(x_2,y_2)\} \\ &= \inf_{y_1\in\mathcal{X}} \left\{ \inf_{y_2\in\mathcal{X}} \{f(y_1,y_2)+c(x_2,y_2)\} + c(x_1,y_1) \right\} \\ &= \inf_{y_1\in\mathcal{X}} \{Q_c(f_{y_1})(x_2)+c(x_1,y_1)\} \end{aligned}$$

where for all $y_1 \in \mathcal{X}$, $f_{y_1}(y_2) = f(y_1,y_2)$, $y_2 \in \mathcal{X}$.

So, applying (27) gives

$$\begin{aligned} \int_{\mathcal{X}\times\mathcal{X}} e^{sQ_c^{(2)} f}\, d\mu^2 &= \int \left(\int e^{s \inf_{y_1\in\mathcal{X}}\{Q_c(f_{y_1})(x_2)+c(y_1,x_1)\}}\, d\mu(x_1) \right) d\mu(x_2) \\ &\le e^{\alpha^{\circledast}(s)} \int e^{s\int Q_c(f_{x_1})(x_2)\, d\mu(x_1)}\, d\mu(x_2). \end{aligned}$$

But,

$$\int Q_c(f_{x_1})(x_2)\, d\mu(x_1) = \int \inf_{y_1\in\mathcal{X}} \{f(x_1,y_1)+c(x_2,y_1)\}\, d\mu(x_1) \le Q_c(\bar f)(x_2),$$

with $\bar f(y_1) = \int f(x_1,y_1)\, d\mu(x_1)$.

Applying (27) again yields

$$\int e^{s\int Q_c(f_{x_1})(x_2)\, d\mu(x_1)}\, d\mu(x_2) \le \int e^{sQ_c(\bar f)(x_2)}\, d\mu(x_2) \le e^{\alpha^{\circledast}(s) + s\int \bar f(x_2)\, d\mu(x_2)}.$$

Since $\int \bar f(x_2)\, d\mu(x_2) = \int f\, d\mu^2$, this completes the proof. □

To conclude this section, let us put the preceding results in an abstract general setting. Our motivation to do that is to consider transport inequalities involving other functionals J than the entropy.

Consider two convex functions on some vector space $\mathcal{U}$ of measurable functions on $\mathcal{X}$, $\Theta : \mathcal{U} \to (-\infty,\infty]$ and $\Upsilon : \mathcal{U} \to (-\infty,\infty]$. Their convex conjugates are defined for all ν in the space $\mathrm{M}_{\mathcal{U}}$ of all measures on $\mathcal{X}$ such that $\int |u|\, d\nu < \infty$, for all $u \in \mathcal{U}$ by

$$\left\{ \begin{array}{lcl} T(\nu) &=& \sup_{u\in\mathcal{U}} \left\{ \int u\, d\nu - \Theta(u) \right\} \\ J(\nu) &=& \sup_{u\in\mathcal{U}} \left\{ \int u\, d\nu - \Upsilon(u) \right\} \end{array} \right.$$

Without loss of generality, one assumes that Υ is a convex and $\sigma(\mathcal{U}, \mathrm{M}_{\mathcal{U}})$-lower semicontinuous function, so that J and Υ are convex conjugate to each other. It is assumed that $\mathcal{U}$ contains the constant functions, $\Theta(0) = \Upsilon(0) = 0$, $\Theta(u + a\mathbf{1}) = \Theta(u) + a$ and $\Upsilon(u + a\mathbf{1}) = \Upsilon(u) + a$ for all real a and all $u \in \mathcal{U}$ and that Θ and Υ are increasing. This implies that T and J are $[0,\infty]$-valued with their effective domain in $\mathrm{P}_{\mathcal{U}} := \{\nu \in \mathrm{P}(\mathcal{X}); \int |u|\, d\nu < \infty, \forall u \in \mathcal{U}\}$. In this setting, we have the following theorem whose proof is a straightforward adaptation of the proof of Theorem 3.2.

Theorem 3.5. *Let $\alpha \in \mathcal{A}$ and $\mathcal{U}, T, J$ as above. For all $u \in \mathcal{U}$ and $s \ge 0$ define $\Upsilon_u(s) := \Upsilon(su) - s\Theta(u)$. The following statements are equivalent.*

(a) For all $\nu \in \mathrm{P}_{\mathcal{U}}$, $\alpha(T(\nu)) \le J(\nu)$.
(b) For all $u \in \mathcal{U}$ and $s \ge 0$, $\Upsilon_u(s) \le \alpha^{\circledast}(s)$.

This general result will be used in Section 10 devoted to transport-information inequalities, where the functional J is the Fisher information.

4. Concentration for product probability measures

Transport-entropy inequalities are intrinsically linked to the concentration of measure phenomenon for product probability measures. This relation was first discovered by K. Marton in [76]. Informally, a concentration of measure inequality quantifies how fast the probability goes to 1 when a set A is enlarged.

Definition 4.1. *Let $\mathcal{X}$ be a Hausdorff topological space and let $\mathcal{G}$ be its Borel σ-field. An enlargement function is a function $\mathrm{enl} : \mathcal{G} \times [0, \infty) \to \mathcal{G}$ such that*

- *For all $A \in \mathcal{G}$, $r \mapsto \mathrm{enl}(A, r)$ is increasing on $[0, \infty)$ (for the set inclusion).*
- *For all $r \geq 0$, $A \mapsto \mathrm{enl}(A, r)$ is increasing (for the set inclusion).*
- *For all $A \in \mathcal{G}$, $A \subset \mathrm{enl}(A, 0)$.*
- *For all $A \in \mathcal{G}$, $\cup_{r \geq 0} \mathrm{enl}(A, r) = \mathcal{X}$.*

If μ is a probability measure on $\mathcal{X}$, one says that it verifies a concentration of measure inequality if there is a function $\beta : [0, \infty) \to [0, \infty)$ such that $\beta(r) \to 0$ when $r \to +\infty$ and such that for all $A \in \mathcal{G}$ with $\mu(A) \geq 1/2$ the following inequality holds

$$\mu(\mathrm{enl}(A, r)) \geq 1 - \beta(r),\ r \geq 0.$$

There are many ways of enlarging sets. If $(\mathcal{X}, d)$ is a metric space, a classical way is to consider the r-neighborhood of A defined by

$$A^r = \{x \in \mathcal{X}; d(x, A) \leq r\},\ r \geq 0,$$

where the distance of x from A is defined by $d(x, A) = \inf_{y \in A} d(x, y)$.

Let us recall the statement of Marton's concentration theorem whose proof was given at Theorem 1.7.

Theorem 4.2 (Marton's concentration theorem)**.** *Suppose that μ verifies the inequality $\alpha\left(\mathcal{T}_d(\nu, \mu)\right) \leq H(\nu|\mu)$, for all $\nu \in \mathrm{P}(\mathcal{X})$. Then for all $A \subset \mathcal{X}$, with $\mu(A) \geq 1/2$, the following holds*

$$\mu(A^r) \geq 1 - e^{-\alpha(r - r_o)},\ r \geq r_o := \alpha^{-1}(\log 2).$$

We already stated at Proposition 1.9 an important tensorization result. Its statement is recalled below at Proposition 4.3.

Proposition 4.3. *Let c be a lower semicontinuous cost function on $\mathcal{X}$ and $\alpha \in \mathcal{A}$ (see Definition 3.1). Suppose that a probability measure μ verifies the transport-entropy inequality $\alpha(\mathcal{T}_c) \leq H$ on $\mathcal{X}$, then μ^n, $n \geq 1$ verifies the inequality*

$$n\alpha\left(\frac{\mathcal{T}_{c^{\oplus n}}(\nu, \mu^n)}{n}\right) \leq H(\nu|\mu^n),\ \nu \in \mathrm{P}(\mathcal{X}^n),$$

where $c^{\oplus n}(x, y) = \sum_{i=1}^n c(x_i, y_i)$.

Other forms of non-product tensorizations have been studied (see [77, 78, 79, 80], [94] or [36, 112]) in the context of Markov chains or Gibbs measures (see Section 11).

Let us recall a first easy consequence of this tensorization property.

Corollary 4.4. *Suppose that a probability measure μ on $\mathcal{X}$ verifies the inequality $\mathbf{T}_2(C)$, then μ^n verifies the inequality $\mathbf{T}_2(C)$ on $\mathcal{X}^n$, for all positive integer n. In particular, the following dimension-free Gaussian concentration property holds: for all positive integer n and for all $A \subset \mathcal{X}^n$ with $\mu^n(A) \geq 1/2$,*

$$\mu^n(A^r) \geq 1 - \exp(-\frac{1}{C}(r-r_o)^2),\ r \geq r_o := \sqrt{\log(2)},$$

where $A^r = \{x \in \mathcal{X}^n : d_2(x, A) \leq r\}$ and $d_2(x,y) = \left[\sum_{i=1}^n d(x_i, y_i)^2\right]^{1/2}$.

Equivalently, when μ verifies $\mathbf{T}_2(C)$,

$$\mu^n(f > m_f + r + r_o) \leq e^{-r^2/C},\ r \geq 0,$$

for all positive integer n and all 1-Lipschitz function $f : \mathcal{X}^n \to \mathbb{R}$ with median m_f.

Proof. According to the tensorization property, μ^n verifies the inequality $\mathbf{T}_2(C)$ on $\mathcal{X}^n$ equipped with the metric d_2 defined above. It follows from Jensen inequality that μ^n also verifies the inequality $(\mathcal{T}_{d_2})^2 \leq CH$. Theorem 4.2 and Proposition 1.6 then give the conclusion. □

Remark 4.5. So, as was already emphasized at Section 1, when μ verifies $\mathbf{T}_2$, it verifies a dimension-free Gaussian concentration inequality. Dimension-free means that the concentration inequality does not depend explicitly on n. This independence on n corresponds to an optimal behavior. Indeed, the constants in concentration inequalities cannot improve when n grows.

More generally, the following proposition explains what kind of concentration inequalities can be derived from a transport-entropy inequality.

Proposition 4.6. *Let μ be a probability measure on $\mathcal{X}$ satisfying the inequality $\alpha(\mathcal{T}_{\theta(d)}) \leq H$, where the function θ is convex and such that $\sup_{t>0} \theta(2t)/\theta(t) < +\infty$.*

Then for all $\lambda \in (0,1)$, there is some constant $a_\lambda > 0$ such that

$$\inf_\pi \int_{\mathcal{X}\times\mathcal{X}} \theta\left(\frac{d(x,y)}{\lambda}\right) d\pi(x,y) \leq a_\lambda \alpha^{-1}\left(H(\nu|\mu)\right),$$

where the infimum is over the set of couplings of ν and μ.

Furthermore, the product probability measure μ^n on $\mathcal{X}^n$, $n \geq 1$ satisfies the following concentration property.
For all $A \subset \mathcal{X}^n$ such that $\mu^n(A) \geq 1/2$,

$$\mu^n(\mathrm{enl}_\theta(A,r)) \geq 1 - \exp\left(-n\alpha\left(\frac{r - r_o^n(\lambda)}{n\lambda a_\lambda}\right)\right),\ r \geq r_o^n(\lambda),\ \lambda \in (0,1),$$

where

$$\mathrm{enl}_\theta(A,r) = \left\{ x \in \mathcal{X}^n; \inf_{y\in A} \sum_{i=1}^n \theta(d(x_i,y_i)) \le r \right\}, \tag{28}$$

and

$$r_o^n(\lambda) = (1-\lambda) a_{1-\lambda} n \alpha^{-1} \left(\frac{\log 2}{n} \right).$$

The proof can be easily adapted from [52, Proposition 3.4].

Remark 4.7.

(1) It is not difficult to check that $a_\lambda \to +\infty$ when $\lambda \to 0$.
(2) If α is linear, the right-hand side does not depend explicitly on n. In this case, the concentration inequality is dimension-free.
(3) For example, if μ verifies $\mathbb{T}_2(C)$ (which corresponds to $\alpha(t) = t/C$ and $\theta(t) = t^2$), then one can take $a_\lambda = \frac{1}{\lambda^2}$. Defining as before $A^r = \{x \in \mathcal{X}^n; \inf_{y\in A} d_2(x,y) \le r\}$ where $d_2(x,y) = \left(\sum_{i=1}^n d(x_i,y_i)^2\right)^{1/2}$ and optimizing over $\lambda \in (0,1)$, yields

$$\mu^n(A^r) \ge 1 - e^{-\frac{1}{C}(r-r_o)^2},\ r \ge r_o = \sqrt{C \log 2}.$$

So we recover the dimension-free Gaussian inequality of Corollary 4.4.

As we said above there are many ways of enlarging sets, and consequently there many ways to describe the concentration of measure phenomenon. In a series of papers [99, 100, 101] Talagrand has deeply investigated the concentration properties of product of probability measures. In particular, he has proposed different families of enlargements which do not enter into the framework of (28). In particular he has obtained various concentration inequalities based on convex hull approximation or q-points control, which have found numerous applications (see [68] or [99]): deviation bounds for empirical processes, combinatoric, percolation, probability on graphs, etc.

The general framework is the following: One considers a product space $\mathcal{X}^n$. For all $A \subset \mathcal{X}^n$, a function $\varphi_A : \mathcal{X}^n \to [0,\infty)$ measures how far is the point $x \in \mathcal{X}^n$ from the set A. The enlargement of A is then defined by

$$\mathrm{enl}(A,r) = \{x \in \mathcal{X}^n; \varphi_A(x) \le r\},\ r \ge 0.$$

Convex hull approximation. Define on $\mathcal{X}^n$ the following weighted Hamming metrics:

$$d_a(x,y) = \sum_{i=1}^n a_i \mathbf{1}_{x_i \ne y_i},\ x,y \in \mathcal{X}^n,$$

where $a \in ([0,\infty))^n$ is such that $|a| = \sqrt{a_1^2 + \cdots + a_n^2} = 1$. The function φ_A is defined as follows:

$$\varphi_A(x) = \sup_{|a|=1} d(x,A),\ x \in \mathcal{X}^n.$$

An alternative definition for φ_A is the following. For all $x \in \mathcal{X}^n$, consider the set

$$U_A(x) = \{(\mathbf{1}_{x_1 \ne y_1}, \ldots, \mathbf{1}_{x_n \ne y_n}); y \in A\},$$

and let $V_A(x)$ be the convex hull of $U_A(x)$. Then it can be shown that

$$\varphi_A(x) = d(0, V_A(x)),$$

where d is the Euclidean distance in $\mathbb{R}^n$.

A basic result related to convex hull approximation is the following theorem by Talagrand ([99, Theorem 4.1.1]).

Theorem 4.8. *For every product probability measure P on $\mathcal{X}^n$, and every $A \subset \mathcal{X}^n$,*

$$\int e^{\varphi_A^2(x)/4} \, dP(x) \leq \frac{1}{P(A)}.$$

In particular,

$$P(\mathrm{enl}(A, r)) \geq 1 - \frac{1}{P(A)} e^{-r^2/4}, \ r \geq 0.$$

This result admits many refinements (see [99]).

In [78], Marton developed transport-entropy inequalities to recover some of Talagrand's results on convex hull approximation. To catch the Gaussian type concentration inequality stated in the above theorem, a natural idea would be to consider a $\mathbf{T}_2$ inequality with respect to the Hamming metric. In fact, it can be shown easily that such an inequality cannot hold. Let us introduce a weaker form of the transport-entropy inequality $\mathbf{T}_2$. Let $\mathcal{X}$ be some polish space, and d a metric on $\mathcal{X}$; define

$$\widetilde{\mathcal{T}}_2(Q, R) = \inf_\pi \int_{\mathcal{X}} \left(\int_{\mathcal{X}} d(x, y) \, d\pi^y(x) \right)^2 \, dR(y), \ Q, R \in \mathrm{P}(\mathcal{X}),$$

where the infimum runs over all the coupling π of Q and R and where $\mathcal{X} \to \mathrm{P}(\mathcal{X}) : y \mapsto \pi^y$ is a regular disintegration of π given y:

$$\int_{\mathcal{X} \times \mathcal{X}} f(x, y) \, d\pi(x, y) = \int_{\mathcal{X}} \left(\int_{\mathcal{X}} f(x, y) d\pi^y(x) \right) \, dR(y),$$

for all bounded measurable $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

According to Jensen inequality,

$$\mathcal{T}_1(Q, R)^2 \leq \widetilde{\mathcal{T}}_2(Q, R) \leq \mathcal{T}_2(Q, R).$$

One will says that $\mu \in \mathrm{P}(\mathcal{X})$ verifies the inequality $\widetilde{\mathbf{T}}_2(C)$ if

$$\widetilde{\mathcal{T}}_2(Q, R) \leq CH(Q|P) + CH(R|P),$$

for all probability measures Q, R on $\mathcal{X}$.

The following theorem is due to Marton.

Theorem 4.9. *Every probability measure P on $\mathcal{X}$, verifies the inequality $\widetilde{\mathbf{T}}_2(4)$ with respect to the Hamming metric. In other words,*

$$\widetilde{\mathcal{T}}_2(Q, R) = \inf_\pi \int \pi^y\{x; x_i \neq y_i\}^2 \, dR(y) \leq 4H(Q|P) + 4H(R|P),$$

for all probability measures Q, R on $\mathcal{X}$.

A proof of this result can be found in [78] or in [68].

Like $\mathbf{T}_2$, the inequality $\widetilde{\mathbf{T}}_2$ admits a dimension-free tensorization property. A variant of Marton's argument can be used to derive dimension-free concentration and recover Talagrand's concentration results for the convex hull approximation distance. We refer to [78] and [68, Chp 6] for more explanations and proofs.

Control by q-points. Here the point of view is quite different: $q \geq 2$ is a fixed integer and a point $x \in \mathcal{X}^n$ will be close from A if it has many coordinates in common with q vectors of A. More generally, consider $A_1, \dots, A_q \subset \mathcal{X}^n$; the function $\varphi_{A_1,\dots,A_q}$ is defined as follows:

$$\varphi_{A_1,\dots,A_q}(x) = \inf_{y^1 \in A_1, \dots, y^q \in A_q} \mathrm{Card}\left\{i; x_i \notin \{y_i^1, \dots, y_i^q\}\right\}.$$

Talagrand's has obtained the following result (see [99, Theorem 3.1.1] for a proof and further refinements).

Theorem 4.10. *For every product probability measure P on $\mathcal{X}^n$, and every family $A_1, \dots, A_q \subset \mathcal{X}^n$, $q \geq 2$, the following inequality holds*

$$\int q^{\varphi_{A_1,\dots,A_q}(x)}\, dP(x) \leq \frac{1}{P(A_1)\cdots P(A_q)}.$$

In particular, defining $\mathrm{enl}(A, r) = \{x \in \mathcal{X}^n; \varphi_{A,\dots,A}(x) \leq r\}$*, one gets*

$$P(\mathrm{enl}(A, r)) \geq 1 - \frac{1}{q^r P(A)^r}, \quad r \geq 0.$$

In [32], Dembo has obtained transport-entropy inequalities giving back Talagrand's results for q-points control. See also [33], for related inequalities.

5. Transport-entropy inequalities and large deviations

In [53], Gozlan and Léonard have proposed an interpretation of transport-entropy inequalities in terms of large deviations theory. To expose this point of view, let us introduce some notations. Suppose that $(X_n)_{n\geq 1}$ is a sequence of independent and identically distributed $\mathcal{X}$ valued random variables with common law μ. Define their empirical measure

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i},$$

where δ_a stands for the Dirac mass at point $a \in \mathcal{X}$. Let $\mathcal{C}_b(\mathcal{X})$ be the set of all bounded continuous functions on $\mathcal{X}$. The set of all Borel probability measures on $\mathcal{X}$, denoted by $\mathrm{P}(\mathcal{X})$, will be endowed with the weak topology, that is the smallest topology with respect to which all functionals $\nu \mapsto \int_{\mathcal{X}} \varphi\, d\nu$ with $\varphi \in \mathcal{C}_b(\mathcal{X})$ are continuous. If $B \subset \mathcal{X}$, let us denote $H(B|\mu) = \inf\{H(\nu|\mu); \nu \in B\}$. According to a famous theorem of large deviations theory (Sanov's theorem), the relative entropy functional governs the asymptotic behavior of $\mathbb{P}(L_n \in A)$, $A \subset \mathcal{X}$ when n goes to ∞.

Theorem 5.1 (Sanov's theorem)**.** *For all $A \subset \mathrm{P}(\mathcal{X})$ measurable with respect to the Borel σ-field,*

$$-H(\mathrm{int}(A)|\mu) \le \liminf_{n\to+\infty} \frac{1}{n} \log \mathbb{P}\left(L_n \in A\right) \le \limsup_{n\to+\infty} \frac{1}{n} \log \mathbb{P}\left(L_n \in A\right) \le -H(\mathrm{cl}(A)|\mu),$$

where $\mathrm{int}(A)$ denotes the interior of A and $\mathrm{cl}(A)$ its closure (for the weak topology).

For a proof of Sanov's theorem, see [34, Thm 6.2.10].

Roughly speaking, $\mathbb{P}(L_n \in A)$ behaves like $e^{-nH(A|\mu)}$ when n is large. We write the statement of this theorem: $\mathbb{P}(L_n \in A) \underset{n\to\infty}{\asymp} e^{-nH(A|\mu)}$ for short.

Let us explain the heuristics upon which rely [53] and also the articles [52, 57]. To interpret the transport-entropy inequality $\alpha\left(\mathcal{T}_c\right) \le H$, let us define $A_t = \{\nu \in \mathrm{P}(\mathcal{X}); \mathcal{T}_c(\nu,\mu) \ge t\}$, for all $t \ge 0$. Note that the transport-entropy inequality can be rewritten as $\alpha(t) \le H(A_t|\mu)$, $t \ge 0$. But, according to Sanov's theorem,

$$\mathbb{P}(\mathcal{T}_c(L_n,\mu) \ge t) = \mathbb{P}(L_n \in A_t) \underset{n\to\infty}{\asymp} e^{-nH(A_t|\mu)}.$$

Consequently, the transport-entropy inequality $\alpha\left(\mathcal{T}_c\right) \le H$ is intimately linked to the large deviation estimate

$$\limsup_{n\to+\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{T}_c(L_n,\mu) \ge t) \le -\alpha(t),\ t \ge 0.$$

Based on this large deviation heuristics, Gozlan and Léonard have obtained in [53] the following estimates for the deviation of the empirical mean.

Theorem 5.2. *Let α be any function in $\mathcal{A}$ and assume that $c(x,x) = 0$, for all x. Then, the following statements are equivalent.*

(a) $\alpha(\mathcal{T}_c(\nu,\mu)) \le H(\nu|\mu), \quad \forall \nu \in \mathrm{P}(\mathcal{X})$
(b) $\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n(u^{cc}) + \mu(u^c) \ge r) \le -\alpha(r), \quad \forall r \ge 0, \forall u \in \mathcal{U}^{\forall}_{\mathrm{exp}}(\mu)$
(c) $\forall n \ge 1,\ \frac{1}{n} \log \mathbb{P}(L_n(u^{cc}) + \mu(u^c) \ge r) \le -\alpha(r), \quad \forall r \ge 0, \forall u \in {\mathcal{U}^{\forall}}_{\mathrm{exp}}(\mu)$

where $\mathcal{U}^{\forall}_{\mathrm{exp}}(\mu) := \left\{u : \mathcal{X} \to \mathbb{R},\ \text{measurable}, \forall s > 0, \int e^{s|u|}\, d\mu < \infty\right\}$. It is supposed that c is such that u^{cc} and u^c are measurable functions for all $u \in \mathcal{U}^{\forall}_{\mathrm{exp}}(\mu)$. This is the case in particular if either $c = d$ is a lower semicontinuous metric cost or c is continuous.

Specializing to the situation where $c = d$, since $u^{dd} = -u^d \in \mathrm{Lip}(1)$, this means:

Corollary 5.3 (Deviation of the empirical mean)**.** *Suppose that $\int_{\mathcal{X}} e^{sd(x_o,\,\cdot\,)}\, d\mu < \infty$ for some $x_o \in \mathcal{X}$ and all $s > 0$. Then, the following statements are equivalent.*

(a) $\alpha(W_1(\nu,\mu)) \le H(\nu|\mu), \quad \forall \nu \in \mathrm{P}(\mathcal{X})$.
(b) $\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\frac{1}{n}\sum_{i=1}^n u(X_i) \ge \int_{\mathcal{X}} u\, d\mu + r) \le -\alpha(r), \quad \forall r \ge 0, \forall u \in \mathrm{Lip}(1)$.
(c) $\forall n \ge 1,\ \frac{1}{n} \log \mathbb{P}(\frac{1}{n}\sum_{i=1}^n u(X_i) \ge \int_{\mathcal{X}} u\, d\mu + r) \le -\alpha(r), \quad \forall r \ge 0, \forall u \in \mathrm{Lip}(1)$.

Sanov's Theorem and concentration inequalities match also well together, since both give asymptotic results for probabilities of events related to an i.i.d sequence.

In [52], Gozlan has established the following converse to Proposition 4.6:

Theorem 5.4. *Let μ be a probability measure on $\mathcal{X}$ and $(r_o^n)_n$ a sequence of nonnegative numbers such that $r_o^n/n \to 0$ when $n \to +\infty$. Suppose that for all integer n the product measure μ^n verifies the following concentration inequality:*

$$\mu^n(\mathrm{enl}_\theta(A,r)) \geq 1 - \exp\left(-n\alpha\left(\frac{r - r_o^n}{n}\right)\right), \ r \geq r_o^n, \tag{29}$$

for all $A \subset \mathcal{X}^n$ with $\mu^n(A) \geq 1/2$, where $\mathrm{enl}_\theta(A,r)$ is defined in Proposition 4.6. Then μ satisfies the transport-entropy inequality $\alpha(\mathcal{T}_{\theta(d)}) \leq H$.

Together with Proposition 4.6, this result shows that the transport-entropy inequality $\alpha\left(\mathcal{T}_{\theta(d)}\right) \leq H$ is an equivalent formulation of the family of concentration inequalities (29).

Let us emphasize a nice particular case.

Corollary 5.5. *Let μ be a probability measure on $\mathcal{X}$; μ enjoys the Gaussian dimension-free concentration property if and only if μ verifies Talagrand inequality $\mathbf{T}_2$. More precisely, μ satisfies $\mathbf{T}_2(C)$ if and only if there is some $K > 0$ such that for all integer n the inequality*

$$\mu^n(A^r) \geq 1 - Ke^{-r^2/C}, \ r \geq 0,$$

holds for all $A \subset \mathcal{X}^n$ with $\mu^n(A) \geq 1/2$ and where $A^r = \{x \in \mathcal{X}^n; \inf_{y\in A} d_2(x,A) \leq r\}$ and $d_2(x,y) = \left(\sum_{i=1}^n d(x_i,y_i)^2\right)^{1/2}$.

To put these results in perspective, let us recall that in recent years numerous functional inequalities and tools were introduced to describe the concentration of measure phenomenon. Besides transport-entropy inequalities, let us mention other recent approaches based on Poincaré inequalities [55, 19], logarithmic Sobolev inequalities [67, 17], modified logarithmic Sobolev inequalities [19, 21, 47, 10], inf-convolution inequalities [82, 66], Beckner-Latała-Oleszkiewicz inequalities [11, 65, 9, 7]...So the interest of Theorem 5.4 is that it says that transport-entropy inequalities are the right point of view, because they are equivalent to concentration estimates for product measures.

Proof of Corollary 5.5. Let us show that dimension-free Gaussian concentration implies Talagrand inequality (the other implication is Corollary 4.4). For every integer n, and $x \in \mathcal{X}^n$, define $L_n^x = n^{-1}\sum_{i=1}^n \delta_{x_i}$. The map $x \mapsto W_2(L_n^x,\mu)$ is $1/\sqrt{n}$-Lipschitz with respect to the metric d_2. Indeed, if $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_n)$ are in $\mathcal{X}^n$, then the triangle inequality implies that

$$|W_2(L_n^x,\mu) - W_2(L_n^y,\mu)| \leq W_2(L_n^x,L_n^y).$$

According to the convexity property of $\mathcal{T}_2(\,\cdot\,,\,\cdot\,)$ (see e.g [104, Theorem 4.8]), one has

$$\mathcal{T}_2(L_n^x,L_n^y) \leq \frac{1}{n}\sum_{i=1}^n \mathcal{T}_2(\delta_{x_i},\delta_{y_i}) = \frac{1}{n}\sum_{i=1}^n d(x_i,y_i)^2 = \frac{1}{n}d_2(x,y)^2,$$

which proves the claim.

Now, let $(X_i)_i$ be an i.i.d sequence of law μ and let L_n be its empirical measure. Let m_n be the median of $W_2(L_n,\mu)$ and define $A = \{x : W_2(L_n^x,\mu) \leq m_n\}$. Then $\mu^n(A) \geq 1/2$ and it is easy to show that $A^r \subset \{x : W_2(L_n^x, \mu) \leq m_n + r/\sqrt{n}\}$. Applying the Gaussian concentration inequality to A gives

$$\mathbb{P}\left(W_2(L_n, \mu) > m_n + r/\sqrt{n}\right) \leq K \exp\left(-r^2/C\right), \ r \geq 0.$$

Equivalently, as soon as $u \geq m_n$, one has

$$\mathbb{P}\left(W_2(L_n, \mu) > u\right) \leq K \exp\left(-n(u - m_n)^2/C\right).$$

Now, it is not difficult to show that $m_n \to 0$ when $n \to \infty$ (see the proof of [52, Theorem 3.4]). Consequently,

$$\limsup_{n \to +\infty} \frac{1}{n} \log \mathbb{P}\left(W_2(L_n, \mu) > u\right) \leq -u^2/C.$$

for all $u \geq 0$.

On the other hand, according to Sanov's Theorem 5.1,

$$\liminf_{n \to +\infty} \frac{1}{n} \log \mathbb{P}\left(W_2(L_n, \mu) > u\right) \geq -\inf\left\{H(\nu|\mu) : \nu \in \mathrm{P}(\mathcal{X}) \text{ s.t. } W_2(\nu, \mu) > u\right\}.$$

This together with the preceding inequality yields

$$\inf\left\{H(\nu|\mu) : \nu \in \mathrm{P}(\mathcal{X}) \text{ s.t. } W_2(\nu, \mu) > u\right\} \geq u^2/C$$

or in other words,

$$W_2(\nu, \mu)^2 \leq C H(\nu|\mu),$$

and this completes the proof. □

6. Integral criteria

Let us begin with a basic observation concerning the integrability.

Proposition 6.1. *Suppose that a probability measure μ on $\mathcal{X}$ verifies the inequality $\alpha(\mathcal{T}_{\theta(d)}) \leq H$, and let $x_o \in \mathcal{X}$; then $\int_{\mathcal{X}} \exp\left(\alpha \circ \theta(\varepsilon d(x, x_o))\right) d\mu(x)$ is finite for all $\varepsilon > 0$ small enough.*

Proof. If μ verifies the inequality $\alpha(\mathcal{T}_{\theta(d)}) \leq H$, then according to Jensen inequality, it verifies the inequality $\alpha \circ \theta\left(\mathcal{T}_d\right) \leq H$ and according to Theorem 4.2, the inequality

$$\mu(A^r) \geq 1 - \exp(-\alpha \circ \theta(r - r_o)), \ r \geq r_o = \theta^{-1} \circ \alpha^{-1}(\log 2),$$

holds for all A with $\mu(A) \geq 1/2$. Let m be a median of the function $x \mapsto d(x, x_o)$; applying the previous inequality to $A = \{x \in \mathcal{X}; d(x, x_o) \leq m\}$ yields

$$\mu(d(x, x_o) > m + r) = \mu(\mathcal{X} \setminus A^r) \leq \exp(-\alpha \circ \theta(r - r_o)), \ r \geq r_o.$$

It follows easily that $\int \exp\left(\alpha \circ \theta(\varepsilon d(x, x_o))\right) d\mu(x) < +\infty$, if ε is sufficiently small. □

The theorem below shows that this integrability condition is also sufficient when the function α is supposed to be subquadratic near 0.

Theorem 6.2. *Let μ be a probability measure on $\mathcal{X}$ and define $\alpha^{\circledast}(s) = \sup_{t \geq 0}\{st - \alpha(t)\}$, for all $s \geq 0$. If the function α is such that $\limsup_{t \to 0} \alpha(t)/t^2 < +\infty$ and $\sup\{\alpha^{\circledast}(t); t : \alpha^{\circledast}(t) < +\infty\} = +\infty$, then the following statements are equivalent:*

1. *There is some $a > 0$ such that $\alpha\left(a\mathcal{T}_{\theta(d)}(\nu, \mu)\right) \leq H(\nu|\mu)$.*
2. *There is some $b > 0$ such that $\int_{\mathcal{X} \times \mathcal{X}} e^{\alpha \circ \theta(bd(x,y))} \, d\mu(x) d\mu(y) < +\infty$.*

Djellout, Guillin and Wu [36] were the first ones to notice that the inequality $\mathbf{T}_1$ is equivalent to the integrability condition $\int_{\mathcal{X}\times\mathcal{X}} e^{bd(x,y)^2}\,d\mu(x)d\mu(y) < +\infty$. After them, this characterization was extended to other functions α and θ by Bolley and Villani [22]. Theorem 6.2 is due to Gozlan [49]. Let us mention that the constants a and b are related to each other in [49, Theorem 1.15].

Again to avoid technical difficulties, we are going to establish a particular case of Theorem 6.2:

Proposition 6.3. *If $M = \int_{\mathcal{X}\times\mathcal{X}} e^{b^2\frac{d(x,y)^2}{2}}\,d\mu(x)d\mu(y)$ is finite for some $b > 0$, then μ verifies the following $\mathbf{T}_1$ inequality:*

$$\mathcal{T}_d(\nu,\mu) \le \frac{1}{b}\sqrt{1+2\log M}\sqrt{2H(\nu|\mu)},$$

for all $\nu \in \mathrm{P}(\mathcal{X})$.

Proof. First one can suppose that $b = 1$ (if this is not the case, just replace the distance d by the distance bd). Let $C = 2(1+2\log(M))$; according to Corollary 3.4, it is enough to prove that

$$\int e^{sf}\,d\mu \le e^{\frac{s^2C}{4}},\ s\ge 0$$

for all 1-Lipschitz function with $\int f\,d\mu = 0$. Let X, Y be two independent variables of law μ; using Jensen inequality, the symmetry of $f(X)-f(Y)$, the inequality $(2i)! \ge 2^i\cdot i!$ and the fact that f is 1-Lipschitz, one gets

$$\begin{aligned}\mathbb{E}[e^{sf(X)}] \le \mathbb{E}\left[e^{s(f(X)-f(Y))}\right] &= \sum_{i=0}^{+\infty}\frac{s^{2i}\mathbb{E}\left[(f(X)-f(Y))^{2i}\right]}{(2i)!}\\ &\le \sum_{i=0}^{+\infty}\frac{s^{2i}\mathbb{E}\left[d(X,Y)^{2i}\right]}{2^i\cdot i!} = \mathbb{E}\left[\exp\left(\frac{s^2d(X,Y)^2}{2}\right)\right].\end{aligned}$$

So, for $s \le 1$, Jensen inequality gives $\mathbb{E}[e^{sf(X)}] \le M^{s^2}$. If $s \ge 1$, then Young inequality implies $\mathbb{E}[e^{sf}] \le \mathbb{E}\left[e^{s(f(X)-f(Y))}\right] \le e^{\frac{s^2}{2}}M$. So in all cases, $\mathbb{E}[e^{sf}] \le e^{\frac{s^2}{2}}M^{s^2} = e^{s^2C/4}$ which completes the proof. □

7. Transport inequalities with uniformly convex potentials

This section is devoted to some results which have been proved by Cordero-Erausquin in [27] and Cordero-Erausquin, Gangbo and Houdré in [28].

Let us begin with a short overview of [27]. The state space is $\mathcal{X} = \mathbb{R}^k$. Let $V : \mathbb{R}^k \to \mathbb{R}$ be a function of class $\mathcal{C}^2$ which is semiconvex, i.e. $\mathrm{Hess}_x V \ge \kappa\mathrm{Id}$ for all x, for some real κ. If $\kappa > 0$, the potential V is said to be uniformly convex. Define

$$d\mu(x) := e^{-V(x)}\,dx$$

and assume that μ is a probability measure. The main result of [27] is the following

Theorem 7.1. *Let f, g be nonnegative compactly supported functions with f of class $\mathcal{C}^1$ and $\int f\,d\mu = \int g\,d\mu = 1$. If $T(x) = x + \nabla\theta(x)$ is the Brenier map pushing forward $f\mu$ to $g\mu$ (see Theorem 2.9), then*

$$H(g\mu|\mu) \geq H(f\mu|\mu) + \int_{\mathbb{R}^k} \nabla f \cdot \nabla\theta\,d\mu + \frac{\kappa}{2}\int_{\mathbb{R}^k} |\nabla\theta|^2\,f d\mu. \tag{30}$$

Before presenting a sketch of the proof of this result, let us make a couple of comments.

- This result is an extension of Talagrand inequality (6).
- About the regularity of θ. As a convex function, θ is differentiable almost everywhere and it admits a Hessian in the sense of Alexandrov almost everywhere (this is the statement of Alexandrov's theorem). A function θ admits a Hessian in the sense of Alexandrov at $x \in \mathbb{R}^k$ if it is differentiable at x and there exists a symmetric linear map H such that
$$\theta(x+u) = \theta(x) + \nabla\theta(x)\cdot u + \frac{1}{2}Hu\cdot u + o(|u|^2).$$
As a definition, this linear map H is the Hessian in the sense of Alexandrov of θ at x and it is denoted $\mathrm{Hess}_x\,\theta$. Its trace is called the Laplacian in the sense of Alexandrov and is denoted $\Delta_A\theta(x)$.

Outline of the proof. The change of variables formula leads us to the Monge-Ampère equation

$$f(x)e^{-V(x)} = g(T(x))e^{-V(T(x))}\det(\mathrm{Id} + \mathrm{Hess}_x\,\theta)$$

Taking the logarithm , we obtain

$$\log g(T(x)) = \log f(x) + V(x + \nabla\theta(x)) - V(x) - \log\det(\mathrm{Id} + \mathrm{Hess}_x\,\theta).$$

Our assumption on V gives us $V(x + \nabla\theta(x)) - V(x) \geq \nabla V(x)\cdot\nabla\theta(x) + \kappa|\nabla\theta|^2/2$. Since $\log(1+t) \leq t$, we have also $\log\det(\mathrm{Id} + \mathrm{Hess}_x\,\theta) \leq \Delta_A\theta(x)$ where $\Delta_A\theta$ stands for the Alexandrov Laplacian. This implies that $f\mu$-almost everywhere

$$\log g(T(x)) \geq \log f(x) + \nabla V(x)\cdot\nabla\theta(x) - \Delta_A\theta(x) + \kappa|\nabla\theta|^2/2$$

and integrating

$$\int_{\mathbb{R}^k} \log g(T)\,f d\mu \geq \int_{\mathbb{R}^k} f\log f\,d\mu + \int_{\mathbb{R}^k} [\nabla V\cdot\nabla\theta - \Delta_A\theta]\,f d\mu + \frac{\kappa}{2}\int_{\mathbb{R}^k} |\nabla\theta|^2\,f d\mu$$

Integrating by parts (at this point, a rigorous proof necessitates to take account of the *almost everywhere* in the definition of Δ_A), we obtain

$$H(g\mu|\mu) \geq H(f\mu|\mu) + \int_{\mathbb{R}^k} \nabla\theta\cdot\nabla f\,d\mu + \frac{\kappa}{2}\int_{\mathbb{R}^k} |\nabla\theta|^2\,f d\mu$$

which is the desired result. □

Next results are almost immediate corollaries of this theorem.

Corollary 7.2 (Transport inequality)**.** *If V is of class $\mathcal{C}^2$ with* $\mathrm{Hess}\,V \geq \kappa\mathrm{Id}$ *and $\kappa > 0$, then the probability measure $d\mu(x) = e^{-V(x)}\,dx$ satisfies the transport inequality $\mathbf{T}_2(2/\kappa)$:*

$$\frac{\kappa}{2}W_2^2(\nu,\mu) \leq H(\nu|\mu),$$

for all $\nu \in \mathrm{P}(\mathbb{R}^k)$.

Outline of the proof. Plug $f = 1$ into (30). □

This transport inequality extends Talagrand's $\mathbf{T}_2$-inequality [102].

In [42], Feyel and Üstünel have derived another type of extension of $\mathbf{T}_2$ from the finite dimension setting to an abstract Wiener space. Their proof is based on Girsanov theorem.

Next result is the well-known Bakry-Emery criterion for the logarithmic Sobolev inequality [4].

Corollary 7.3 (Logarithmic Sobolev inequality)**.** *If V is of class $\mathcal{C}^2$ with* $\mathrm{Hess}\, V \geq \kappa \mathrm{Id}$ *and $\kappa > 0$, then the probability measure $d\mu(x) = e^{-V(x)}\, dx$ satisfies the logarithmic Sobolev inequality* $\mathbf{LS}(2/\kappa)$ *(see Definition 8.9 below):*

$$H(f\mu|\mu) \leq \frac{2}{\kappa} \int_{\mathcal{X}} |\nabla \sqrt{f}|^2\, d\mu$$

for all sufficiently regular f such that $f\mu \in \mathrm{P}(\mathbb{R}^k)$.

Outline of the proof. Plugging $g = 1$ into (30) yields

$$H(f\mu|\mu) \leq -\int_{\mathbb{R}^k} \nabla f \cdot \nabla \theta\, d\mu - \frac{\kappa}{2} \int_{\mathbb{R}^k} |\nabla \theta|^2\, f d\mu \tag{31}$$

where $T(x) = x + \nabla(x)$ is the Brenier map pushing forward $f\mu$ to μ. Since $\nabla\theta$ is unknown to us, we are forced to optimize as follows

$$H(f\mu|\mu) \leq \sup_{\nabla\theta} \left\{ -\int_{\mathbb{R}^k} \nabla f \cdot \nabla \theta\, d\mu - \frac{\kappa}{2} \int_{\mathbb{R}^k} |\nabla \theta|^2\, f d\mu \right\} = \frac{2}{\kappa} I(f\mu|\mu),$$

which is the desired inequality. □

The next inequality has been discovered by Otto and Villani [89]. It will be used in Section 8 for comparing transport and logarithmic Sobolev inequalities. More precisely, Otto-Villani's Theorem 8.12 states that if μ satisfies the logarithmic Sobolev inequality, then it satisfies $\mathbf{T}_2$.

Let us define the (usual) Fisher information with respect to μ by

$$I_F(f|\mu) = \int |\nabla \log f|^2\, f d\mu$$

for all positive and sufficiently smooth function f.

Corollary 7.4 (HWI inequality)**.** *If V is of class $\mathcal{C}^2$ with* $\mathrm{Hess}\, V \geq \kappa \mathrm{Id}$ *for some* real κ, *the probability measure $d\mu(x) = e^{-V(x)}\, dx$ satisfies the HWI inequality*

$$H(f\mu|\mu) \leq W_2(f\mu, \mu) \sqrt{I_F(f|\mu)} - \frac{\kappa}{2} W_2^2(f\mu, \mu)$$

for all nonnegative smooth compactly supported function f with $\int_{\mathbb{R}^k} f\, d\mu = 1$.

Note that the HWI inequality gives back the celebrated Bakry-Emery criterion.

Outline of the proof. Start from (31), use $W_2^2(f\mu,\mu) = \int_{\mathbb{R}^k} |\nabla\theta|^2\, f d\mu$ and

$$\begin{aligned} -\int \nabla\theta\cdot\nabla f\, d\mu &= -\int \nabla\theta\cdot\nabla\log f\, f d\mu \\ &\le \left(\int |\nabla\theta|^2\, f d\mu \int |\nabla\log f|^2\, f d\mu\right)^{1/2} = W_2(f\mu|\mu)\sqrt{I_F(f|\mu)}, \end{aligned}$$

and here you are. □

Now, let us have a look at the results of [28]. They extend Theorem 7.1 and its corollaries. Again, the state space is $\mathcal{X} = \mathbb{R}^k$ and the main ingredients are

- An entropy profile: $r \in [0,\infty) \mapsto \mathrm{s}(r) \in \mathbb{R}$;
- A cost function: $v \in \mathbb{R}^k \mapsto c(v) \in [0,\infty)$;
- A potential: $x \in \mathbb{R}^k \mapsto V(x) \in \mathbb{R}$.

The framework of our previous Theorem 7.1 corresponds to the entropy profile $\mathrm{s}(r) = r\log r - r$ and the quadratic transport cost $c(y-x) = |y-x|^2/2$.

We are only interested in probability measures $d\rho(x) = \rho(x)\,dx$ which are absolutely continuous and we identify ρ and its density. The *free energy functional* is

$$F(\rho) := \int_{\mathbb{R}^k} [\mathrm{s}(\rho) + \rho V](x)\,dx$$

and our reference measure μ is the *steady state*: the unique minimizer of F. Since s will be assumed to be strictly convex, μ is the unique solution of

$$\mathrm{s}'(\mu) = -V, \tag{32}$$

which, by (20) is

$$\mu = \mathrm{s}^{*\prime}(-V).$$

As $\mathrm{s}(\rho)+\mathrm{s}^*(-V) \ge -V\rho$, see (19), in order that F is a well-defined $(-\infty,\infty]$-valued function, it is enough to assume that $\int_{\mathbb{R}^k} \mathrm{s}^*(-V)(x)\,dx < \infty$. One also requires that $\int_{\mathbb{R}^k} \mathrm{s}^{*\prime}(-V)(x)\,dx = 1$ so that μ is a probability density.
The free energy is the sum of the *entropy* $S(\rho)$ and the *internal energy* $U(\rho)$ which are defined by

$$S(\rho) := \int_{\mathbb{R}^k} \mathrm{s}(\rho)(x)\,dx, \qquad U(\rho) := \int_{\mathbb{R}^k} V(x)\rho(x)\,dx.$$

It is assumed that

($\mathrm{A_s}$) (a) $\mathrm{s} \in \mathcal{C}^2(0,\infty) \cap \mathcal{C}([0,\infty))$ is strictly convex, $\mathrm{s}(0) = 0$, $\mathrm{s}'(0) = -\infty$ and
(b) $r \in (0,\infty) \mapsto r^d \mathrm{s}(r^{-d})$ is convex increasing;
(A_c) c is convex, of class $\mathcal{C}^1$, even, $c(0) = 0$ and $\lim_{|v|\to\infty} c(v)/|v| = \infty$;
(A_V) For some real number κ, $V(y) - V(x) \ge \nabla V(x)\cdot(y-x) + \kappa c(y-x)$, for all x, y.

If $\kappa > 0$, the potential V is said to be uniformly c-convex.

We see with Assumption (A_V) that the cost function c is a tool for quantifying the curvature of the potential V. Also note that if $\kappa > 0$ and $c(y-x) = \|y-x\|^p$ for some p in (A_V), letting y tend to x, one sees that it is necessary that $p \ge 2$.

The transport cost associated with c is $\mathcal{T}_c(\rho_0, \rho_1)$. Theorem 2.9 admits an extension to the case of strictly convex transport cost $c(y-x)$ (instead of the quadratic cost). Under the assumption (A_c) on c, if the transport cost $\mathcal{T}_c(\rho_0, \rho_1)$ between two absolutely continuous probability measures ρ_0 and ρ_1 is finite, there exists a unique (generalized) Brenier map T which pushes forward ρ_0 to ρ_1 and it is represented by

$$T(x) = x + \nabla c^*(\nabla\theta(x))$$

for some function θ such that $\theta(x) = -\inf_{y\in\mathbb{R}^k}\{c(y-x) + \eta(y)\}$ for some function η. This has been proved by Gangbo and McCann in [44] and T will be named later the Gangbo-McCann map.

The fundamental result of [28] is the following extension of Theorem 7.1.

Theorem 7.5. *For any ρ_0, ρ_1 which are compactly supported and such that $\mathcal{T}_c(\rho_0, \rho_1) < \infty$, we have*

$$F(\rho_1) - F(\rho_0) \geq \kappa\mathcal{T}_c(\rho_0, \rho_1) + \int_{\mathbb{R}^k} (T(x) - x) \cdot \nabla[\mathrm{s}'(\rho_0) - \mathrm{s}'(\mu)](x)\, \rho_0(x)\, dx \tag{33}$$

where T is the Gangbo-McCann map which pushes forward ρ_0 to ρ_1.

Outline of the proof. Let us first have a look at S. Thanks to the assumption (A_s) one can prove that S is displacement convex. This formally means that if $\mathcal{T}_c(\rho_0, \rho_1) < \infty$,

$$S(\rho_1) - S(\rho_0) \geq \frac{d}{dt} S(\rho_t)_{|t=0}$$

where $(\rho_t)_{0\leq t\leq 1}$ is the displacement interpolation of ρ_0 to ρ_1 which is defined by

$$\rho_t := [(1-t)\mathrm{Id} + tT]_\# \rho_0, \quad 0 \leq t \leq 1$$

where T pushes forward ρ_0 to ρ_1. Since $\frac{\partial\rho}{\partial t}(t,x) + \nabla\cdot[\rho(t,x)(T(x)-x)] = 0$, another way of writing this convex inequality is

$$S(\rho_1) - S(\rho_0) \geq \int_{\mathbb{R}^k} (T(x) - x) \cdot \nabla[\mathrm{s}'(\rho_0)](x)\, \rho_0(x)\, dx. \tag{34}$$

Note that when the cost is quadratic, by Theorem 2.9 the Brenier-Gangbo-McCann map between the uniform measures ρ_0 and ρ_1 on the balls $B(0, r_0)$ and $B(0, r_1)$ is given by $T = (r_1/r_0)\mathrm{Id}$ so that the image $\rho_t = T_{t\#}\rho_0$ of ρ_0 by the displacement $T_t = (1-t)\mathrm{Id} + tT$ at time $0 \leq t \leq 1$ is the uniform measure on the ball $B(0, r_t)$ with $r_t = (1-t)r_0 + tr_1$. Therefore, $t \in [0,1] \mapsto S(\rho_t) = r_t^d \mathrm{s}(r_t^{-d})$ is convex for all $0 < r_0 \leq r_1$ if and only if $r^d \mathrm{s}(r^{-d})$ is convex: i.e. assumption (A_s-b).
It is immediate that under the assumption (A_V) we have

$$U(\rho_1) - U(\rho_0) \geq \int_{\mathbb{R}^k} \nabla V(x) \cdot [T(x) - x]\, \rho_0(x) dx + \kappa\mathcal{T}_c(\rho_0, \rho_1).$$

One can also prove that this is a necessary condition for assumption (A_V) to hold true. Summing (34) with this inequality, and taking (32) into account, leads us to (33). □

Let us define the *generalized relative entropy*

$$S(\rho|\mu) := F(\rho) - F(\mu) = \int_{\mathbb{R}^k} [\mathrm{s}(\rho) - \mathrm{s}(\mu) - \mathrm{s}'(\mu)(\rho - \mu)](x)\, dx$$

on the set $\mathrm{P}^{\mathrm{ac}}(\mathbb{R}^k)$ of all absolutely continuous probability measures on $\mathbb{R}^k$. It is a $[0,\infty]$-valued convex function which admits μ as its unique minimum.

Theorem 7.6 (Transport-entropy inequality). *Assume that the constant κ in assumption* (A_V) *is positive: $\kappa > 0$. Then, μ satisfies the following transport-entropy inequality*

$$\kappa \mathcal{T}_c(\rho,\mu) \le S(\rho|\mu),$$

for all $\rho \in \mathrm{P}^{\mathrm{ac}}(\mathbb{R}^k)$.

Outline of the proof. If ρ and μ are compactly supported, plug $\rho_0 = \mu$ and $\rho_1 = \rho$ into (33) to obtain the desired result. Otherwise, approximate ρ and μ by compactly supported probability measures. □

Let us define the *generalized relative Fisher information* for all $\rho \in \mathrm{P}^{\mathrm{ac}}(\mathbb{R}^k)$ by

$$\mathcal{I}(\rho|\mu) := \int_{\mathbb{R}^k} \kappa\, c^*\Big(-\kappa^{-1}\nabla[\mathrm{s}'(\rho) - \mathrm{s}'(\mu)](x)\Big)\, d\rho(x) \in [0,\infty]$$

with $\mathcal{I}(\rho|\mu) = \infty$ if $\nabla\rho$ is undefined on a set with positive Lebesgue measure.

Theorem 7.7 (Entropy-information inequality). *Assume that $\kappa > 0$. Then, μ satisfies the following entropy-information inequality*

$$S(\rho|\mu) \le \mathcal{I}(\rho|\mu),$$

for all $\rho \in \mathrm{P}^{\mathrm{ac}}(\mathbb{R}^k)$.

Outline of the proof. Change c into κc so that the constant κ becomes $\kappa = 1$. If ρ and μ are compactly supported, plug $\rho_0 = \rho$ and $\rho_1 = \mu$ into (33) to obtain

$$\begin{aligned} F(\rho) - F(\mu) + \mathcal{T}_c(\rho,\mu) &\le \int_{\mathbb{R}^k} (T(x) - x)\cdot \nabla[\mathrm{s}'(\mu) - \mathrm{s}'(\rho)](x)\, d\rho(x) \\ &\le \int_{\mathbb{R}^k} c(T(x) - x)\, d\rho(x) + \int_{\mathbb{R}^k} c^*\Big(\nabla[\mathrm{s}'(\mu) - \mathrm{s}'(\rho)](x)\Big)\, d\rho(x) \end{aligned}$$

where the last inequality is a consequence of Fenchel inequality (19). Since T is the Gangbo-McCann map between ρ and μ, we have $\int_{\mathbb{R}^k} c(T(x)-x)\, d\rho(x) = \mathcal{T}_c(\rho,\mu)$. Therefore,

$$F(\rho) - F(\mu) \le \int_{\mathbb{R}^k} c^*\Big(-\nabla[\mathrm{s}'(\rho) - \mathrm{s}'(\mu)](x)\Big)\, d\rho(x)$$

which is the desired result for compactly supported measures. For the general case, approximate ρ and μ by compactly supported probability measures. □

A direct application of these theorems allow us to recover inequalities which were first proved by Bobkov and Ledoux in [20].

Corollary 7.8. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^k$ and V a convex potential. Suppose that V is uniformly p-convex with respect to $\|\cdot\|$ for some $p \ge 2$; this means that there exists a constant $\kappa > 0$ such that for all $x, y \in \mathbb{R}^k$*

$$V(x) + V(y) - 2V\left(\frac{x+y}{2}\right) \ge \frac{\kappa}{p}\|y - x\|^p. \tag{35}$$

Denote $d\mu(x) := e^{-V(x)}\,dx$ *(where it is understood that* μ *is a probability measure) and* $H(\rho|\mu)$ *the usual relative entropy.*

(1) μ *verifies the transport-entropy inequality*

$$\kappa\mathcal{T}_{c_p}(\rho,\mu) \le H(\rho|\mu),$$

for all $\rho \in \mathrm{P}^{\mathrm{ac}}(\mathbb{R}^k)$*, where* $c_p(y-x) = \|y-x\|^p/p$*.*

(2) μ *verifies the entropy-information inequality*

$$H(f\mu|\mu) \le \frac{1}{q\kappa^{q-1}}\int_{\mathbb{R}^k} \|\nabla \log f\|_*^q\, f\,d\mu$$

for all smooth nonnegative function f *such that* $\int_{\mathbb{R}^k} f\,d\mu = 1$*, where* $\|\cdot\|_*$ *is the dual norm of* $\|\cdot\|$ *and* $1/p+1/q=1$*.*

8. Links with other functional inequalities

In this section, we investigate the position of transport-entropy inequalities among the relatively large class of functional inequalities (mainly of Sobolev type) appearing in the literature. We will be concerned with transport-entropy inequalities of the form $\mathcal{T}_{\theta(d)} \le H$ since inequalities of the form $\alpha\left(\mathcal{T}_{\theta(d)}\right) \le H$ with a nonlinear function α are mainly described in terms of integrability conditions according to Theorem 6.2.

8.1. Links with the Property (τ).

In [82], Maurey introduced the Property (τ) which we describe now. Let c be a cost function on $\mathcal{X}$. Recall that for all $f \in \mathcal{B}_b(\mathcal{X})$, the function Q_cf is defined by $Q_cf(x) := \inf_y\{f(y)+c(x,y)\}$. If a probability measure $\mu \in \mathrm{P}(\mathcal{X})$ satisfies

$$\left(\int e^{Q_cf}\,d\mu\right)\left(\int e^{-f}\,d\mu\right) \le 1, \tag{τ}$$

for all $f \in \mathcal{B}_b(\mathcal{X})$, one says that the couple (μ,c) satisfies the Property (τ).

The basic properties of this class of functional inequalities are summarized in the following result.

Proposition 8.1. *Suppose that* (μ,c) *verifies the Property* (τ)*, then the following holds.*

(1) *The probability measure* μ *verifies the transport-entropy inequality* $\mathcal{T}_c \le H$*.*
(2) *For all positive integer* n*, the couple* $(\mu^n, c^{\oplus n})$*, with* $c^{\oplus n}(x,y) = \sum_{i=1}^n c(x_i,y_i)$*,* $x,y \in \mathcal{X}^n$*, verifies the Property* (τ)*.*
(3) *For all positive integer* n*, and all Borel set* $A \subset \mathcal{X}^n$ *with* $\mu^n(A) > 0$*,*

$$\mu^n(\mathrm{enl}_{c^{\oplus n}}(A,r)) \ge 1 - \frac{1}{\mu^n(A)}e^{-r},\ \ r \ge 0$$

where

$$\mathrm{enl}_{c^{\oplus n}}(A,r) = \left\{x \in \mathcal{X}^n;\ \inf_{y\in A}\sum_{i=1}^n c(x_i,y_i) \le r\right\}$$

The third point of the proposition above was the main motivation for the introduction of this class of inequalities. In [82], Maurey established that the symmetric exponential probability measure $d\nu(x) = \frac{1}{2}e^{-|x|}\,dx$ on $\mathbb{R}$ satisfies the Property (τ) with the cost function $c(x,y) = a\min(|x|^2, |x|)$ for some constant $a > 0$. It enables him to recover Talagrand's concentration results for the multidimensional exponential distribution with sharp constants. Moreover, using the Prekopa-Leindler inequality (see Theorem 13.1 below), he showed that the standard Gaussian measure γ verifes the Property (τ) with the cost function $c(x,y) = \frac{1}{4}|x-y|^2$. The constant $1/4$ is sharp.

Proof. (1) According to the dual formulation of transport-entropy inequalities, see Corollary 3.3, μ verifies the inequality $\mathcal{T}_c \le H$ if and only if $\int e^{Q_c f}\,d\mu \cdot e^{-\int f\,d\mu} \le 1$ for all $f \in \mathcal{B}_b(\mathcal{X})$. Jensen inequality readily implies that this condition is weaker than the Property (τ).
(2) The proof of this tensorization property follows the lines of the proof of Theorem 4.3. We refer to [82] or [68] for a complete proof.
(3) Applying the Property (τ) satisfied by $(\mu^n, c^{\oplus n})$ to the function $u(x) = 0$ if $x \in A$ and $u(x) = t$, if $x \notin A$ and letting $t \to \infty$ yields the inequality:

$$\int e^{\inf_{y\in A} c^{\oplus n}(x,y)}\,d\mu^n(x) \le \frac{1}{\mu^n(A)},$$

which immediately implies the concentration inequality. □

In fact, the Property (τ) can be viewed as a symmetric transport-entropy inequality.

Proposition 8.2. *The couple (μ, c) verifies the Property (τ) if and only if μ verifies the following inequality*

$$\mathcal{T}_c(\nu_1, \nu_2) \le H(\nu_1|\mu) + H(\nu_2|\mu),$$

for all $\nu_1, \nu_2 \in \mathrm{P}(\mathcal{X})$.

Proof. According to the Kantorovich dual equality, see Theorem 2.2, the symmetric transport-entropy inequality holds if and only if for all couple (u, v) of bounded functions such that $u \oplus v \le c$, the inequality

$$\int u\,d\nu_1 - H(\nu_1|\mu) + \int v\,d\nu_2 - H(\nu_2|\mu) \le 0$$

holds for all $\nu_1, \nu_2 \in \mathrm{P}(\mathcal{X})$. Since $\sup_\nu\{\int u\,d\nu - H(\nu|\mu)\} = \log \int e^u\,d\mu$ (this is the convex conjugate of the result of Proposition B.1), the symmetric transport-entropy inequality holds if and only if

$$\int e^u\,d\mu \int e^v\,d\mu \le 1,$$

for all couple (u, v) of bounded functions such that $u \oplus v \le c$. One concludes by observing that for a given $f \in \mathcal{B}_b(\mathcal{X})$, the best function u such that $u \oplus (-f) \le c$ is $u = Q_c f$. □

As we have seen above, the Property (τ) is always stronger than the transport inequality $\mathcal{T}_c \le H$. Actually, when the cost function is of the form $c(x,y) = \theta(d(x,y))$ with a convex θ, the transport-entropy inequality and the Property (τ) are qualitatively equivalent as shown in the following.

Proposition 8.3. *Let μ be a probability measure on $\mathcal{X}$ and $\theta : [0,\infty) \to [0,\infty)$ be a convex function such that $\theta(0) = 0$. If μ verifies the transport-entropy inequality $\mathcal{T}_c \leq H$, then the couple $(\mu, \tilde{c})$, with $\tilde{c}(x,y) = 2\theta\left(\frac{d(x,y)}{2}\right)$ verifies the Property (τ).*

Proof. According to the dual formulation of the transport inequality $\mathcal{T}_c \leq H$, one has

$$\int e^{Q_c f}\, d\mu \cdot e^{-\int f\, d\mu} \leq 1$$

Applying this inequality with $\pm Q_c f$ instead of f, one gets

$$\int e^{Q_c(Q_c f)}\, d\mu \cdot e^{-\int Q_c f\, d\mu} \leq 1 \quad \text{and} \quad \int e^{Q_c(-Q_c f)}\, d\mu \cdot e^{\int Q_c f\, d\mu} \leq 1.$$

Multiplying these two inequalities yields to

$$\int e^{Q_c(Q_c f)}\, d\mu \cdot \int e^{Q_c(-Q_c f)}\, d\mu \leq 1.$$

Now, for all $x, y \in \mathcal{X}$, one has: $-f(y) + Q_c f(x) \leq \theta(d(x,y))$, and consequently, $-f \leq Q_c(-Q_c f)$. On the other hand, the convexity of θ easily yields

$$Q_c(Q_c f)(x) \leq \inf_{y \in \mathcal{X}} \left\{ f(y) + 2\theta\left(\frac{d(x,y)}{2}\right) \right\} = Q_{\tilde{c}} f.$$

This completes the proof. □

We refer to the works [95, 96] by Samson and [66] by Latała and Wojtaszczyk for recent advances in the study of the Property (τ).
In [96], Samson established different variants of the Property (τ) in order to derive sharp deviation results *à la* Talagrand for supremum of empirical processes.
In [66], Latała and Wojtaszczyk have considered a cost function naturally associated to a probability μ on $\mathbb{R}^k$. A symmetric probability measure μ is said to satisfy the inequality IC(β) for some constant $\beta > 0$ if it verifies the Property (τ) with the cost function $c(x,y) = \Lambda_\mu^*\left(\frac{x-y}{\beta}\right)$, where Λ_μ^* is the Cramér transform of μ defined by

$$\Lambda_\mu^*(x) = \sup_{y \in \mathbb{R}^k} \left\{ x \cdot y - \log \int e^{u \cdot y}\, d\mu(u) \right\},\ x \in \mathbb{R}^k,$$

For many reasons, this corresponds to an optimal choice for c. They have shown that isotropic log-concave distributions on $\mathbb{R}$ (mean equals zero and variance equals one) satisfy the inequality IC(48). They conjectured that isotropic log-concave distributions in all dimensions verify the inequality IC(β) with a universal constant β. This conjecture is stronger than the Kannan-Lovasz-Simonovits conjecture on the Poincaré constant of isotropic log-concave distributions [62].

8.2. **Definitions of the Poincaré and logarithmic Sobolev inequalities.** Let $\mu \in \mathrm{P}(\mathcal{X})$ be a given probability measure and $(P_t)_{t\geq 0}$ be the semigroup on $L^2(\mu)$ of a μ-reversible Markov process $(X_t)_{t\geq 0}$. The generator of $(P_t)_{t\geq 0}$ is $\mathcal{L}$ and its domain on $L^2(\mu)$ is $\mathbb{D}_2(\mathcal{L})$. Define the Dirichlet form

$$\mathcal{E}(g,g) := \langle -\mathcal{L}g, g\rangle_\mu, \quad g \in \mathbb{D}_2(\mathcal{L}).$$

Under the assumptions that

(a) $(X_t)_{t\geq 0}$ is μ-reversible,
(b) $(X_t)_{t\geq 0}$ is μ-ergodic,

$\mathcal{E}$ is closable in $L^2(\mu)$ and its closure $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$ admits the domain $\mathbb{D}(\mathcal{E}) = \mathbb{D}_2(\sqrt{-\mathcal{L}})$ in $L^2(\mu)$.

Remark 8.4. About these assumptions.

(a) means that the semigroup $(P_t)_{t\geq 0}$ is μ-symmetric.
(b) means that if $f \in \mathcal{B}_b(\mathcal{X})$ satisfies $P_t f = f$, μ-a.e. for all $t \geq 0$, then f is constant μ-a.e.

Definition 8.5 (Fisher information and Donsker-Varadhan information)**.**

(1) The Fisher information *of f with respect to μ (and the generator $\mathcal{L}$) is defined by*

$$I_\mu(f) = \mathcal{E}(\sqrt{f}, \sqrt{f})$$

for all $f \geq 0$ such that $\sqrt{f} \in \mathbb{D}(\mathcal{E})$.

(2) The Donsker-Varadhan information $I(\nu|\mu)$ *of the measure ν with respect to μ is defined by*

$$I(\nu|\mu) = \begin{cases} \mathcal{E}(\sqrt{f}, \sqrt{f}) & \textit{if } \ \nu = f\mu \in \mathrm{P}(\mathcal{X}),\ \sqrt{f} \in \mathbb{D}(\mathcal{E}) \\ +\infty & \textit{otherwise} \end{cases} \tag{36}$$

Example 8.6 (Standard situation). As a typical example, considering a probability measure $\mu = e^{-V(x)}dx$ with V of class $\mathcal{C}^1$ on a complete connected Riemannian manifold $\mathcal{X}$, one takes $(X_t)_{t\geq 0}$ to be the diffusion generated by

$$\mathcal{L} = \Delta - \nabla V \cdot \nabla$$

where Δ, ∇ are respectively the Laplacian and the gradient on $\mathcal{X}$. The Markov process $(X_t)_{t\geq 0}$ is μ-reversible and the corresponding Dirichlet form is given by

$$\mathcal{E}(g, g) = \int_{\mathcal{X}} |\nabla g|^2 \, d\mu,\ g \in \mathbb{D}(\mathcal{E}) = H^1(\mathcal{X}, \mu)$$

where $H^1(\mathcal{X}, \mu)$ is the closure with respect to the norm $\sqrt{\int_{\mathcal{X}} (|g|^2 + |\nabla g|^2) \, d\mu}$ of the space of infinitely differentiable functions on $\mathcal{X}$ with bounded derivatives of all orders. It also matches with the space of these $g \in L^2(\mathcal{X})$ such that $\nabla g \in L^2(\mathcal{X} \to T\mathcal{X}; \mu)$ in distribution.

Remark 8.7. The Fisher information in this example

$$I_\mu(f) = \int_{\mathcal{X}} |\nabla \sqrt{f}|^2 \, d\mu \tag{37}$$

differs from the usual Fisher information

$$I_F(f|\mu) := \int_{\mathbb{R}^k} |\nabla \log f|^2 \, f d\mu$$

by a multiplicative factor. Indeed, we have $I_\mu = I_F/4$. The reason for preferring I_μ to I_F in these notes is that $I(\cdot|\mu)$ is the large deviation rate function of the occupation measure of the Markov process $(X_t)_{t\geq 0}$ as will be seen in Section 10.

Let us introduce the 1-homogenous extension of the relative entropy H:

$$\mathrm{Ent}_\mu(f) := \int f \log f \, d\mu - \int f \, d\mu \, \log \int f \, d\mu,$$

for all nonnegative function f. The following relation holds:

$$\mathrm{Ent}_\mu(f) = \int f \, d\mu \; H\left(\frac{f\mu}{\int f \, d\mu} \middle| \mu \right).$$

As usual, $\mathrm{Var}_\mu(f) := \int_{\mathcal{X}} f^2 \, d\mu - (\int_{\mathcal{X}} f \, d\mu)^2$.

Definition 8.8 (General Poincaré and logarithmic Sobolev inequalities)**.**

(1) A probability $\mu \in \mathrm{P}(\mathcal{X})$ is said to satisfy the Poincaré inequality with constant C if

$$\mathrm{Var}_\mu(f) \le C I_\mu(f^2)$$

for any function $f \in \mathbb{D}(\mathcal{E})$.

(2) A probability μ on $\mathcal{X}$ is said to satisfy the logarithmic Sobolev inequality with a constant $C > 0$, if

$$H(f\mu|\mu) \le C I(f\mu|\mu)$$

holds for all $f : \mathcal{X} \to \mathbb{R}^+$. Equivalently, μ verifies this inequality if

$$\mathrm{Ent}_\mu(f^2) \le C I_\mu(f^2)$$

for any function $f \in \mathbb{D}(\mathcal{E})$.

In the special important case where $\mathcal{L} = \Delta - \nabla V \cdot \nabla$, the Fisher information is given by (37) and we say that the corresponding Poincaré and logarithmic Sobolev inequalities are *usual*.

Definition 8.9 (Usual Poincaré and logarithmic Sobolev inequalities, $\mathbf{P}(C)$ and $\mathbf{LS}(C)$)**.**

(1) A probability $\mu \in \mathrm{P}(\mathcal{X})$ is said to satisfy the (usual) Poincaré inequality with constant C, $\mathbf{P}(C)$ for short, if

$$\mathrm{Var}_\mu(f) \le C \int |\nabla f|^2 \, d\mu$$

for any function smooth enough function f.

(2) A probability μ on $\mathcal{X}$ is said to satisfy the (usual) logarithmic Sobolev inequality with a constant $C > 0$, $\mathbf{LS}(C)$ for short, if

$$\mathrm{Ent}_\mu(f^2) \le C \int |\nabla f|^2 \, d\mu \tag{38}$$

for any function smooth enough function f.

Remark 8.10 (Spectral gap). The Poincaré inequality $\mathbf{P}(C)$ can be made precise by means of the Dirichlet form $\mathcal{E}$:

$$\mathrm{Var}_\mu(f) \le C \, \mathcal{E}(f, f), \; \forall f \in \mathbb{D}_2(\mathcal{L})$$

for some finite $C \ge 0$. The best constant C in the above Poincaré inequality is the inverse of the spectral gap of $\mathcal{L}$.

8.3. **Links with Poincaré inequalities.** We noticed in Section 1 that $\mathbf{T}_2$ is stronger than $\mathbf{T}_1$. The subsequent proposition enables us to make precise the gap between these two inequalities.

Proposition 8.11. *Let μ be a probability measure on $\mathbb{R}^k$ and d be the Euclidean distance ; if μ verifies the inequality $\mathcal{T}_{\theta(d)} \leq H$ with a function $\theta(t) \geq t^2/C$ near 0 with $C > 0$, then μ verifies the Poincaré inequality $\mathbf{P}(C/2)$.*

So in particular, $\mathbf{T}_2$ implies Poincaré inequality with the constant $C/2$, while $\mathbf{T}_1$ doesn't imply it. The result above was established by Otto and Villani in [89]. Below is a proof using the Hamilton-Jacobi semigroup.

Proof. According to Corollary 3.3, for all bounded continuous function $f : \mathbb{R}^k \to \mathbb{R}$, $\int e^{Rf}\, d\mu \leq e^{\int f\, d\mu}$, where $Rf(x) = \inf_{y\in\mathbb{R}^k}\{f(y)+\theta(|x-y|_2)\}$. For all $t > 0$, define $R_t f(x) = \inf_{y\in\mathbb{R}^k}\{f(y)+\frac{1}{t}\theta(|y-x|_2)\}$. Suppose that $\theta(u) \geq u^2/C$, for all $0 \leq u \leq r$, for some $r > 0$. If $M > 0$, is such that $|f(x)| \leq M$ for all $x \in \mathbb{R}^k$, then it is not difficult to see that the infimum in the definition of $R_t f(x)$ is attained in the ball of center x and radius r as soon as $t \leq \theta(r)/(2M)$. So, for all $t \leq \theta(r)/(2M)$,

$$R_t f(x) = \inf_{|y-x|\leq r}\{f(y) + \frac{1}{t}\theta(|y-x|_2)\} \geq \inf_{|y-x|\leq r}\{f(y) + \frac{1}{Ct}|y-x|_2^2\} \geq Q_t f(x),$$

with $Q_t f(x) = \inf_{y\in\mathbb{R}^k}\{f(y) + \frac{1}{tC}|x-y|_2^2\}$, $t > 0$. Consequently, the inequality

$$\int e^{tQ_t f}\, d\mu \leq e^{t\int f\, d\mu} \tag{39}$$

holds for all $t \geq 0$ small enough.

If f is smooth enough (say of class $\mathcal{C}^2$), then defining $Q_0 f = f$, the function $(t,x) \mapsto Q_t f(x)$ is solution of the Hamilton-Jacobi partial differential equation :

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) + \frac{C}{4}|\nabla_x u|^2(t,x) = 0, & t \geq 0, x \in \mathbb{R}^k \\ u(0,x) = f, & x \in \mathbb{R}^k \end{cases} \tag{40}$$

(see for example [104, Theorem 22.46]).

So if f is smooth enough, it is not difficult to see that

$$\int e^{tQ_t f}\, d\mu = 1 + t\int f\, d\mu + \frac{t^2}{2}\left(\int f^2\, d\mu - \frac{C}{2}\int |\nabla f|^2\, d\mu\right) + o(t^2).$$

So (39) implies that $\mathrm{Var}_\mu(f) \leq \frac{C}{2}\int |\nabla f|^2\, d\mu$, which completes the proof. □

8.4. **Around Otto-Villani theorem.** We now present the famous Otto-Villani theorem and its generalizations.

Theorem 8.12 (Otto-Villani)**.** *Let μ be a probability measure on $\mathbb{R}^k$. If μ verifies the logarithmic Sobolev inequality $\mathbf{LS}(C)$ with a constant $C > 0$, then it verifies the inequality then $\mathbf{T}_2(C)$.*

Let us mention that Otto-Villani theorem is also true on a Riemannian manifold. This result was conjectured by Bobkov and Götze in [17] and first proved by Otto and Villani in [89]. The proof by Otto and Villani was a rather sophisticated combination of optimal transport and partial differential equation results. It was adapted to other situations (in particular to path spaces) by Wang in [107, 110, 109]. Soon after [89], Bobkov, Gentil and Ledoux have proposed in [16] a much more elementary proof relying on simple computations on the Hamilton-Jacobi semigroup. This approach is at the origin of many subsequent developments (see for instance [48], [47], [25], [73] or [54]). In [52], Gozlan gives yet another proof which is build on the characterization of dimension-free Gaussian concentration exposed in Corollary 5.5. It is very robust and works as well if $\mathbb{R}^k$ is replaced by an (almost) arbitrary polish space (see [52, Theorems 4.9 and 4.10]).

First proof of Theorem 8.12 following [16]. In this proof we explain the Hamilton-Jacobi method of Bobkov, Gentil and Ledoux. Consider the semigroup Q_t defined for all bounded function f by

$$Q_tf(x) = \inf_{y\in\mathbb{R}^k}\{f(y) + \frac{1}{Ct}|x-y|^2\},\ t>0, \qquad Q_0f = f.$$

If f is smooth enough, $(t,x)\mapsto Q_tf(x)$ solves the Hamilton-Jacobi equation (40).

According to (38), and (40)

(41)

$$\mathrm{Ent}_\mu(e^{tQ_tf}) \le \frac{Ct^2}{4}\int |\nabla_x Q_tf|^2(t,x)e^{tQ_tf(t,x)}\,d\mu(x) = -t^2\int \frac{\partial Q_tf}{\partial t}(t,x)e^{tQ_tf(t,x)}\,d\mu(x).$$

Let $Z_t = \int e^{tQ_tf(t,x)}\,d\mu(x)$, $t\ge 0$; then

$$Z_t' = \int Q_tf(t,x)e^{tQ_tf(t,x)}\,d\mu(x) + t\int \frac{\partial Q_tf}{\partial t}(t,x)e^{tQ_tf(t,x)}\,d\mu(x).$$

Consequently,

$$\begin{aligned}\mathrm{Ent}_\mu(e^{tQ_tf}) &= t\int Q_tf(t,x)e^{tQ_tf(t,x)}\,d\mu(x) - \int e^{tQ_tf(t,x)}\,d\mu(x)\log\int e^{tQ_tf(t,x)}\,d\mu(x)\\ &= tZ_t' - Z_t\log Z_t - t^2\int \frac{\partial Q_tf}{\partial t}(t,x)e^{tQ_tf(t,x)}\,d\mu(x).\end{aligned}$$

This together with (41), yields $tZ_t' - Z_t\log Z_t \le 0$ for all $t\ge 0$. In other words, the function $t\mapsto \frac{\log Z_t}{t}$ is decreasing on $(0,+\infty)$. As a result,

$$\log Z_1 = \log\int e^{Q_1f}\,d\mu \le \lim_{t\to 0}\frac{\log Z_t}{t} = \int f\,d\mu,$$

which is Bobkov-Götze dual version of $\mathbf{T}_2(C)$ stated in Corollary 3.3. □

Second proof of Theorem 8.12 following [52]. Now let us explain how to use concentration to prove Otto-Villani theorem. First let us recall the famous Herbst argument. Take g a 1-Lipschitz function such that $\int g\,d\mu = 0$ and apply (38) to $f = e^{tg/2}$, with $t\ge 0$; then letting $Z_t = \int e^{tg}\,d\mu$, one gets

$$Z_t' - Z_t\log Z_t \le \frac{Ct^2}{4}\int |\nabla g|^2 e^{tg}\,d\mu \le \frac{C}{4}t^2 Z_t,\ t>0$$

where the inequality follows from the fact that g is 1-Lipschitz. In other word,

$$\frac{d}{dt}\left(\frac{\log Z_t}{t}\right) \leq \frac{C}{4},\ t > 0.$$

Since $\frac{\log Z_t}{t} \to 0$ when $t \to 0$, integrating the inequality above yields

$$\int e^{tg}\, d\mu \leq e^{\frac{C}{4}t^2},\ t > 0.$$

Since this holds for all centered and 1-Lipschitz function g, one concludes from Corollary 3.4 that μ verifies the inequality $\mathbf{T}_1(C)$ on $(\mathbb{R}^k, |\cdot|_2)$.

The next step is a tensorization argument. Let us recall that the logarithmic Sobolev inequality enjoys the following well known tensorization property : if μ verifies $\mathbf{LS}(C)$ on $\mathbb{R}^k$, then for all positive integer n, the product probability measure μ^n satisfies $\mathbf{LS}(C)$ on $\left(\mathbb{R}^k\right)^n$. As a consequence, the argument above shows that for all positive integer n, μ^n verifies the inequality $\mathbf{T}_1(C)$ on $\left(\left(\mathbb{R}^k\right)^n, |\cdot|_2\right)$. According to Marton's argument (Theorem 4.2), there is some constant $K > 0$ such that for all positive integer n and all $A \subset \left(\mathbb{R}^k\right)^n$, with $\mu^n(A) \geq 1/2$, it holds

$$\mu^n(A^r) \geq 1 - Ke^{-Cr^2},\ r \geq 0.$$

The final step is given by Corollary 5.5 : this dimension-free Gaussian concentration inequality implies $\mathbf{T}_2(C)$ and this completes the proof. □

Otto-Villani theorem admits the following natural extension which appears in [16] and [47, Theorem 2.10].

For all $p \in [1,2]$, define $\theta_p(x) = x^2$ if $|x| \leq 1$ and $\frac{2}{p}x^p + 1 - \frac{2}{p}$ if $|x| \geq 1$.

Theorem 8.13. *Suppose that a probability μ on $\mathbb{R}^k$ verifies the following modified logarithmic Sobolev inequality*

$$\mathrm{Ent}_\mu(f^2) \leq C_1 \int \sum_{i=1}^k \theta_p^*\left(\frac{\partial f}{\partial x_i}\frac{1}{f}\right) f^2\, d\mu,$$

for all $f : \mathbb{R}^k \to \mathbb{R}$ smooth enough, where θ_p^* is the convex conjugate of θ_p. Then there is a constant C_2 such that μ verifies the transport-entropy inequality $\mathcal{T}_{\theta_p(|\cdot|_2)} \leq C_2 H$.

The theorem above is stated in a very lazy way ; the relation between C_1 and C_2 is made clear in [47, Theorem 2.10].

Sketch of proof. We shall only indicate that two proofs can be made. The first one uses the following Hamilton Jacobi semigroup

$$Q_t f(x) = \inf_{y \in \mathbb{R}^k}\left\{f(y) + t\sum_{i=1}^k \theta_p\left(\frac{y_i - x_i}{t}\right)\right\},$$

which solves the following Hamilton-Jacobi equation

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) + \sum_{i=1}^k \theta_p^*\left(\frac{\partial u}{\partial x_i}\right)(t,x) = 0, & t \geq 0, x \in \mathbb{R}^k \\ u(0,x) = f, & x \in \mathbb{R}^k \end{cases}$$

The second proof uses concentration : according to a result by Barthe and Roberto [10, Theorem 27], the modified logarithmic Sobolev inequality implies dimension-free concentration for the enlargement $\mathrm{enl}_{\theta_p}(A,r) = \{x \in (\mathbb{R}^k)^n ; \inf_{y\in A} \sum_{i=1}^n \theta_p(|x-y|_2) \le r\}$ and according to Theorem 5.4, this concentration property implies the transport-entropy inequality $\mathcal{T}_{\theta_p(|\cdot|_2)} \le CH$, for some $C>0$. □

The case $p=1$ is particularly interesting ; namely according to a result by Bobkov and Ledoux, the modified logarithmic Sobolev inequality with the function θ_1 is equivalent to Poincaré inequality (see [19, Theorem 3.1] for a precise statement). Consequently, the following results holds

Corollary 8.14. *Let μ be a probability measure on $\mathbb{R}^k$; the following propositions are equivalent:*

(1) The probability μ verifies Poincaré inequality for some constant $C>0$;
(2) The probability μ verifies the transport-entropy inequality $\mathcal{T}_c \le H$, with a cost function of the form $c(x,y) = \theta_1(a|x-y|_2)$, for some $a>0$.

Moreover the constants are related as follows: (1) implies (2) with $a = \frac{1}{\tau\sqrt{C}}$, where τ is some universal constant and (2) implies (1) with $C = \frac{1}{2a^2}$.

Again a precise result can be found in [16, Corollary 5.1].

Let us recall that μ is said to satisfy a super-Poincaré inequality if there is a decreasing function $\beta : [1,+\infty) \to [0,\infty)$ such that

$$\int f^2\,d\mu \le \beta(s)\int |\nabla f|^2\,d\mu + s\left(\int |f|\,d\mu\right)^2,\ s\ge 1$$

holds true for all sufficiently smooth f. This class of functional inequalities was introduced by Wang in [106] with applications in spectral theory. Many functional inequalities (Beckner-Latała-Oleszkiewicz inequalities for instance [65, 108]) can be represented as a super-Poincaré inequality for a specific choice of the function β. Recently efforts have been made to see which transport-entropy inequalities can be derived from super-Poincaré inequalities. We refer to Wang [109, Theorem 1.1] (in a Riemannian setting) and Gozlan [51, Theorem 5.4] for these very general extensions of Otto-Villani theorem.

8.5. **$\mathbf{T}_2$ and LS under curvature assumptions.** In [89], Otto and Villani proved that the logarithmic Sobolev inequality was sometimes implied by the inequality $\mathbf{T}_2$. The key argument for this converse is the so called HWI inequality (see [89, Theorem 3] or Corollary 7.4 of the present paper) which is recalled below. If μ is an absolutely continuous probability measure with a density of the form $d\mu(x) = e^{-V(x)}\,dx$, with V of class $\mathcal{C}^2$ on $\mathbb{R}^k$ and such that $\mathrm{Hess}\,V \ge \kappa \mathrm{Id}$, with $\kappa\in\mathbb{R}$, then for all probability measure ν on $\mathbb{R}^k$ having a smooth density with respect to μ,

$$H(\nu|\mu) \le 2W_2(\nu,\mu)\sqrt{I(\nu|\mu)} - \frac{\kappa}{2}W_2^2(\nu,\mu), \tag{42}$$

where $I(\cdot|\mu)$ is the Donsker-Varadhan information.

Proposition 8.15. *Let $d\mu(x) = e^{-V(x)}\,dx$, with V of class $\mathcal{C}^2$ on $\mathbb{R}^k$ and such that* $\operatorname{Hess} V \geq \kappa\mathrm{Id}$, *with $\kappa \leq 0$; if μ verifies the inequality $\mathbf{T}_2(C)$, with $C < -2/\kappa$, then it verifies the inequality* $\mathbf{LS}\left(\frac{4C}{(1+\kappa C/2)^2}\right)$. *In particular, when V is convex then μ verifies* $\mathbf{LS}(4C)$.

Proof. Applying (42) together with the assumed $\mathbf{T}_2(C)$ inequality, yields

$$H(\nu|\mu) \leq 2\sqrt{CH(\nu|\mu)}\sqrt{I(\nu|\mu)} - \frac{\kappa C}{2} H(\nu|\mu),$$

for all ν. Thus, if $1 + \frac{\kappa C}{2} > 0$, one has $H(\nu|\mu) \leq \frac{4C}{\left(1+\frac{\kappa C}{2}\right)^2} I(\nu|\mu)$. Taking $d\nu(x) = f^2(x)\,dx$ with a smooth f yields

$$\mathrm{Ent}_\mu(f^2) \leq \frac{4C}{\left(1+\frac{\kappa C}{2}\right)^2} \int |\nabla f|^2\,d\mu,$$

which completes the proof. $\square$

So in the range $C + 2/\kappa < 0$, $\mathbf{T}_2$ and $\mathbf{LS}$ are equivalent. In fact under the condition $\operatorname{Hess} V \geq \kappa$, a strong enough Gaussian concentration property implies the logarithmic Sobolev inequality, as shown in the following theorem by Wang [105].

Theorem 8.16. *Let $d\mu(x) = e^{-V(x)}\,dx$, with V of class $\mathcal{C}^2$ on $\mathbb{R}^k$ and such that* $\operatorname{Hess} V \geq \kappa\mathrm{Id}$, *with $\kappa \leq 0$; if there is some $C < -2/\kappa$, such that $\int e^{\frac{1}{C}d^2(x_o,x)}\,d\mu(x)$ is finite, for some (and thus all) point x_o, then μ verifies the logarithmic Sobolev inequality for some constant $\tilde{C}$.*

Recently Barthe and Kolesnikov have generalized Wang's theorem to different functional inequalities and other convexity defects [8]. Their proofs rely on Theorem 7.1. A drawback of Theorem 8.16 is that the constant $\tilde{C}$ depends too heavily on the dimension k. In a series of papers [87, 86, 85], E. Milman has shown that under curvature conditions concentration inequalities and isoperimetric inequalities are in fact equivalent with a *dimension-free* control of constants. Let us state a simple corollary of Milman's results.

Corollary 8.17. *Let $d\mu(x) = e^{-V(x)}\,dx$, with V of class $\mathcal{C}^2$ on $\mathbb{R}^k$ and such that* $\operatorname{Hess} V \geq \kappa\mathrm{Id}$, *with $\kappa \leq 0$; if there is some $C < -2/\kappa$ and $M > 1$, such that μ verifies the following Gaussian concentration inequality*

$$\mu(A^r) \geq 1 - Me^{-\frac{1}{C}r^2},\ r \geq 0 \tag{43}$$

for all A such that $\mu(A) \geq 1/2$ and with $A^r = \{x \in \mathbb{R}^k; \exists y \in A, |x-y|_2 \leq r\}$, then μ verifies the logarithmic Sobolev inequality with a constant $\tilde{C}$ depending only on C, κ and M. In particular, the constant $\tilde{C}$ is independent on the dimension k of the space.

The conclusion of the preceding results is that when $C + 2/\kappa < 0$, it holds

$$\mathbf{LS}(C) \Rightarrow \mathbf{T}_2(C) \Rightarrow \text{ Gaussian concentration (43) with constant } 1/C \Rightarrow \mathbf{LS}(\tilde{C}),$$

and so these three inequalities are qualitatively equivalent in this range of parameters. Nevertheless, the equivalence between $\mathbf{LS}$ and $\mathbf{T}_2$ is no longer true when $\operatorname{Hess} V$ is unbounded from below. In [25], Cattiaux and Guillin were able to give an example of a probability μ on $\mathbb{R}$ verifying $\mathbf{T}_2$, but not $\mathbf{LS}$. Cattiaux and Guillin's counterexample is discussed in Theorem 9.5 below.

8.6. **A refined version of Otto-Villani theorem.** We close this section with a recent result by Gozlan, Roberto and Samson [54] which completes the picture showing that $\mathbf{T}_2$ (and in fact many other transport-entropy inequalities) is equivalent to a logarithmic Sobolev inequality restricted to a subclass of functions.

Let us say that a function $f : \mathbb{R}^k \to \mathbb{R}$ is λ-semiconvex, $\lambda \geq 0$, if the function $x \mapsto f(x) + \frac{\lambda}{2}|x|^2$ is convex. If f is $\mathcal{C}^2$ this is equivalent to the condition $\mathrm{Hess}\, f(x) \geq -\lambda \mathrm{Id}$. Moreover, if f is λ-semiconvex, it is almost everywhere differentiable, and for all x where $\nabla f(x)$ is well defined, one has

$$f(y) \geq f(x) + \nabla f(x) \cdot (y - x) - \frac{\lambda}{2}|y - x|^2.$$

for all $y \in \mathbb{R}^k$.

Theorem 8.18. *Let μ be a probability measure on $\mathbb{R}^k$. The following propositions are equivalent:*

(1) There exists $C_1 > 0$ such that μ verify the inequality $\mathbf{T}_2(C_1)$.
(2) There exists $C_2 > 0$ such that for all $0 \leq \lambda < \frac{2}{C_2}$ and all λ-semiconvex $f : \mathbb{R}^k \to \mathbb{R}$,

$$\mathrm{Ent}_\mu(e^f) \leq \frac{C_2}{\left(1 - \frac{\lambda C_2}{2}\right)^2} \int |\nabla f|^2 e^f \, d\mu.$$

The constants C_1 and C_2 are related in the following way:

$$(1) \Rightarrow (2) \text{ with } C_2 = C_1.$$
$$(2) \Rightarrow (1) \text{ with } C_1 = 8C_2.$$

More general results can be found in [54, Theorem 1.8]. Let us emphasize the main difference between this theorem and Proposition 8.15 : in the result above the curvature assumption is made on the functions f and not on the potential V. A nice corollary of Theorem 8.18, is the following perturbation result:

Theorem 8.19. *Let μ be a probability measure on $\mathbb{R}^k$ and consider $d\tilde{\mu}(x) = e^{\varphi(x)}\, dx$, where $\varphi : \mathbb{R}^k \to \mathbb{R}$ is bounded. If μ verifies $\mathbf{T}_2(C)$, then $\tilde{\mu}$ verifies $\mathbf{T}_2(8e^{\mathrm{Osc}(\varphi)}C)$, where $\mathrm{Osc}(\varphi) = \sup \varphi - \inf \varphi$.*

Many functional inequalities of Sobolev type enjoy the same bounded perturbation property (without the factor 8). For the Poincaré inequality or the logarithmic Sobolev inequality, the proof (due to Holley and Stroock) is almost straightforward (see e.g [2, Theorems 3.4.1 and 3.4.3]). For transport-entropy inequalities, the question of the perturbation was raised in [89] and remained open for a long time. The proof of Theorem 8.19 relies on the representation of $\mathbf{T}_2$ as a restricted logarithmic Sobolev inequality provided by Theorem 8.18. Contrary to Sobolev type inequalities, no direct proof of Theorem 8.19 is known.

9. Workable sufficient conditions for transport-entropy inequalities

In this section, we review some of the known sufficient conditions on $V : \mathbb{R}^k \to \mathbb{R}$ under which $d\mu = e^{-V}\, dx$ verifies a transport-entropy inequality of the form $\mathcal{T}_{\theta(d)} \leq H$. Unlike Section 7, the potential V is not supposed to be (uniformly) convex.

9.1. **Cattiaux and Guillin's restricted logarithmic Sobolev method.** Let μ be a probability measure on $\mathbb{R}^k$ such that $\int e^{\varepsilon|x|^2}\,d\mu(x) < +\infty$, for some $\varepsilon > 0$. Following Cattiaux and Guillin in [25], let us say that μ verifies the restricted logarithmic Sobolev inequality $\mathbf{rLS}(C,\eta)$ if

$$\mathrm{Ent}_\mu(f^2) \leq C \int |\nabla f|^2\,d\mu,$$

for all smooth $f : \mathbb{R}^k \to \mathbb{R}$ such that

$$f^2(x) \leq \left(\int f^2\,d\mu\right) e^{\eta|x_o - x|^2 + \int |x_o - y|^2\,d\mu(y)},\ x \in \mathbb{R}^k.$$

Using Bobkov-Gentil-Ledoux proof of Otto-Villani theorem, Cattiaux and Guillin obtained the following result (see [25, Theorem 1.17]).

Theorem 9.1. *Let μ be a probability measure on $\mathbb{R}^k$ such that $\int e^{\varepsilon|x|^2}\,d\mu(x) < +\infty$, for some $\varepsilon > 0$. If the restricted logarithmic Sobolev inequality $\mathbf{rLS}(C,\eta)$ holds for some $\eta < \varepsilon/2$, then μ verifies the inequality $\mathbf{T}_2(\tilde{C})$, for some $\tilde{C} > 0$.*

The interest of this theorem is that the restricted logarithmic Sobolev inequality above is strictly weaker than the usual one. Moreover, workable sufficient conditions for the **rLS** can be given. Let us start with the case of the real axis.

Theorem 9.2. *Let $d\mu(x) = e^{-V(x)}\,dx$ be a probability measure on $\mathbb{R}$ with $\int e^{\varepsilon|x|^2}\,d\mu(x) < +\infty$ for some $\varepsilon > 0$. If μ is such that*

$$A^+ = \sup_{x \geq 0} \int_x^{+\infty} t^2 e^{-V(t)}\,dt \int_0^x e^{V(t)}\,dt \quad \text{and} \quad A^- = \sup_{x \leq 0} \int_{-\infty}^x t^2 e^{-V(t)}\,dt \int_x^0 e^{V(t)}\,dt$$

are finite then μ verifies $\mathbf{rLS}(C,\eta)$, for some $C, \eta > 0$ and so it verifies also $\mathbf{T}_2(\tilde{C})$ for some $\tilde{C} > 0$.

The finiteness of A^+ and A^- can be determined using the following proposition (see [25, Proposition 5.5]).

Proposition 9.3. *Suppose that $d\mu(x) = e^{-V(x)}\,dx$ be a probability measure on $\mathbb{R}$ with V of class $\mathcal{C}^2$ such that $\frac{V''}{(V')^2}(x) \to 0$ when $x \to \infty$. If V verifies*

$$\limsup_{x \to \pm\infty} \left|\frac{x}{V'(x)}\right| < +\infty, \tag{44}$$

then A^+ and A^- are finite (and there is $\varepsilon > 0$ such that $\int e^{\varepsilon|x|^2}\,d\mu(x) < +\infty$).

The condition $\frac{V''}{(V')^2}(x) \to 0$ when $x \to \infty$ is not very restrictive and appears very often in results of this type (see [2, Corollary 6.4.2 and Theorem 6.4.3] for instance).

Now let us recall the following result by Bobkov and Götze (see [17, Theorem 5.3] and [2, Theorems 6.3.4 and 6.4.3]) dealing this time with the logarithmic Sobolev inequality.

Theorem 9.4. *Let $d\mu(x) = e^{-V(x)}\,dx$ be a probability measure on $\mathbb{R}$, and m a median of μ. If V is such that*

$$D^- = \sup_{x<m} \int_{-\infty}^{x} e^{-V(t)}\,dt \log\left(\frac{1}{\int_{-\infty}^{x} e^{-V(t)}\,dt}\right) \int_x^m e^{V(t)}\,dt$$

$$D^+ = \sup_{x>m} \int_x^{+\infty} e^{-V(t)}\,dt \log\left(\frac{1}{\int_x^{+\infty} e^{-V(t)}\,dt}\right) \int_m^x e^{V(t)}\,dt$$

are finite, then μ verifies the logarithmic Sobolev inequality, and the optimal constant C_{opt} verifies

$$\tau_1 \max(D^-, D^+) \le C_{opt} \le \tau_2 \max(D^-, D^+),$$

where τ_1 and τ_2 are known universal constants.

Moreover if V is of class $\mathcal{C}^2$ and verifies $\lim_{x\to\infty} \frac{V''}{(V')^2}(x) = 0$, then D^- and D^+ are finite if and only if V verifies the following conditions:

$$\liminf_{x\to\infty} |V'(x)| > 0 \qquad \text{and} \qquad \limsup_{x\to\infty} \frac{V(x)}{(V')^2(x)} < +\infty$$

Theorem 9.5 (Cattiaux-Guillin's counterexample)**.** *The probability measure $d\mu(x) = \frac{1}{Z}e^{-V(x)}\,dx$ defined on $\mathbb{R}$ with $V(x) = |x|^3 + 3x^2\sin^2 x + |x|^\beta$, with Z a normalizing constant and $2 < \beta < 5/2$ satisfies the inequality $\mathbf{T}_2$ but not the logarithmic Sobolev inequality.*

Proof. For all $x > 0$,

$$\begin{aligned} V'(x) &= 3x^2 + 6x\sin^2 x + 6x^2\cos x \sin x + \beta x^{\beta-1} \\ &= 3x^2(1+\cos 2x) + 6x\sin^2 x + \beta x^{\beta-1}. \end{aligned}$$

and

$$V''(x) = 6x^2\cos 2x + 6x(1 + 2\sin 2x) + 6\sin^2 x + \beta(\beta-1)x^{\beta-2}.$$

First, observe that $V'(x) > 0$ for all $x > 0$ and $V'(x) \to \infty$ when $x \to +\infty$. Moreover, for x large enough, $\left|\frac{V''}{V'^2}(x)\right| \le D\frac{x^2}{x^{2\beta-2}}$, and $0 \le \frac{x}{V'} \le D\frac{x}{x^{\beta-1}}$, for some numerical constant $D > 0$. Since, $\beta > 2$, it follows that $\frac{V''}{V'^2}(x) \to 0$, and $\frac{x}{V'(x)} \to 0$ when $x \to +\infty$. Consequently, it follows from Proposition 9.3, that μ verifies $\mathbf{T}_2(C)$, for some $C > 0$. On the other hand, consider the sequence $x_k = \frac{\pi}{4} + k\pi$, then $V'^2(x_k) = (6x_k + \beta x_k^{\beta-1})^2 \sim \beta^2(\pi k)^{2\beta-2}$ and $V(x_k) \sim (k\pi)^3$. So $\frac{V(x_k)}{V'^2(x_k)} \sim \beta^2(\pi k)^{5-2\beta}$, and since $\beta < 5/2$, one concludes that $\limsup_{x\to+\infty} \frac{V}{V'^2}(x) = +\infty$. According to Theorem 9.4, it follows that μ does not verify the logarithmic- Sobolev inequality. □

Recently, Cattiaux, Guillin and Wu have obtained in [26] different sufficient conditions for the restricted logarithmic Sobolev inequality **rLS** in dimension $k \ge 1$.

Theorem 9.6. *Let μ be a probabilikty measure on $\mathbb{R}^k$ with a density of the form $d\mu(x) = e^{-V(x)}\,dx$, with $V : \mathbb{R}^k \to \mathbb{R}$ of class $\mathcal{C}^2$. If one of the following conditions*

$$\exists a < 1, R, c > 0, \text{ such that } \forall |x| > R, \qquad (1-a)|\nabla V(x)|^2 - \Delta V(x) \geq c|x|^2$$

or

$$\exists R, c > 0, \text{such that } \forall |x| > R, \qquad x \cdot \nabla V(x) \geq c|x|^2$$

is satisfied, then **rLS** *holds.*

We refer to [26, Corollary 2.1] for a proof relying on the so called Lyapunov functions method.

9.2. **Contraction methods.** In [50], Gozlan recovered Cattiaux and Guillin's sufficient condition (44) for $\mathbf{T}_2$ and extended it to other transport-entropy inequalities on the real axis. The proof relies on a simple contraction argument, we shall now explain it in a general setting.

Contraction of transport-entropy inequalities. In the sequel, $\mathcal{X}$ and $\mathcal{Y}$ will be polish spaces. If μ is a probability measure on $\mathcal{X}$ and $T : \mathcal{X} \to \mathcal{Y}$ is a measurable map, the image of μ under T will be denoted by $T_\#\mu$; by definition, it is the probability measure on $\mathcal{Y}$ defined by

$$T_\#\mu(A) = \mu\left(T^{-1}(A)\right),$$

for all measurable subset A of Y.

The result below shows that if μ verifies a transport-entropy inequality on $\mathcal{X}$ then the image $T_\#\mu$ verifies verifies a transport-entropy inequality on $\mathcal{Y}$ with a new cost function expressed in terms on T.

Theorem 9.7. *Let μ_o be a probability measure on $\mathcal{X}$ and $T : \mathcal{X} \to \mathcal{Y}$ be a measurable bijection. If μ_o satisfies the transport-entropy inequality $\alpha(\mathcal{T}_c) \leq H$ with a cost function c on $\mathcal{X}$, then $T_\#\mu_o$ satisfies the transport-entropy inequality $\alpha(\mathcal{T}_{c^T}) \leq H$ with the cost function c^T defined on $\mathcal{Y}$ by*

$$c^T(y_1, y_2) = c(T^{-1}y_1, T^{-1}y_2),\ y_1, y_2 \in \mathcal{Y}.$$

Proof. Let us define $Q(y_1, y_2) = (T^{-1}y_1, T^{-1}y_2),\ y_1, y_2 \in \mathcal{Y}$, and $\mu_1 = T_\#\mu_o$. Let $\nu \in \mathrm{P}(\mathcal{Y})$ and take $\pi \in \Pi(\nu, \mu_1)$, the subset of $\mathrm{P}(\mathcal{Y}^2)$ consisting of the probability π with their marginal measures $\pi_0 = \nu$ and $\pi_1 = \mu_1$. Then $\int c^T(y_1, y_2)\,d\pi = \int c(x, y)\,dQ_\#\pi$, so

$$\mathcal{T}_{c^T}(\nu, \mu_1) = \inf_{\pi \in Q_\#\Pi(\nu,\mu_1)} \int c(x, y) d\pi.$$

But it is easily seen that $Q_\#\Pi(\nu, \mu_1) = \Pi(T^{-1}{}_\#\nu, \mu_o)$. Consequently

$$\mathcal{T}_{c^T}(\nu, \mu_1) = \mathcal{T}_c(T^{-1}{}_\#\nu, \mu_o).$$

Since μ_o satisfies the transport-entropy inequality $\alpha(\mathcal{T}_c) \leq H$, it holds

$$\alpha\left(\mathcal{T}_c(T^{-1}{}_\#\nu, \mu_o)\right) \leq H(T^{-1}{}_\#\nu|\mu_o).$$

But it is easy to check , with Proposition B.1 and the fact that T is one-one, that

$$H(T^{-1}{}_\#\nu|\mu_o) = H(\nu|T_\#\mu_o).$$

Hence

$$\alpha\left(\mathcal{T}_{c^T}(\nu, \mu_1)\right) \leq H(\nu|\mu_1),$$

for all $\nu \in \mathrm{P}(Y)$. $\square$

Remark 9.8. This contraction property was first observed by Maurey (see [82, Lemma 2]) in the context of inf-convolution inequalities. Theorem 9.7 is a simple but powerful tool to derive new transport-entropy inequalities from already known ones.

Sufficient conditions on $\mathbb{R}$. Let us recall that a probability measure μ on $\mathbb{R}$ is said to satisfy Cheeger inequality with the constant $\lambda > 0$ if

$$\int |f(x) - m(f)|\, d\mu(x) \leq \lambda \int |f'(x)|\, d\mu(x), \tag{45}$$

for all $f : \mathbb{R} \to \mathbb{R}$ sufficiently smooth, where $m(f)$ denotes a median of f under μ.

Using the contraction argument presented above, Gozlan obtained the following theorem ([50, Theorem 2]).

Theorem 9.9. *Let $\theta : [0, \infty) \to [0, \infty)$ be such that $\theta(t) = t^2$ for all $t \in [0, 1]$, $t \mapsto \frac{\theta(t)}{t}$ is increasing and $\sup_{t>0} \frac{\theta(2t)}{\theta(t)} < +\infty$ and let μ be a probability measure on $\mathbb{R}$ which verifies Cheeger inequality for some $\lambda_o > 0$. The following propositions are equivalent :*

(1) The probability measure μ verifies the transport cost inequality $\mathcal{T}_\theta \leq CH$, for some $C > 0$.
(2) The constants $K^+(\varepsilon)$ and $K^-(\varepsilon)$ defined by

$$K^+(\varepsilon) = \sup_{x \geq m} \frac{\int_x^{+\infty} e^{\varepsilon\theta(u-x)}\, d\mu(u)}{\mu[x, +\infty)} \quad \textit{and} \quad K^-(\varepsilon) = \sup_{x \leq m} \frac{\int_{-\infty}^{x} e^{\varepsilon\theta(x-u)}\, d\mu(u)}{\mu(-\infty, x]}$$

are finite for some $\varepsilon > 0$, where m denotes the median of μ.

The condition K^+ and K^- finite is always necessary to have the transport-entropy inequality (see [50, Corollary 15]). This condition is sufficient if Cheeger inequality holds. Cheeger inequality is slightly stronger than Poincaré inequality. On the other hand transport-entropy inequalities of the form $\mathcal{T}_\theta \leq H$, with a function θ as above, imply Poincaré inequality (Theorem 8.11). So Theorem 9.9 offers a characterization of transport-entropy inequalities except perhaps on the "small" set of probability measures verifying Poincaré but not Cheeger inequality.

Sketch of proof. We will only prove the sufficiency of the condition K^+ and K^- finite. Moreover, to avoid technical difficulties, we shall only consider the case $\theta(t) = t^2$. Let $d\mu_o(x) = \frac{1}{2}e^{-|x|}\, dx$ be the two-sided exponential measure on $\mathbb{R}$. According to a result by Talagrand [102] the probability μ_o verifies the transport-entropy inequality $\mathcal{T}_{c_o} \leq C_o H$, with the cost function $c_o(x, y) = \min(|x - y|^2, |x - y|)$, for some $C_o > 0$.

Consider the cumulative distribution functions of μ and μ_o defined by $F(x) = \mu(-\infty, x]$ and $F_o(x) = \mu_o(-\infty, x]$, $x \in \mathbb{R}$. The monotone rearrangement map $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = F^{-1} \circ F_o$, see (22), transports the probability μ_o onto the probability μ : $T_\# \mu_o = \mu$. Consequently, by application of the contraction Theorem 9.7, the probability μ verifies the transport-cost inequality $\mathcal{T}_{c_o^T} \leq C_o H$, with $c_o^T(x - y) = c_o(T^{-1}(x) - T^{-1}(y))$. So, all we have to do is to show that there is some constant $a > 0$ such that $c_o(T^{-1}(x) - T^{-1}(y)) \geq \frac{1}{a^2}|x-y|^2$, for all $x, y \in \mathbb{R}$. This condition is equivalent to the following

$$|T(x) - T(y)| \leq a \min(|x-y|, |x-y|^{1/2}),\ x, y \in \mathbb{R}.$$

In other words, we have to show that T is a-Lipschitz and a-Hölder of order $1/2$.

According to a result by Bobkov and Houdré, μ verifies Cheeger inequality (45) with the constant λ_o if and only if T is λ_o-Lipschitz (see [18, Theorem 1.3]).

To deal with the Hölder condition, observe that if T is a-Hölder on $[0, \infty)$ and on $\mathbb{R}^-$, then it is $\sqrt{2}a$-Hölder on $\mathbb{R}$. Let us treat the case of $[0, \infty)$, the other case being similar. The condition T is a-Hölder on $[0, \infty)$ is equivalent to

$$T^{-1}(x+u) - T^{-1}(x) \geq \frac{u^2}{a^2},\ x \geq m, u \geq 0.$$

But a simple computation gives : $T^{-1}(x) = -\log(2(1-F(x)))$, for all $x \geq m$. So the condition above reads

$$\frac{1-F(x+u)}{1-F(x)} \leq e^{-\frac{u^2}{a^2}},\ x \geq m, u \geq 0. \tag{46}$$

Since, $K^+(\varepsilon) = \sup_{x\geq m} \frac{\int_x^{+\infty} e^{\varepsilon(u-x)^2}\, d\mu(u)}{\mu[x,+\infty)}$ is finite, an application of Markov inequality yields

$$\frac{1-F(x+u)}{1-F(x)} \leq K^+(\varepsilon)e^{-\varepsilon u^2},\ x \geq m, u \geq 0.$$

On the other hand the Lipschitz continuity of T can be written

$$\frac{1-F(x+u)}{1-F(x)} \leq e^{-\frac{u}{\lambda_o}},\ x \geq m, u \geq 0.$$

So, if $a > 0$ is chosen so that $\frac{u^2}{a^2} \leq \max\left(\frac{u}{\lambda_o}, \varepsilon u^2 - \log K^+(\varepsilon)\right)$, then (46) holds and this completes the proof. □

The following corollary gives a concrete criterion to decide whether a probability measure on $\mathbb{R}$ verifies a given transport-entropy inequality. It can be deduced from Theorem 9.9 thanks to an estimation of the integrals defining K^+ and K^-. We refer to [50] for this technical proof.

Corollary 9.10. *Let $\theta : [0, \infty) \to [0, \infty)$ of class $\mathcal{C}^2$ be as in Theorem 9.9 and let μ be a probability measure on $\mathbb{R}$ with a density of the form $d\mu(x) = e^{-V(x)}\, dx$, with V of class $\mathcal{C}^2$. Suppose that $\frac{\theta''}{\theta'^2}(x) \to 0$ and $\frac{V''}{V'^2}(x) \to 0$ when $x \to \infty$. If there is some $a > 0$ such that*

$$\limsup_{x\to\pm\infty} \frac{\theta'(a|x|)}{|V'(m+x)|} < +\infty,$$

with m the median of μ, then μ verifies the transport-entropy inequality $\mathcal{T}_\theta \leq CH$, for some $C > 0$.

Note that this corollary generalizes Cattiaux and Guillin's condition (44).

Poincaré inequalities for non-Euclidean metrics. Our aim is now to partially generalize to the multidimensional case the approach explained in the preceding section. The two main ingredients of the proof of Theorem 9.9 were the following :

- The fact that $d\mu_o(x) = \frac{1}{2}e^{-|x|}\,dx$ verifies the transport-entropy inequality $\mathcal{T}_c \le CH$ with the cost function $c(x,y) = \min(|x-y|^2, |x-y|)$. Let us define the cost function $c_1(x,y) = \min(|x-y|_2, |x-y|_2^2)$ on $\mathbb{R}^k$ equipped with its Euclidean distance. We have seen in Corollary 8.14 that a probability measure on $\mathbb{R}^k$ verifies the transport-entropy inequality $\mathcal{T}_{c_1} \le C_1 H$ for some $C_1 > 0$ if and only if it verifies Poincaré inequality with a constant $C_2 > 0$ related to C_1.
- The fact that the application T sending μ_o on μ was both Lipschitz and 1/2-Hölder. Consequently, the application $\omega = T^{-1}$ which maps μ on μ_o, behaves like x for small values of x and like x^2 for large values of x.

So we can combine the two ingredients above by saying that "the image of μ by an application ω which resembles $\pm\max(|x|, |x|^2)$ verifies Poincaré inequality." It appears that this gets well in higher dimension and gives a powerful way to prove transport-entropy inequalities.

Let us introduce some notation. In the sequel, $\omega : \mathbb{R} \to \mathbb{R}$ will denote an application such that $x \mapsto \omega(x)/x$ is increasing on $(0, +\infty)$, $\omega(x) \ge 0$ for all $x \ge 0$, and $\omega(-x) = -\omega(x)$ for all $x \in \mathbb{R}$. It will be convenient to keep the notation ω to denote the application $\mathbb{R}^k \to \mathbb{R}^k$: $(x_1, \ldots, x_k) \mapsto (\omega(x_1), \ldots, \omega(x_k))$. We will consider the metric d_ω defined on $\mathbb{R}^k$ by

$$d_\omega(x,y) = |\omega(x) - \omega(y)|_2 = \sqrt{\sum_{i=1}^k |\omega(x_i) - \omega(y_i)|^2},\ x, y \in \mathbb{R}^k.$$

Theorem 9.11. *Let μ be a probability measure on $\mathbb{R}^k$. The following statements are equivalent.*

(1) The probability $\tilde{\mu} = \omega_\#\mu$ verifies Poincaré inequality with the constant C:

$$\mathrm{Var}_{\tilde{\mu}}(f) \le C \int |\nabla f|_2^2\, d\tilde{\mu},$$

for all $f : \mathbb{R}^k \to \mathbb{R}$ smooth enough.

(2) The probability μ verifies the following weighted Poincaré inequality with the constant $C > 0$:

$$\mathrm{Var}_\mu(f) \le C \int \sum_{i=1}^k \frac{1}{\omega'(x_i)^2} \left(\frac{\partial f}{\partial x_i}(x)\right)^2 d\mu(x), \tag{47}$$

for all $f : \mathbb{R}^k \to \mathbb{R}$ smooth enough.

(3) The probability μ verifies the transport-entropy inequality $\mathcal{T}_c \le H$, with the cost function $c(x,y) = \theta_1(a d_\omega(x,y))$ for some $a > 0$, with $\theta_1(t) = \min(t^2, t)$, $t \ge 0$. More precisely,

$$\inf_{\pi:\pi_0=\nu,\pi_1=\mu} \int_{\mathbb{R}^k\times\mathbb{R}^k} \min\left(a^2|\omega(x)-\omega(y)|_2^2, a|\omega(x)-\omega(y)|_2\right) d\pi(x,y) \le H(\nu|\mu), \tag{48}$$

for all $\nu \in \mathrm{P}(\mathbb{R}^k)$.

The constants C and a are related in the following way: (1) implies (3) with $a = \frac{1}{\tau\sqrt{C}}$, where τ is a universal constant, and (3) implies (1) with $C = \frac{1}{2a^2}$.

Proof. The equivalence between (1) and (2) is straightforward.
Let us show that (1) implies (3). Indeed, according to Corollary 8.14, $\tilde{\mu}$ verifies the transport-entropy inequality $\mathcal{T}_{\tilde{c}} \leq H$ with $\tilde{c}(x,y) = \theta_1(a|x-y|_2)$, and $a = \frac{1}{\tau\sqrt{C}}$. Consequently, according to the contraction Theorem 9.7, μ which is the image of $\tilde{\mu}$ under the map ω^{-1} verifies the transport-entropy inequality $\mathcal{T}_c \leq H$ where $c(x,y) = \tilde{c}(\omega(x),\omega(y)) = \theta_1(ad_\omega(x,y))$. The proof of the converse is similar. □

Definition 9.12. *When μ verifies (47), one says that the inequality $\mathbf{P}(\omega, C)$ holds.*

Remark 9.13. If $f : \mathbb{R}^k \to \mathbb{R}$ let us denote by $|\nabla f|_\omega(x)$ the "length of the gradient" of f at point x with respect to the metric d_ω defined above. By definition,

$$|\nabla f|_\omega(x) = \limsup_{y\to x} \frac{|f(y)-f(x)|}{d_\omega(x,y)},\ x \in \mathbb{R}^k.$$

It is not difficult to see that μ verifies the inequality $\mathbf{P}(\omega, C)$ if and only if it verifies the following Poincaré inequality

$$\mathrm{Var}_\mu(f) \leq C \int |\nabla f|_\omega^2\, d\mu,$$

for all f smooth enough. So, the inequality $\mathbf{P}(\omega, \cdot\,)$ is a true Poincaré inequality for the non-Euclidean metric d_ω.

So according to Theorem 9.11, the Poincaré inequality (47) is qualitatively equivalent to the transport cost inequality (48). Those transport-entropy inequalities are rather unusual, but can be compared to more classical transport-entropy inequalities using the following proposition.

Proposition 9.14. *The following inequality holds*

$$\theta_1(ad_\omega(x,y)) \geq \theta_1\left(\frac{a}{\sqrt{k}}\right) \sum_{i=1}^k \theta_1 \circ \omega\left(\frac{|x_i - y_i|}{2}\right),\ x,y \in \mathbb{R}^k. \tag{49}$$

We skip the technical proof of this inequality and refer to [51, Lemma 2.6 and Proof of Proposition 4.2]. Let us emphasize an important particular case. In the sequel, $\omega_2 : \mathbb{R} \to \mathbb{R}$ will be the function defined by $\omega_2(x) = \max(x, x^2)$, for all $x \geq 0$ and such that $\omega_2(-x) = -\omega_2(x)$, for all $x \in \mathbb{R}$.

Corollary 9.15. *If a probability measure μ on $\mathbb{R}^k$ verifies the inequality $\mathbf{P}(\omega_2, C)$ for some $C > 0$ then it verifies the inequality $\mathbf{T}_2(4\omega_2(\tau\sqrt{kC}))$, where τ is some universal constant.*

In other words, a sufficient condition for μ to verify $\mathbf{T}_2$ is that the image of μ under the map ω_2 verifies Poincaré inequality. We do not know if this condition is also necessary.

Proof. According to Theorem 9.11, if μ verifies $\mathbf{P}(\omega_2, C)$ then it verifies the transport-entropy inequality $\mathcal{T}_c \leq H$ with the cost function $c(x,y) = \theta_1(ad_{\omega_2}(x,y))$, with $a = \frac{1}{\tau\sqrt{C}}$. According to (49), one has

$$\theta_1(ad_{\omega_2}(x,y)) \geq \theta_1\left(\frac{a}{\sqrt{k}}\right)\sum_{i=1}^{k}\theta_1\circ\omega_2\left(\frac{|x_i-y_i|}{2}\right) = \frac{\theta_1\left(\frac{1}{\tau\sqrt{kC}}\right)}{4}|x-y|_2^2,$$

since $\theta_1\circ\omega_2(t)=t^2$, for all $t\in\mathbb{R}$. Observing that $\frac{1}{\theta_1(1/t)}=\omega_2(t)$, $t>0$, one concludes that μ verifies the inequality $\mathbf{T}_2(4\omega_2(\tau\sqrt{kC}))$, which completes the proof. □

Poincaré inequality has been deeply studied by many authors and several necessary or sufficient conditions are now available for this functional inequality. Using the equivalence

$$\mu \text{ verifies } \mathbf{P}(\omega,C) \Leftrightarrow \omega_{\#}\mu \text{ verifies } \mathbf{P}(C), \tag{50}$$

it is an easy job to convert the known criteria for Poincaré inequality into criteria for the $\mathbf{P}(\omega,\cdot)$ inequality.

In dimension one, one has a necessary and sufficient condition.

Proposition 9.16. *An absolutely continuous probability measure μ on $\mathbb{R}$ with density $h>0$ satisfies the inequality $\mathbf{P}(\omega,C)$ for some $C>0$ if and only if*

$$D_\omega^- = \sup_{x\leq m}\mu(-\infty,x]\int_x^m \frac{\omega'(u)^2}{h(u)}\,du < +\infty \quad \text{and} \quad D_\omega^+ = \sup_{x\geq m}\mu[x,+\infty)\int_m^x \frac{\omega'(u)^2}{h(u)}\,du < +\infty, \tag{51}$$

where m denotes the median of μ. Moreover the optimal constant C denoted by C_{opt} verifies

$$\max(D_\omega^-,D_\omega^+) \leq C_{\mathrm{opt}} \leq 4\max(D_\omega^-,D_\omega^+).$$

Proof. This proposition follows at once from the celebrated Muckenhoupt condition for Poincaré inequality (see [88]). According to Muckenhoupt condition, a probability measure $d\nu = h\,dx$ having a positive density with respect to Lebesgue measure, satisfies Poincaré inequality if and only if

$$D^- = \sup_{x\leq m}\nu(-\infty,x]\int_x^m \frac{1}{h(u)}\,du < +\infty \quad \text{and} \quad D^+ = \sup_{x\geq m}\nu[x,+\infty)\int_m^x \frac{1}{h(u)}\,du < +\infty,$$

and the optimal constant C_{opt} verifies $\max(D^-,D^+) \leq C_{opt} \leq 4\max(D^-,D^+)$. Now, according to (50), μ satisfies $\mathbf{P}(\omega,C)$ if and only if $\tilde{\mu}=\omega_{\#}\mu$ satisfies Poincaré inequality with the constant C. The density of $\tilde{\mu}$ is $\tilde{h}=\frac{h\circ\omega^{-1}}{\omega'\circ\omega^{-1}}$. Plugging $\tilde{h}$ into Muckenhoupt conditions immediately gives us the announced result. □

Estimating the integrals defining D^- and D^+ by routine arguments, one can obtain the following workable sufficient conditions (see [51, Proposition 3.3] for a proof).

Proposition 9.17. *Let μ be an absolutely continuous probability measure on $\mathbb{R}$ with density $d\mu(x)=e^{-V(x)}\,dx$. Assume that the potential V is of class $\mathcal{C}^1$ and that ω verifies the following regularity condition:*

$$\frac{\omega''(x)}{\omega'^2(x)} \xrightarrow[x\to+\infty]{} 0.$$

If V is such that

$$\limsup_{x\to\pm\infty}\frac{|\omega'(x)|}{|V'(x)|} < +\infty,$$

then the probability measure μ verifies the inequality $\mathbf{P}(\omega, C)$ for some $C > 0$.

Observe that this proposition together with the inequality (49) furnishes another proof of Corollary 9.10 and enables us to recover (as a particular instance, taking $\omega = \omega_2$) Cattiaux and Guillin's condition for $\mathbf{T}_2$.

In dimension k, it is well known that a probability $d\nu(x) = e^{-W(x)}\,dx$ on $\mathbb{R}^k$ satisfies Poincaré inequality if W verifies the following condition:

$$\liminf_{|x|\to+\infty} \frac{1}{2}|\nabla W|_2^2(x) - \Delta W(x) > 0.$$

This condition is rather classical in the functional inequality literature. The interested reader can find a nice elementary proof in [3]. Using (50) again, it is not difficult to derive a similar multidimensional condition for the inequality $\mathbf{P}(\omega,\,\cdot\,)$ (see [51, Proposition 3.5] for a proof).

10. Transport-information inequalities

Instead of the transport-entropy inequality $\alpha(\mathcal{T}_c) \le H$, Guillin, Léonard, Wu and Yao have investigated in [57] the following transport-information inequality

$$(\mathcal{T}_c I) \qquad \alpha(\mathcal{T}_c(\nu, \mu)) \le I(\nu|\mu),$$

for all $\nu \in \mathrm{P}(\mathcal{X})$, where the relative entropy $H(\nu|\mu)$ is replaced by the Donsker-Varadhan information $I(\nu|\mu)$ of ν with respect to μ which was defined at (36).

This section reports some results of [57].

Background material from large deviation theory. We have seen in Section 5 that any transport-entropy inequality satisfied by a probability measure μ is connected to the large deviations of the empirical measure $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ of the sequence $(X_i)_{i\ge1}$ of independent copies of μ-distributed random variables. The link between these two notions is Sanov's theorem which asserts that L_n obey the large deviation principle with the relative entropy $\nu \mapsto H(\nu|\mu)$ as its rate function. In this section, we are going to play the same game replacing $(X_i)_{i\ge1}$ with an $\mathcal{X}$-valued time-continuous Markov process $(X_t)_{t\ge0}$ with a unique invariant probability measure μ. Instead of the large deviations of L_n, it is natural to consider the large deviations of the occupation measure

$$L_t := \frac{1}{t}\int_0^t \delta_{X_s}\,ds$$

as the length of observation t tends to infinity. The random probability measure L_t describes the ratio of time the random path $(X_s)_{0\le s\le t}$ has spent in each subset of $\mathcal{X}$. If $(X_t)_{t\ge0}$ is μ-ergodic, then the ergodic theorem states that, almost surely, L_t tends to μ as t tends to infinity. If in addition $(X_t)_{t\ge0}$ is μ-reversible, then L_t obeys the large deviation principle with some rate function $I(\,\cdot\,|\mu)$. Roughly speaking:

$$\mathbb{P}(L_t \in A) \underset{t\to\infty}{\asymp} e^{-t\inf_{\nu\in A} I(\nu|\mu)}. \tag{52}$$

The functional $\nu \in \mathrm{P}(\mathcal{X}) \mapsto I(\nu|\mu) \in [0,\infty]$ measures some kind of difference between ν and μ, i.e. some quantity of information that ν brings out with respect to the prior knowledge of

μ. With the same strategy as in Section 5, based on similar heuristics, we are lead to a new class of transport inequalities which are called transport-information inequalities.

We give now a rigorous statement of (52) which plays the same role as Sanov's theorem played in Section 5.

Let the Markov process $(X_t)_{t\geq 0}$ satisfy the assumptions which have been described at Section 8.2. Recall that the Donsker-Varadhan information $I(\cdot|\mu)$ is defined at (36).

Theorem 10.1 (Large deviations of the occupation measure)**.** *Denoting* $\mathbb{P}_\beta(\,\cdot\,) := \int_{\mathcal{X}} \mathbb{P}_x(\,\cdot\,)\, d\beta(x)$ *for any initial probability measure* β*, suppose as in Remark 8.4 that* $((X_t)_{t\geq 0}, \mathbb{P}_\mu)$ *is a stationary ergodic process.*
In addition to these assumptions on the Markov process, suppose that the initial law $\beta \in \mathrm{P}(\mathcal{X})$ *is absolutely continuous with respect to* μ *and* $d\beta/d\mu$ *is in* $L^2(\mu)$*. Then,* L_t *obeys the large deviation principle in* $\mathrm{P}(\mathcal{X})$ *with the rate function* $I(\,\cdot\,|\mu)$*, as* t *tends to infinity. This means that, for all Borel measurable* $A \subset \mathrm{P}(\mathcal{X})$*,*

$$-\inf_{\nu\in \mathrm{int}(A)} I(\nu|\mu) \leq \liminf_{t\to+\infty} \frac{1}{t}\log \mathbb{P}_\beta\left(L_t \in A\right) \leq \limsup_{t\to+\infty} \frac{1}{t}\log \mathbb{P}_\beta\left(L_t \in A\right) \leq -\inf_{\nu\in \mathrm{cl}(A)} I(\nu|\mu)$$

where $\mathrm{int}(A)$ *denotes the interior of* A *and* $\mathrm{cl}(A)$ *its closure (for the weak topology).*

This was proved by Donsker and Varadhan [39] under some conditions of absolute continuity and regularity of $P_t(x, dy)$ but without any restriction on the initial law. The present statement has been proved by Wu [111, Corollary B.11].

The inequalities $\mathbf{W_1I}$ and $\mathbf{W_2I}$. The derivation of the large deviation results for L_t as t tends to infinity is intimately related to the Feynman-Kac semigroup

$$P_t^u g(x) := \mathbb{E}^x\left[g(X_t)\exp\left(\int_0^t u(X_s)\, ds\right)\right].$$

When u is bounded, (P_t^u) is a strongly continuous semigroup of bounded operators on $L^2(\mu)$ whose generator is given by $\mathcal{L}^u g = \mathcal{L}g + ug$, for all $g \in \mathbb{D}_2(\mathcal{L}^u) = \mathbb{D}_2(\mathcal{L})$.

Theorem 10.2 (Deviation of the empirical mean, [57])**.** *Let* d *be a lower semicontinuous metric on the polish space* $\mathcal{X}$*,* (X_t) *be a* μ*-reversible and ergodic Markov process on* $\mathcal{X}$ *and* α *a function in the class* $\mathcal{A}$*, see Definition 3.1.*

(1) The following statements are equivalent:
- $\forall \nu \in \mathrm{P}(\mathcal{X})$, $I(\nu|\mu) < \infty \Rightarrow \int_{\mathcal{X}} d(x_o, \cdot\,)\, d\nu < \infty$;
- $\mathbb{E}_\mu \exp\left(\lambda_o \int_0^1 d(x_o, X_t)\, dt\right) < \infty$ *for some* $\lambda_o > 0$.

(2) Under this condition, the subsequent statements are equivalent.

(a) The following inequality holds true:

$$\alpha(W_1(\nu,\mu)) \leq I(\nu|\mu), \qquad (W_1I)$$

for all $\nu \in \mathrm{P}(\mathcal{X})$*.*

(b) For all Lipschitz function u *on* $\mathcal{X}$ *with* $\|u\|_{\mathrm{Lip}} \leq 1$ *and all* $\lambda, t \geq 0$*,*

$$\|P_t^{\lambda u}\|_{L^2(\mu)} \leq \exp\left(t[\lambda \int_{\mathcal{X}} u\, d\mu + \alpha^{\circledast}(\lambda)]\right);$$

(c) *For all Lipschitz function u on $\mathcal{X}$ with $\|u\|_{\mathrm{Lip}} \le 1$, $\int_{\mathcal{X}} u\,d\mu = 0$ and all $\lambda \ge 0$,*

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{E}_\mu \exp\left(\lambda \int_0^t u(X_s)\,ds\right) \le \alpha^{\circledast}(\lambda);$$

(d) *For all Lipschitz function u on $\mathcal{X}$, $r, t > 0$ and $\beta \in \mathrm{P}(\mathcal{X})$ such that $d\beta/d\mu \in L^2(\mu)$,*

$$\mathbb{P}_\beta\left(\frac{1}{t}\int_0^t u(X_s)\,ds \ge \int_{\mathcal{X}} u\,d\mu + r\right) \le \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\Big(-t\alpha\left(r/\|u\|_{\mathrm{Lip}}\right)\Big).$$

Remark 10.3. The Laplace-Varadhan principle allows us to identify the left-hand side of the inequality stated at (c), so that (c) is equivalent to: For all Lipschitz function u on $\mathcal{X}$ with $\|u\|_{\mathrm{Lip}} \le 1$, $\int_{\mathcal{X}} u\,d\mu = 0$, all $\lambda \ge 0$ and all $\nu \in \mathrm{P}(\mathcal{X})$,

$$\lambda \int_{\mathcal{X}} u\,d\nu - I(\nu|\mu) \le \alpha^{\circledast}(\lambda).$$

The proof of statement (1) follows the proof of (26) once one knows that $\nu \mapsto I(\nu|\mu)$ and $u \mapsto \Upsilon(u) := \log \|P_1^u\|_{L^2(\mu)} = \frac{1}{t}\log\|P_t^u\|_{L^2(\mu)}$ (for all $t > 0$) are convex conjugate to each other. The idea of the proof of the second statement is pretty much the same as in Section 5. As was already mentioned, one has to replace Sanov's theorem with Theorem 10.1. The equivalence of (a) and (c) can be obtained without appealing to large deviations, but only invoking the duality of inequalities stated at Theorem 3.5 and the fact that I and Υ are convex conjugate to each other, as was mentioned a few lines above.

Let us turn our attention to the analogue of $\mathbf{T}_2$.

Definition 10.4. *The probability measure $\mu \in \mathrm{P}_2(\mathcal{X})$ satisfies the inequality $\mathbf{W}_2\mathbf{I}(C)$ with constant C if*

$$W_2^2(\nu,\mu) \le C^2 I(\nu|\mu),\ \forall \nu \in \mathrm{P}(\mathcal{X}).$$

Theorem 10.5 ($\mathbf{W}_2\mathbf{I}$, [57])**.** *The statements below are equivalent.*

(a) *The probability measure $\mu \in \mathrm{P}(\mathcal{X})$ verifies $\mathbf{W}_2\mathbf{I}(C)$.*
(b) *For any $v \in \mathcal{B}_b(\mathcal{X})$, $\|P_t^{\frac{Qv}{C^2}}\|_{L^2(\mu)} \le e^{\frac{t}{C^2}\mu(v)}$, $\forall t \ge 0$ where $Qv(x) = \inf_{y\in\mathcal{X}}\{v(y) + d^2(x,y)\}$.*
(c) *For any $u \in \mathcal{B}_b(\mathcal{X})$, $\|P_t^{\frac{u}{C^2}}\|_{L^2(\mu)} \le e^{\frac{t}{C^2}\mu(Su)}$, $\forall t \ge 0$ where $Su(y) = \sup_{x\in\mathcal{X}}\{u(y) - d^2(x,y)\}$.*

Proposition 10.6 ($\mathbf{W}_2\mathbf{I}$ in relation with $\mathbf{LS}$ and $\mathbf{P}$, [57])**.** *In the framework of the Riemannian manifold as above, the following results hold.*

(a) $\mathbf{LS}(C)$ *implies* $\mathbf{W}_2\mathbf{I}(C)$.
(b) $\mathbf{W}_2\mathbf{I}(C)$ *implies* $\mathbf{P}(C/2)$.
(c) *Assume that* $\mathrm{Ric} + \mathrm{HessV} \ge \kappa\mathrm{Id}$ *with $\kappa \in \mathbb{R}$. If $C\kappa \le 2$, Then,*
$\mathbf{W}_2\mathbf{I}(C)$ *implies* $\mathbf{LS}(2C - C^2\kappa/2)$.

Note that $\mathbf{W}_2\mathbf{I}(C)$ with $C\kappa \le 2$ is possible. This follows from Part (a) and the Bakry-Emery criterion in the case $\kappa > 0$, see Corollary 7.3.

Proof. • Proof of (a). By Theorem 8.12, we know that $\mathbf{LS}(C)$ implies $\mathbf{T}_2(C)$. Hence, $W_2(\nu,\mu) \le \sqrt{2CH(\nu|\mu)} \le 2C\sqrt{I(\nu|\mu)}$.

• Proof of (b). The proof follows from the usual linearization procedure. Set $\mu_\varepsilon = (1+\varepsilon g)\mu$ for some smooth and compactly supported g with $\int g\,d\mu = 0$, we easily get $\lim_{\varepsilon\to 0} I(\mu_\varepsilon|\mu)/\varepsilon^2 = \frac{1}{4}\mathcal{E}(g,g)$ and by Otto-Villani [89, p.394], there exists r such that $\int g^2\,d\mu \le \sqrt{\mathcal{E}(g,g)}\frac{W_2(\mu_\varepsilon,\mu)}{\varepsilon} + \frac{r}{\varepsilon}W_2^2(\mu_\varepsilon,\mu)$. Using now $\mathbf{W}_2\mathbf{I}(C)$ we get

$$\int g^2\,d\mu \le C\sqrt{\mathcal{E}(g,g)}\sqrt{\frac{I(\mu_\varepsilon|\mu)}{\varepsilon^2}} + \frac{rC^2}{\varepsilon}I(\mu_\varepsilon|\mu).$$

Letting $\varepsilon \to 0$ gives the result.

• Proof of (c). It is a direct application of the HWI inequality, see Corollary 7.4. □

Tensorization. In Section 1 we have already seen how transport-entropy inequalities tensorize. We revisit tensorization, but this time we replace the relative entropy $H(\cdot|\mu)$ with the Donsker-Varadhan $I(\cdot|\mu)$. This will be quite similar in spirit to what as already been done in Section 1, but we are going to use alternate technical lemmas which will prepare the road to Section 11 where a Gibbs measure will replace our product measure. This approach which is partly based on Gozlan & Léonard [53] is developed in Guillin, Léonard, Wu & Yao's article [57].

On the polish product space $\mathcal{X}^{(n)} := \prod_{i=1}^n \mathcal{X}_i$ equipped with the product measure $\mu := \otimes_{i=1}^n \mu_i$, consider the cost function

$$\oplus_i c_i(x,y) := \sum_{i=1}^n c_i(x_i,y_i),\ \forall x,y \in \mathcal{X}^{(n)}$$

where for each index i, c_i is lower semicontinuous on $\mathcal{X}_i^2$ and assume that for each $1 \le i \le n$, $\mu_i \in \mathrm{P}(\mathcal{X}_i)$ satisfies the transport-information inequality

$$\alpha_i(\mathcal{T}_{c_i}(\nu,\mu_i)) \le I_{\mathcal{E}_i}(\nu|\mu_i),\ \forall \nu \in \mathrm{P}(\mathcal{X}_i) \tag{53}$$

where $I_{\mathcal{E}_i}(\nu|\mu_i)$ is the Donsker-Varadhan information related to some Dirichlet form $(\mathcal{E}_i, \mathbb{D}(\mathcal{E}_i))$, and α_i stands in the class $\mathcal{A}$, see Definition 3.1. Define the global Dirichlet form $\oplus_i^\mu \mathcal{E}_i$ by

$$\mathbb{D}(\oplus_i^\mu \mathcal{E}_i) := \left\{ g \in L^2(\mu) : g_i^{\widetilde{x}_i} \in \mathbb{D}(\mathcal{E}_i), \text{for } \mu\text{-a.e. } \widetilde{x}_i \text{ and } \int_{\mathcal{X}^{(n)}} \sum_{i=1}^n \mathcal{E}_i(g_i^{\widetilde{x}_i}, g_i^{\widetilde{x}_i})\,d\mu(x) < +\infty \right\}$$

where $g_i^{\widetilde{x}_i} : x_i \mapsto g_i^{\widetilde{x}_i}(x_i) := g(x)$ with $\widetilde{x}_i := (x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_n)$ considered as fixed and

$$\oplus_i^\mu \mathcal{E}_i(g,g) := \int_{\mathcal{X}^{(n)}} \sum_{i=1}^n \mathcal{E}_i(g_i^{\widetilde{x}_i}, g_i^{\widetilde{x}_i})\,d\mu(x), \quad g \in \mathbb{D}(\oplus_i^\mu \mathcal{E}_i). \tag{54}$$

Let $I_{\oplus_i \mathcal{E}_i}(\nu|\mu)$ be the Donsker-Varadhan information associated with $(\oplus_i^\mu \mathcal{E}_i, \mathbb{D}(\oplus_i^\mu \mathcal{E}_i))$, see (36). We denote $\alpha_1 \square \cdots \square \alpha_n$ the inf-convolution of $\alpha_1,\dots,\alpha_n$ which is defined by

$$\alpha_1 \square \cdots \square \alpha_n(r) = \inf\{\alpha_1(r_1) + \cdots + \alpha_n(r_n); r_1,\dots,r_n \ge 0, r_1 + \cdots + r_n = r\}, \quad r \ge 0.$$

Theorem 10.7 ([57]). *Assume that for each $i = 1, \cdots, n$, μ_i satisfies the transport-entropy inequality (53). Then, the product measure μ satisfies the following transport-entropy inequality*

$$\alpha_1 \Box \cdots \Box \alpha_n (\mathcal{T}_{\oplus c_i}(\nu, \mu)) \leq I_{\oplus_i \mathcal{E}_i}(\nu|\mu), \quad \forall \nu \in \mathrm{P}(\mathcal{X}^{(n)}). \tag{55}$$

This result is similar to Proposition 1.8. But its proof will be different. It is based on the following sub-additivity result for the transport cost of a product measure.

Let $(X_i)_{1\leq i\leq n}$ be the canonical process on $\mathcal{X}^{(n)}$. For each i, $\widetilde{X}_i = (X_j)_{1\leq j\leq n; j\neq i}$ is the configuration without its value at index i.

Given a probability measure ν on $\mathcal{X}^{(n)}$,

$$\nu_i^{\widetilde{x}_i} = \nu(X_i \in \cdot | \widetilde{X}_i = \widetilde{x}_i)$$

denotes the regular conditional distribution of X_i knowing that $\widetilde{X}_i = \widetilde{x}_i$ under ν and

$$\nu_i = \nu(X_i \in \cdot)$$

denotes the i-th marginal of ν.

Proposition 10.8 ([57]). *Let $\mu = \bigotimes_{i=1}^n \mu_i$ be a product probability measure on $\mathcal{X}^{(n)}$. For all $\nu \in \mathrm{P}(\mathcal{X}^{(n)})$,*

$$\mathcal{T}_{\oplus c_i}(\nu, \mu) \leq \int_{\mathcal{X}^{(n)}} \left(\sum_{i=1}^n \mathcal{T}_{c_i}(\nu_i^{\widetilde{x}_i}, \mu_i) \right) d\nu(x).$$

The proof of this proposition is given in the Appendix at Proposition A.2.

The following additivity property of the Fisher information will be needed. It holds true even in the dependent case.

Lemma 10.9 ([57]). *Let ν, μ be probability measures on $\mathcal{X}^{(n)}$ such that $I_{\oplus_i \mathcal{E}_i}(\nu|\mu) < +\infty$. Then,*

$$I_{\oplus_i \mathcal{E}_i}(\nu|\mu) = \int_{\mathcal{X}^{(n)}} \sum_{i=1}^n I_{\mathcal{E}_i}(\nu_i^{\widetilde{x}_i} | \mu_i^{\widetilde{x}_i}) \, d\nu(x).$$

Sketch of proof. Let f be a regular enough function. This why this is only a sketch of proof, because an approximation argument which we do not present here, is needed to obtain the result for any f in the domain $\mathbb{D}(\oplus_i \mathcal{E}_i)$.

Then, $\dfrac{d\nu_i^{\widetilde{x}_i}}{d\mu_i^{\widetilde{x}_i}}(x_i) = \dfrac{f_i^{\widetilde{x}_i}(x_i)}{\mu_i^{\widetilde{x}_i}(f_i^{\widetilde{x}_i})}$, ν-a.s. where $f_i^{\widetilde{x}_i}$ is the function f of x_i with $\widetilde{x}_i$ fixed. For ν-a.e. $\widetilde{x}_i$,

$$I_{\mathcal{E}_i}(\nu_i^{\widetilde{x}_i} | \mu_i^{\widetilde{x}_i}) = \mathcal{E}_i \left(\sqrt{\frac{f_i^{\widetilde{x}_i}}{\mu_i^{\widetilde{x}_i}(f_i^{\widetilde{x}_i})}}, \sqrt{\frac{f_i^{\widetilde{x}_i}}{\mu_i^{\widetilde{x}_i}(f_i^{\widetilde{x}_i})}} \right) = \frac{1}{\mu_i^{\widetilde{x}_i}(f_i^{\widetilde{x}_i})} \mathcal{E}_i(\sqrt{f_i^{\widetilde{x}_i}}, \sqrt{f_i^{\widetilde{x}_i}}).$$

Thus,

$$\begin{aligned}\int_{\mathcal{X}^{(n)}} \sum_{i=1}^n I_{\mathcal{E}_i}(\nu_i^{\widetilde{x}_i}|\mu_i^{\widetilde{x}_i})\,d\nu(x) &= \int_{\mathcal{X}^{(n)}} f(x)\sum_{i=1}^n \frac{1}{\mu_i^{\widetilde{x}_i}(f_i^{\widetilde{x}_i})}\mathcal{E}_i\Big(\sqrt{f_i^{\widetilde{x}_i}},\sqrt{f_i^{\widetilde{x}_i}}\Big)\,d\mu(x)\\ &= \int_{\mathcal{X}^{(n)}} \sum_{i=1}^n \mathcal{E}_i\Big(\sqrt{f_i^{\widetilde{x}_i}},\sqrt{f_i^{\widetilde{x}_i}}\Big)\,d\mu(x)\\ &= \oplus_i^\mu \mathcal{E}_i(\sqrt{f},\sqrt{f})\\ &= I_{\oplus_i\mathcal{E}_i}(\nu|\mu),\end{aligned}$$

which completes the sketch of the proof. □

This additivity is different from the super-additivity of the Fisher information for a product measure obtained by Carlen [24].

We are now ready to write the proof of Theorem 10.7.

Proof of Theorem 10.7. Without loss of generality we may assume that $I(\nu|\mu) < +\infty$. By Proposition 10.8, Jensen inequality and the definition of $\alpha_1\square\cdots\square\alpha_n$,

$$\begin{aligned}\alpha_1\square\cdots\square\alpha_n(\mathcal{T}_{\oplus c_i}(\nu,\mu)) &\le \alpha_1\square\cdots\square\alpha_n\left(\int_{\mathcal{X}^{(n)}}\sum_{i=1}^n \mathcal{T}_{c_i}(\nu_i^{\widetilde{x}_i},\mu_i)\right)d\nu(x)\\ &\le \int_{\mathcal{X}^{(n)}} \alpha_1\square\cdots\square\alpha_n\left(\sum_{i=1}^n \mathcal{T}_{c_i}(\nu_i^{\widetilde{x}_i},\mu_i)\right)d\nu(x)\\ &\le \int_{\mathcal{X}^{(n)}}\sum_{i=1}^n \alpha_i(\mathcal{T}_{c_i}(\nu_i^{\widetilde{x}_i},\mu_i))\,d\nu(x)\\ &\le \int_{\mathcal{X}^{(n)}}\sum_{i=1}^n I_{\mathcal{E}_i}(\nu_i^{\widetilde{x}_i}|\mu_i)\,d\nu(x).\end{aligned}$$

The last quantity is equal to $I_{\oplus\mathcal{E}_i}(\nu|\mu)$, by Lemma 10.9. □

As an example of application, let $(X_t^i)_{t\ge0}, i=1,\cdots,n$ be n Markov processes with the same transition semigroup (P_t) and the same symmetrized Dirichlet form $\mathcal{E}$ on $L^2(\rho)$, and conditionally independent once the initial configuration $(X_0^i)_{i=1,\cdots,n}$ is fixed. Then $X_t := (X_t^1,\cdots,X_t^n)$ is a Markov process with the symmetrized Dirichlet form given by

$$\oplus_n^{\rho^n}\mathcal{E}(g,g) = \int\sum_{i=1}^n \mathcal{E}(g_i^{\widetilde{x}_i},g_i^{\widetilde{x}_i})\,\rho(dx_1)\cdots\rho(dx_n).$$

Corollary 10.10 ([57])**.**

(1) Assume that ρ satisfies the transport-information inequality $\alpha(\mathcal{T}_c)\le I_{\mathcal{E}}$ on $\mathcal{X}$ with α in the class $\mathcal{A}$. Then ρ^n satisfies

$$n\alpha\left(\frac{\mathcal{T}_{\oplus_n c}(\nu,\rho^n)}{n}\right)\le I_{\oplus_n\mathcal{E}}(\nu|\rho^n),\quad \forall\nu\in\mathrm{P}(\mathcal{X}^n).$$

(2) *Suppose in particular that ρ verifies $\alpha(\mathcal{T}_d) \le I_{\mathcal{E}}$ for the metric lower semicontinuous cost d. Then, for any Borel measurable d-Lipschitz(1) function u, any initial measure β on $\mathcal{X}^n$ with $d\beta/d\rho^n \in L^2(\rho^n)$ and any $t, r > 0$,*

$$\mathbb{P}_\beta\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{t}\int_0^t u(X_s^i)\,ds \ge \rho(u) + r\right) \le \left\|\frac{d\beta}{d\rho^n}\right\|_2 e^{-nt\alpha(r)}.$$

(3) *If ρ satisfies $\mathbf{W}_2\mathbf{I}(C) : \mathcal{T}_{d^2} \le C^2 I_{\mathcal{E}}$, then ρ^n satisfies $\mathbf{W}_2\mathbf{I}(C) : \mathcal{T}_{\oplus_n d^2} \le C^2 I_{\oplus_n \mathcal{E}}$.*

Proof. As $\alpha^{\square n}(r) = n\alpha(r/n)$, the first part (1) follows from Theorem 10.7. The second part (2) follows from Theorem 10.2 and the third part (3) is a direct application of (1). □

Transport-information inequalities in the literature. Several integral criteria are worked out in [57], mostly in terms of Lyapunov functions. Note also that the assumptions in [57] are a little less restrictive than those of the present section, in particular the Markov process might not be reversible, but it is required that its Dirichlet form is closable.

For further relations between $\alpha(\mathcal{T}) \le I$ and other functional inequalities, one can read the paper [56] by Guillin, Léonard, Wang and Wu.

In [45], Gao, Guillin and Wu have refined the above concentration results in such a way that Bernstein inequalities are accessible. The strategy remains the same since it is based on the transport-information inequalities of Theorem 10.2, but the challenge is to express the constants in terms of asymptotic variances. Lyapunov function conditions allow to derive explicit rates.

An interesting feature with Theorem 10.2 is that it allows to treat time-continuous Markov processes with jumps. This is widely done in [45]. But processes with jumps might not verify a Poincaré inequality even in presence of good concentration properties, for instance when considering processes with strong pulling-back drifts. In such cases, even the $\alpha(\mathcal{T}) \le I$ strategy fails. An alternative attack of the problem of finding concentration estimates for the empirical means (of Lipschitz observables) has been performed by Wu in [113] where usual transport inequalities $\alpha(\mathcal{T}) \le H$ at the level of the Markov transition kernel are successfully exploited.

Gibbs measures are also investigated by Gao and Wu [46] by means of transport-information inequalities. This is developed in the next Section 11.

11. Transport inequalities for Gibbs measures

We have seen transport inequalities with respect to a reference measure μ and how to derive transport inequalities for the product measure $\mu = \rho^n$ from transport inequalities for ρ. A step away from this product measure structure, one is naturally lead to consider Markov structures. This is the case with Gibbs measures, a description of equilibrium states in statistical physics. Three natural problem encountered with Gibbs measures are:

(1) Find criteria for the uniqueness/non-uniqueness of the solutions to the Dobrushin-Lanford-Ruelle (DLR) problem associated with the local specifications (see next subsection below). This uniqueness corresponds to the absence of phase coexistence of the physical system and the unique solution is our Gibbs measure μ.

(2) Obtain concentration estimates for the Gibbs measures.
(3) In case of uniqueness, estimate the speed of convergence of the Glauber dynamics (see below) towards the equilibrium μ.

A powerful tool for investigating this program is the logarithmic Sobolev inequality. This is known since the remarkable contribution in 1992 of Zegarlinski [114], see also the papers [97, 98] by Stroock & Zegarlinski. Lecture notes on the subject have been written by Martinelli [75], Royer [93] and Guionnet & Zegarlinski [59]. An alternate approach is to exchange logarithmic Sobolev inequalities with Poincaré inequality. Indeed, in some situations both these inequalities are equivalent [97, 98].

Recently, another approach of this problem has been proposed which consists of replacing logarithmic Sobolev inequalities by transport inequalities. This is what this section is about. The main recent contributions in this area are due to Marton [80], Wu [112], Gao & Wu [46] and Ma, Shen, Wang & Wu [74].

Gibbs measures. The configuration space is $\mathcal{X}^{\mathrm{I}}$ where $\mathcal{X}$ is the spin space and I is a countable set of sites, for instance a finite set with a graph structure or the lattice $\mathrm{I} = \mathbb{Z}^d$. A configuration is $x = (x_i)_{i\in\mathrm{I}}$ where $x_i \in \mathcal{X}$ is the spin value at site $i \in \mathrm{I}$. The spin space might be finite, for instance $\mathcal{X} = \{-1, 1\}$ as in the Ising model, or infinite, for instance $\mathcal{X} = S^k$ the k-dimensional sphere or $\mathcal{X} = \mathbb{R}^k$. It is assumed that $\mathcal{X}$ is a polish space furnished with its Borel σ-field. Consequently, any conditional probability measure admits a regular version.

Let us introduce some notation. For any $i \in \mathrm{I}$, $\widetilde{x}_i$ is the restriction of the configuration x to $\{i\}^c := \mathrm{I} \setminus \{i\}$. Given $\nu \in \mathrm{P}(\mathcal{X}^{\mathrm{I}})$, one can consider the family of conditional probability laws of X_i knowing $\widetilde{X}_i$ where $X = (X_i)_{i\in\mathrm{I}}$ is the canonical configuration. We denote these conditional laws:

$$\nu_i^{\widetilde{x}_i} := \mu(X_i \in \cdot | \widetilde{X}_i = \widetilde{x}_i), \quad i \in \mathrm{I}, x \in \mathcal{X}^{\mathrm{I}}.$$

As different projections of the same ν, these conditional laws satisfy a collection of compatibility conditions.

The DLR problem is the following inverse problem. Consider a family of prescribed local specifications $\mu_i^{\widetilde{x}_i}$, $i \in \mathrm{I}$, $x \in \mathcal{X}^{\mathrm{I}}$ which satisfy the appropriate collection of compatibility conditions. Does there exist some $\mu \in \mathrm{P}(\mathcal{X}^{\mathrm{I}})$ whose conditional distributions are precisely these prescribed local specifications? Is there a unique such μ?
The solutions of the DLR problem are called Gibbs measures.

Glauber dynamics. It is well-known that $d\mu(x) = Z^{-1}e^{-V(x)}\,dx$ where Z is a normalizing constant, is the invariant probability measure of the Markov generator $\Delta - \nabla V \cdot \nabla$. This fact is extensively exploited in the semigroup approach of the Poincaré and logarithmic Sobolev inequalities. Indeed, these inequalities exhibit on their right-hand side the Dirichlet form $\mathcal{E}$ associated with this Markov generator.

This differs from the **WH** inequalities such as $\mathbf{T}_1$ or $\mathbf{T}_2$ which do not give any role to any Dirichlet form: it is the main reason why we didn't encounter the semigroup approach in these notes up to now. But replacing the entropy H by the information $I(\cdot|\mu)$, one obtains transport-information inequalities **WI** and the semigroups might have something to tell us.

Why should one introduce some dynamics related to a Gibbs measure? Partly because in practice the normalizing constant Z is inaccessible to computation in very high dimension, so that simulating a Markov process $(X_t)_{t\geq 0}$ admitting our Gibbs measure as its (unique) invariant measure during a long period of time allows us to compute estimates for average quantities. Another reason is precisely the semigroup approach which helps us deriving functional inequalities dealing with Dirichlet forms. This relevant dynamics, which is often called the Glauber dynamics, is precisely the Markov dynamics associated with the closure of the Dirichlet form which admits our Gibbs measure as its invariant measure.

Now, let us describe the Glauber dynamics precisely.

Let μ be a Gibbs measure (solution of the DLR problem) with the local specifications $\{\mu_i^{\widetilde{x}_i} \in \mathrm{P}(\mathcal{X});\ i \in \mathrm{I}, x \in \mathcal{X}^{\mathrm{I}}\}$. For each $i \in \mathrm{I}, x \in \mathcal{X}^{\mathrm{I}}$, consider a Dirichlet form $(\mathcal{E}_i^{\widetilde{x}_i}, \mathbb{D}(\mathcal{E}_i^{\widetilde{x}_i}))$ and define the global Dirichlet form $\mathcal{E}^\mu$ by

$$\mathbb{D}(\mathcal{E}^\mu) := \Big\{ f \in L^2(\mu) : \text{for all } i \in \mathrm{I},\ f_i^{\widetilde{x}_i} \in \mathbb{D}(\mathcal{E}_i^{\widetilde{x}_i}), \text{for } \mu\text{-a.e. } \widetilde{x}_i \\ \text{and } \int_{\mathcal{X}^{\mathrm{I}}} \sum_{i\in\mathrm{I}} \mathcal{E}_i^{\widetilde{x}_i}(f_i^{\widetilde{x}_i}, f_i^{\widetilde{x}_i})\, d\mu(x) < +\infty \Big\}$$

where $f_i^{\widetilde{x}_i} : x_i \mapsto f_i^{\widetilde{x}_i}(x_i) := f(x)$ with $\widetilde{x}_i$ considered as fixed and

$$\mathcal{E}^\mu(f,f) := \int_{\mathcal{X}^{\mathrm{I}}} \sum_{i\in\mathrm{I}} \mathcal{E}_i^{\widetilde{x}_i}(f_i^{\widetilde{x}_i}, f_i^{\widetilde{x}_i})\, d\mu(x), \quad f \in \mathbb{D}(\mathcal{E}^\mu). \tag{56}$$

Assume that $\mathcal{E}^\mu$ is closable. Then, the Glauber dynamics is the Markov process associated with the closure of $\mathcal{E}^\mu$.

Example 11.1. An interesting example is given by the following extension of the standard Example 8.6. Let $\mathcal{X}$ be a complete connected Riemannian manifold. Consider a Gibbs measure μ solution to the DLR problem as above. For each $i \in \mathrm{I}$ and $x \in \mathcal{X}^{\mathrm{I}}$, the one-site Dirichlet form $\mathcal{E}_i^{\widetilde{x}_i}$ is defined for any smooth enough function f on $\mathcal{X}$ by

$$\mathcal{E}_i^{\widetilde{x}_i}(f,f) = \int_{\mathcal{X}} |\nabla f|^2\, d\mu_i^{\widetilde{x}_i}$$

and the global Dirichlet form $\mathcal{E}^\mu$ which is defined by (56) is given for any smooth enough cylindrical function f on $\mathcal{X}^{\mathrm{I}}$ by

$$\mathcal{E}^\mu(f,f) = \int_{\mathcal{X}^{\mathrm{I}}} |\nabla_{\mathrm{I}} f|^2\, d\mu$$

where ∇_{I} is the gradient on the product manifold $\mathcal{X}^{\mathrm{I}}$. The corresponding Markov process is a family indexed by I of interacting diffusion processes, all of them sharing the same fixed temperature (diffusion coefficient=2). This process on $\mathcal{X}^{\mathrm{I}}$ admits the Gibbs measure μ as an invariant measure.

Dimension-free tensorization property. It is well known that the Poincaré inequality **P** implies an exponentially fast L^2-convergence as t tends to infinity of the law of X_t to the invariant measure μ. Similarly, a logarithmic Sobolev inequality **LS** implies a stronger convergence in entropy. Moreover, both **P** and **LS** enjoy a dimension-free tensorization property which is of fundamental importance when working in an infinite dimensional setting.

This dimension-free tensorization property is also shared by $\mathbf{T}_2 = \mathbf{W}_2\mathbf{H}$, see Corollary 4.4, and by $\mathbf{W}_2\mathbf{I}$, see Corollary 10.10-(3).

Now, suppose that each one-site specification $\mu_i^{\widetilde{x}_i}$, for any $i \in \mathrm{I}$ and any $x \in \mathcal{X}^{\mathrm{I}}$, satisfies a functional inequality with the dimension-free tensorization property. One can reasonably expect that, provided that the constants $C_i^{\widetilde{x}_i}$ in these inequalities enjoy some uniformity property in i and x, any Gibbs measure built with the local specifications $\mu_i^{\widetilde{x}_i}$ also shares some non-trivial functional inequality (in the same family of inequalities). This is what Zegarlinski [114] discovered with bounded spin systems and **LS**. On the other hand, this inequality (say **P** or **LS**) satisfied by the Gibbs measures μ entails an exponentially fast convergence as t tends to infinity of the global Glauber dynamics to μ. By standard arguments, one can prove that this implies the uniqueness of the invariant measure and therefore, the uniqueness of the solution of the DLR problem.

In conclusion, some uniformity property in i and x of the inequality constants $C_i^{\widetilde{x}_i}$ is a sufficient condition for the uniqueness of the DLR problem and an exponentially fast convergence of the Glauber dynamics.

Recently, Marton [80] and Wu [112] considered the "dimension-free" transport-entropy inequality $\mathbf{T}_2$ and Gao & Wu [46] the "dimension-free" transport-information inequality $\mathbf{W}_2\mathbf{I}$ in the setting of Gibbs measures.

Dobrushin coefficients. Let d be a lower semicontinuous metric on $\mathcal{X}$ and let $\mathrm{P}_p(\mathcal{X})$ be the set of all Borel probability measures ρ on $\mathcal{X}$ such that $\int_{\mathcal{X}} d^p(\xi_o, \xi)\, d\rho(\xi) < \infty$ with $p \geq 1$. Assume that for each site $i \in \mathrm{I}$ and each boundary condition $\widetilde{x}_i$, the specification $\mu_i^{\widetilde{x}_i}$ is in $\mathrm{P}_p(\mathcal{X})$. For any $i, j \in \mathrm{I}$, the Dobrushin interaction W_p-coefficient is defined by

$$c_p(i,j) := \sup_{x,y;\ x=y \text{ off } j} \frac{W_p(\mu_i^{\widetilde{x}_i}, \mu_i^{\widetilde{y}_i})}{d(x_j, y_j)}$$

where W_p is the Wasserstein metric of order p on $\mathrm{P}_p(\mathcal{X})$ which is built on the metric d. Let $\mathbf{c}_p = (c_p(i,j))_{i,j \in \mathrm{I}}$ denote the corresponding matrix which is seen as an endomorphism of $\ell^p(\mathrm{I})$. Its operator norm is denoted by $\|\mathbf{c}_p\|_p$.

Dobrushin [37, 38] obtained a criterion for the uniqueness of the Gibbs measure (cf. Question (1) above) in terms of the coefficients $c_1(i,j)$ with $p = 1$. It is

$$\sup_{j \in \mathrm{I}} \sum_{i \in \mathrm{I}} c_1(i,j) < 1.$$

This quantity is $\|\mathbf{c}_1\|_1$, so that Dobrushin's condition expresses that $\mathbf{c}_1$ is contractive on $\ell^1(\mathrm{I})$ and the uniqueness follows from a fixed point theorem, see Föllmer's lecture notes [43] for this well-advised proof.

Wasserstein metrics on $\mathrm{P}(\mathcal{X}^{\mathrm{I}})$. Let $p \geq 1$ be fixed. The metric on $\mathcal{X}^{\mathrm{I}}$ is

$$d_{p,\mathrm{I}}(x,y) := \Big(\sum_{i \in \mathrm{I}} d^p(x_i, y_i) \Big)^{1/p}, \quad x, y \in \mathcal{X}^{\mathrm{I}} \tag{57}$$

and the Wasserstein metric $W_{p,\mathrm{I}}$ on $\mathrm{P}(\mathcal{X}^{\mathrm{I}})$ is built upon $d_{p,\mathrm{I}}$. One sees that it corresponds to the tensor cost $d_{p,\mathrm{I}}^p = \oplus_{i \in \mathrm{I}} d_i^p$ with an obvious notation.

Gao and Wu [46] have proved the following tensorization result for the Wasserstein distance between Gibbs measures.

Proposition 11.2. *Assume that* $\mu_i^{\widetilde{x}_i} \in \mathrm{P}_p(\mathcal{X})$ *for all* $i \in \mathrm{I}$ *and* $x \in \mathcal{X}^\mathrm{I}$ *and also suppose that* $\|\mathbf{c}_p\|_p < 1$. *Then,* $\mu \in \mathrm{P}_p(\mathcal{X}^\mathrm{I})$ *and for all* $\nu \in \mathrm{P}_p(\mathcal{X}^\mathrm{I})$,

$$W_{p,\mathrm{I}}^p(\nu,\mu) \le (1-\|\mathbf{c}_p\|_p)^{-1} \int_{\mathcal{X}^\mathrm{I}} \sum_{i\in\mathrm{I}} W_p^p(\nu_i^{\widetilde{x}_i}, \mu_i^{\widetilde{x}_i})\, d\nu(x).$$

Sketch of proof. As a first step, let us follow exactly the beginning of the proof of Proposition A.2 in the Appendix. Keeping the notation of Proposition A.2, we have

$$\mathbb{E}\sum_{i\in\mathrm{I}} d^p(U_i,V_i) = W_{p,\mathrm{I}}^p(\nu,\mu). \tag{58}$$

and we arrive at (66) which, with $c_i = d^p$, is

$$\mathbb{E}d^p(U_i,V_i) \le \mathbb{E}d^p(\widehat{U}_i,\widehat{V}_i) = \mathbb{E}W_p^p\left(\nu_i^{\widetilde{U}_i}, \mu_i^{\widetilde{V}_i}\right).$$

As in Marton's paper [80], we can use the triangular inequality for W_p and the definition of $\mathbf{c}_p$ to obtain for all $i \in \mathrm{I}$,

$$W_p\left(\nu_i^{\widetilde{U}_i}, \mu_i^{\widetilde{V}_i}\right) \le W_p\left(\nu_i^{\widetilde{U}_i}, \mu_i^{\widetilde{U}_i}\right) + \sum_{j\in\mathrm{I},\ j\neq i} c_p(i,j) d(U_j,V_j).$$

Putting both last inequalities together, we see that

$$\mathbb{E}d^p(U_i,V_i) \le \mathbb{E}\left(W_p\left(\nu_i^{\widetilde{U}_i}, \mu_i^{\widetilde{U}_i}\right) + \sum_{j\in\mathrm{I},\ j\neq i} c_p(i,j) d(U_j,V_j)\right)^p, \quad \text{for all } i\in\mathrm{I}, \tag{59}$$

and summing them over all the sites i gives us

$$\mathbb{E}\sum_{i\in\mathrm{I}} d^p(U_i,V_i) \le \mathbb{E}\sum_{i\in\mathrm{I}}\left(W_p\left(\nu_i^{\widetilde{U}_i}, \mu_i^{\widetilde{U}_i}\right) + \sum_{j\in\mathrm{I},\ j\neq i} c_p(i,j) d(U_j,V_j)\right)^p.$$

Consider the norm $\|A\| := (\mathbb{E}\sum_{i\in\mathrm{I}} |A_i|^p)^{1/p}$ of the random vector $A = (A_i)_{i\in\mathrm{I}}$. With $A_i = d(U_i,V_i)$ and $B_i = W_p\left(\nu_i^{\widetilde{U}_i}, \mu_i^{\widetilde{U}_i}\right)$, this inequality is simply

$$\|A\| \le \|\mathbf{c}_p A + B\|,$$

since $c_p(i,i) = 0$ for all $i \in \mathrm{I}$. This implies that

$$(1-\|\mathbf{c}_p\|_p)\|A\| \le \|B\|$$

which, with (58), is the announced result.

Similarly to the first step of the proof of Proposition A.2, this proof contains a measurability bug and one has to correct it exactly as in the complete proof of Proposition A.2. □

Recall that the global Dirichlet form $\mathcal{E}^\mu$ is defined at (56). The corresponding Donsker-Varadhan information is defined by

$$I_{\mathcal{E}^\mu}(\nu|\mu) = \begin{cases} \mathcal{E}_\mu(\sqrt{f},\sqrt{f}) & \text{if } \nu = f\mu \in \mathrm{P}(\mathcal{X}^\mathrm{I}),\ f\in\mathbb{D}(\mathcal{E}_\mu) \\ +\infty & \text{otherwise.} \end{cases}$$

Similarly, we define for each $i \in \mathrm{I}$ and $x \in \mathcal{X}^{\mathrm{I}}$,

$$I_{\mathcal{E}_i^{\widetilde{x}_i}}(\rho|\mu_i^{\widetilde{x}_i}) = \begin{cases} \mathcal{E}_i^{\widetilde{x}_i}(\sqrt{g},\sqrt{g}) & \text{if } \rho = g\mu_i^{\widetilde{x}_i} \in \mathrm{P}(\mathcal{X}), g \in \mathbb{D}(\mathcal{E}_i^{\widetilde{x}_i}) \\ +\infty & \text{otherwise.} \end{cases}$$

We are now ready to present a result of tensorization of one-site $\mathbf{W}_2\mathbf{I}$ inequalities in the setting of Gibbs measures.

Theorem 11.3 ([46])**.** *Assume that for each site $i \in \mathrm{I}$ and any configuration $x \in \mathcal{X}^{\mathrm{I}}$ the local specifications are in $\mathrm{P}_2(\mathcal{X})$ and satisfy the following one-site $\mathbf{W}_2\mathbf{I}$ inequality*

$$W_2^2(\rho, \mu_i^{\widetilde{x}_i}) \le C^2 I_{\mathcal{E}_i^{\widetilde{x}_i}}(\rho|\mu_i^{\widetilde{x}_i}), \quad \forall \rho \in \mathrm{P}_2(\mathcal{X}),$$

the constant C being uniform in i and x.
It is also assumed that the Dobrushin W_2-coefficients satisfy $\|\mathbf{c}_2\|_2 < 1$.
Then, any Gibbs measure μ is in $\mathrm{P}_2(\mathcal{X}^{\mathrm{I}})$ and satisfies the following $\mathbf{W}_2\mathbf{I}$ inequality:

$$W_{2,\mathrm{I}}^2(\nu,\mu) \le \frac{C^2}{1-\|\mathbf{c}_2\|_2} I_{\mathcal{E}^\mu}(\nu|\mu), \quad \forall \nu \in \mathrm{P}_2(\mathcal{X}^{\mathrm{I}}).$$

Proof. By Proposition 11.2, we have for all $\nu \in \mathrm{P}_2(\mathcal{X}^{\mathrm{I}})$

$$W_{2,\mathrm{I}}^2(\nu,\mu) \le (1-\|\mathbf{c}_2\|_2)^{-1} \int_{\mathcal{X}^{\mathrm{I}}} \sum_{i\in\mathrm{I}} W_2^2(\nu_i^{\widetilde{x}_i}, \mu_i^{\widetilde{x}_i})\, d\nu(x).$$

Since the local specifications satisfy a uniform inequality $\mathbf{W}_2\mathbf{I}$, we obtain

$$\begin{aligned} W_{2,\mathrm{I}}^2(\nu,\mu) &\le \frac{C^2}{1-\|\mathbf{c}_2\|_2} \int_{\mathcal{X}^{\mathrm{I}}} \sum_{i\in\mathrm{I}} I_{\mathcal{E}_i^{\widetilde{x}_i}}(\nu_i^{\widetilde{x}_i}|\mu_i^{\widetilde{x}_i})\, d\nu(x) \\ &= \frac{C^2}{1-\|\mathbf{c}_2\|_2} I_{\mathcal{E}^\mu}(\nu|\mu) \end{aligned}$$

where the last equality is Lemma 10.9. □

We decided to restrict our attention to the case $p = 2$ because of its free-dimension property, but a similar result still holds with $p > 1$ under the additional requirement that I is a finite set.

As a direct consequence, under the assumptions of Theorem 11.3, μ satisfies a fortiori the $\mathbf{W}_1\mathbf{I}$ equality

$$W_{1,\mathrm{I}}^2(\nu,\mu) \le \frac{C^2}{1-\|\mathbf{c}_2\|_2} I_{\mathcal{E}^\mu}(\nu|\mu), \quad \forall \nu \in \mathrm{P}_1(\mathcal{X}^{\mathrm{I}}).$$

Therefore, we can derive from Theorem 10.2 the following deviation estimate for the Glauber dynamics.

Corollary 11.4 (Deviation of the Glauber dynamics)**.** *Suppose that μ is the unique Gibbs measure (for instance if $\|\mathbf{c}_1\|_1 < 1$) and that the assumptions of Theorem 11.3 are satisfied. Then, the Glauber dynamics $(X_t)_{t\ge 0}$ verifies the following deviation inequality.*

For all $d_{1,\mathrm{I}}$-Lipschitz function u on $\mathcal{X}^{\mathrm{I}}$ (see (57) for the definition of $d_{1,\mathrm{I}}$), for all $r,t>0$ and all $\beta \in \mathrm{P}(\mathcal{X}^{\mathrm{I}})$ such that $d\beta/d\mu \in L^2(\mu)$,

$$\mathbb{P}_\beta\left(\frac{1}{t}\int_0^t u(X_s)\,ds \geq \int_{\mathcal{X}^{\mathrm{I}}} u\,d\mu + r\right) \leq \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\Big(-\frac{1-\|\mathbf{c}_2\|_2}{C^2\|u\|_{\mathrm{Lip}}^2}\,tr^2\Big).$$

12. Free transport inequalities

The semicircular law is the probability distribution σ on $\mathbb{R}$ defined by

$$d\sigma(x) = \frac{1}{2\pi}\sqrt{4-x^2}\,\mathbb{1}_{[-2,2]}(x)\,dx.$$

This distribution plays a fundamental role in the asymptotic theory of Wigner random matrices.

Definition 12.1 (Wigner matrices). *Let N be a positive integer; a (complex) $N\times N$ Wigner matrix M is an Hermitian random matrix such that the entries $M(i,j)$ with $i<j$ are i.i.d $\mathbb{C}$-valued random variables with $\mathbb{E}[M(i,j)]=0$ and $\mathbb{E}[|M(i,j)|^2]=1$ and such that the diagonal entries $M(i,i)$ are i.i.d centered real random variables independent of the off-diagonal entries and having finite variance. When the entries of M are Gaussian random variables and $\mathbb{E}[M(1,1)^2]=1$, M is referred to as the Gaussian Unitary Ensemble (GUE).*

Let us recall the famous Wigner theorem (see e.g [1] or [58] for a proof).

Theorem 12.2 (Wigner theorem). *Let $(M_N)_{N\geq 0}$ be a sequence of complex Wigner matrices such that $\max_{N\geq 0}(\mathbb{E}[M_N(1,1)^2]) < +\infty$ and let L_N be the empirical distribution of $X_N := \frac{1}{\sqrt{N}}M_N$, that is to say*

$$L_N = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i^N},$$

where $\lambda_1^N \leq \lambda_2^N \leq \ldots \leq \lambda_N^N$ are the (real) eigenvalues of X_N. Then the sequence of random probability measures L_N converges almost surely to the semicircular law (for the weak topology).

In [15], Biane and Voiculescu have obtained the following transport inequality for the semicircular distribution σ

$$\mathcal{T}_2(\nu,\sigma) \leq 2\widetilde{\Sigma}(\nu|\sigma), \tag{60}$$

which holds for all $\nu\in\mathrm{P}(\mathbb{R})$ with compact support (see [15, Theorem 2.8]). The functional appearing in the left-hand side of (60) is the *relative free entropy* defined as follows:

$$\widetilde{\Sigma}(\nu|\sigma) = E(\nu) - E(\sigma),$$

where

$$E(\nu) = \int \frac{x^2}{2}\,d\nu(x) - \iint \log(|x-y|)\,d\nu(x)d\nu(y).$$

The relative free entropy $\widetilde{\Sigma}(\,\cdot\,|\sigma)$ is a natural candidate to replace the relative entropy $H(\,\cdot\,|\sigma)$, because it governs the large deviations of L_N when M_N is drawn from the GUE, as was shown by Ben Arous and Guionnet in [13]. More precisely, we have the following: for every open (resp. closed) subset O (resp. F) of $\mathrm{P}(\mathbb{R})$,

$$\liminf_{N\to\infty} \frac{1}{N^2} \log \mathbb{P}\left(L_N \in O\right) \geq -\inf\{\widetilde{\Sigma}(\nu|\sigma); \nu \in O\},$$
$$\limsup_{N\to\infty} \frac{1}{N^2} \log \mathbb{P}\left(L_N \in F\right) \leq -\inf\{\widetilde{\Sigma}(\nu|\sigma); \nu \in F\}.$$

Different approaches were considered to prove (60) and to generalize it to other compactly supported probability measures. The original proof by Biane and Voiculescu was inspired by [89]. Then Hiai, Petz and Ueda [60] proposed a simpler proof relying on Ben Arous and Guionnet large deviation principle. Later Ledoux gave alternative arguments based on a free analogue of the Brunn-Minkowski inequality. Recently, Ledoux and Popescu [91, 70] proposed yet another approach using optimal transport tools. Here, we will sketch the proof of Hiai, Petz and Ueda.

We need to introduce some supplementary material. Define $\mathcal{H}_N$ as the set of Hermitian $N \times N$ matrices. We will identify $\mathcal{H}_N$ with the space $\mathbb{R}^{N^2}$ using the map

$$H \in \mathcal{H}_N \mapsto \left((H(i,i))_i, (\mathrm{Re}(H(i,j)))_{i<j}, (\mathrm{Im}(H(i,j)))_{i<j}\right). \tag{61}$$

The Lebesgue measure dH on $\mathcal{H}_N$ is

$$dH := \prod_{i=1}^{N} dH_{i,i} \prod_{i<j} d\left(\mathrm{Re}(H_{i,j})\right) \prod_{i<j} d\left(\mathrm{Im}(H_{i,j})\right).$$

For all continuous function $Q : \mathbb{R} \to \mathbb{R}$, let us define the probability measure $P_{N,Q}$ on $\mathcal{H}_N$ by

$$\int f \, dP_{N,Q} := \frac{1}{Z_N(Q)} \int f(H) e^{-N\mathrm{Tr}(Q(H))} \, dH \tag{62}$$

for all bounded and measurable $f : \mathcal{H}_N \to \mathbb{R}$, where $Q(H)$ is defined using the basic functional calculus, and Tr is the trace operator. In particular, when M_N is drawn from the GUE, then it is easy to check that the law of $X_N = N^{-1/2}M_N$ is $P_{N,x^2/2}$.

The following theorem is due to Ben Arous and Guionnet.

Theorem 12.3. *Assume that $Q : \mathbb{R} \to \mathbb{R}$ is a continuous function such that*

$$\liminf_{|x|\to\infty} \frac{Q(x)}{\log|x|} > 2, \tag{63}$$

and for all $N \geq 1$ consider a random matrix $X_{N,Q}$ distributed according to $P_{N,Q}$. Let $\lambda_1^N \leq \ldots \leq \lambda_N^N$ be the ordered eigenvalues of $X_{N,Q}$ and define $L_N = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i^N} \in \mathrm{P}(\mathbb{R})$. The sequence of random measures $(L_N)_{N\geq 1}$ obeys a large deviation principle, in $\mathrm{P}(\mathbb{R})$ equipped with the weak topology, with speed N^2 and the good rate function I_Q defined by

$$I_Q(\nu) = E_Q(\nu) - \inf_\nu E_Q(\nu), \;\; \nu \in \mathrm{P}(\mathbb{R})$$

where

$$E_Q(\nu) = \int Q(x)\, d\nu(x) - \iint \log|x-y| \, d\nu(x)d\nu(y), \;\; \nu \in \mathrm{P}(\mathbb{R}).$$

In other words, for all open (resp. closed) O (resp. F) of $\mathrm{P}(\mathbb{R})$, it holds

$$\liminf_{N\to\infty} \frac{1}{N^2} \log \mathbb{P}\left(L_N \in O\right) \geq -\inf\{I_Q(\nu); \nu \in O\},$$
$$\limsup_{N\to\infty} \frac{1}{N^2} \log \mathbb{P}\left(L_N \in F\right) \leq -\inf\{I_Q(\nu); \nu \in F\}.$$

Moreover, the functional I_Q admits a unique minimizer denoted by μ_Q. The probability measure μ_Q is compactly supported and is characterized by the following two conditions: there is a constant $C_Q \in \mathbb{R}$ such that

$$Q(x) \geq 2\int \log|x-y|\, d\mu_Q(y) + C_Q, \quad \textit{for all } x \in \mathbb{R}$$

and

$$Q(x) = 2\int \log|x-y|\, d\mu_Q(y) + C_Q, \quad \textit{for all } x \in \mathrm{Supp}(\mu_Q).$$

Finally, the asymptotic behavior of the normalizing constant $Z_N(Q)$ in (62) is given by:

$$\lim_{N\to\infty} \frac{1}{N^2} \log Z_N(Q) = -E_Q(\mu_Q) = -\inf_{\nu} E_Q(\nu).$$

Remark 12.4. Let us make a few comments on this theorem.

(1) When $Q(x) = x^2/2$, then μ_Q is the semicircular law σ.
(2) For a general Q, one has the identity $I_Q(\nu) = E_Q(\nu) - E_Q(\mu_Q)$. So, to be coherent with the notation given at the beginning of this section, we will denote $I_Q(\nu) = \widetilde{\Sigma}(\nu|\mu_Q)$ in the sequel.
(3) As a by-product of the large deviation principle, we can conclude that the sequence of random measures L_N converges almost surely to μ_Q (for the weak topology). When $Q(x) = x^2/2$, this provides a proof of Wigner theorem in the particular case of the GUE.

Now we can prove the transport inequality (60).

Proof of (60). We will prove the inequality (60) only in the case where ν is a probability measure with support included in $[-A; A]$, $A > 0$ and such that the function

$$S_\nu(x) := 2\int \log|x-y|\, d\nu(y)$$

is finite and continuous over $\mathbb{R}$. The general case is then obtained by approximation (see [60] for explanations).

First step. To prove that $\mathcal{T}_2(\nu, \sigma) \leq 2\widetilde{\Sigma}(\nu|\sigma)$, the first idea is to use Theorem 12.3 to provide a matrix approximation of ν and σ.

Let $Q_\nu : \mathbb{R} \to \mathbb{R}$ be a continuous function such that $Q_\nu = S_\nu$ on $[-A, A]$, $Q_\nu \geq S_\nu$ and $Q_\nu(x) = \frac{x^2}{2}$ when $|x|$ is large. Let $X_{N,\nu}$, $N \geq 1$ be a sequence of random matrices distributed according to the probability $P_{N,\nu}$ associated to Q_ν in (62) (we shall write in the sequel $P_{N,\nu}$ instead of P_{N,Q_ν}). The characterization of the equilibrium measure μ_{Q_ν} easily implies that $\mu_{Q_\nu} = \nu$. So, the random empirical measures $L_{N,\nu}$ of $X_{N,\nu}$ follows the large deviation principle with the good rate function $\widetilde{\Sigma}(\,\cdot\,|\nu)$. In particular, $L_{N,\nu}$ converges almost surely to ν (for the weak topology). Let us consider the probability measure ν_N defined for all bounded measurable function f by

$$\int f\, d\nu_N := \mathbb{E}\left[\int f\, dL_{N,\nu}\right].$$

The almost sure convergence of $L_{N,\nu}$ to ν easily implies that ν_N converges to ν for the weak topology. We do the same construction with $Q_\sigma(x) = x^2/2$ yielding a sequence σ_N converging to σ (note that in this case, the sequences $X_{N,\sigma}$ and $P_{N,\sigma}$ correspond to the GUE rescaled by a factor $\sqrt{N}$).

Second step. Now we compare the Wasserstein distance between ν_N and σ_N to the one between $P_{N,\nu}$ and $P_{N,\sigma}$. To define the latter, we equip $\mathcal{H}_N$ with the Frobenius norm defined as follows:

$$\|A - B\|_F^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} |A(i,j) - B(i,j)|^2.$$

By definition, if P_1, P_2 are probability measures on $\mathcal{H}_N$, then

$$\mathcal{T}_2(P_1, P_2) := \inf \mathbb{E}\left[\|X - Y\|_F^2\right],$$

where the infimum is over all the couples of $N \times N$ random matrices (X, Y) such that X is distributed according to P_1 and Y according to P_2. According to the classical Hoffman-Wielandt inequality (see e.g [61]), if $A, B \in \mathcal{H}_N$ then,

$$\sum_{i=1}^{N} |\lambda_i(A) - \lambda_i(B)|^2 \le \|A - B\|_F^2,$$

where $\lambda_1(A) \le \lambda_2(A) \le \ldots \le \lambda_N(A)$ (resp. $\lambda_1(B) \le \lambda_2(B) \le \ldots \le \lambda_N(B)$) are the eigenvalues of A (resp. B) in increasing order. So, if $(X_{N,\nu}, X_{N,\sigma})$ is an optimal coupling between $P_{N,\nu}$ and $P_{N,\sigma}$, we have

$$\begin{aligned}\mathcal{T}_2(P_{N,\nu}, P_{N,\sigma}) = \mathbb{E}[\|X_{N,\nu} - X_{N,\sigma}\|_F^2] &\ge \mathbb{E}\left[\sum_{i=1}^{N} |\lambda_i(X_{N,\nu}) - \lambda_i(X_{N,\sigma})|^2\right]\\ &= N\mathbb{E}\left[\iint |x - y|^2\, dR_N\right],\end{aligned}$$

where R_N is the random probability measure on $\mathbb{R} \times \mathbb{R}$ defined by

$$R_N := \frac{1}{N}\sum_{i=1}^{N} \delta_{(\lambda_i(X_{N,\nu}),\lambda_i(X_N,\sigma))}.$$

It is clear that $\pi_N := \mathbb{E}[R_N]$ has marginals ν_N and σ_N. Hence, applying Fubini theorem in the above inequality yields

$$\mathcal{T}_2(P_{N,\nu}, P_{N,\sigma}) \ge N\mathcal{T}_2(\nu_N, \sigma_N).$$

Third step. If we identify the space $\mathcal{H}_N$ to the space $\mathbb{R}^{N^2}$ using the map defined in (61), then $P_{N,\sigma}$ is a product of Gaussian measures:

$$P_{N,\sigma} = \mathcal{N}(0, 1/N)^N \otimes \mathcal{N}(0, 1/(2N))^{N(N-1)/2} \otimes \mathcal{N}(0, 1/(2N))^{N(N-1)/2}.$$

Each factor verifies Talagrand inequality $\mathbf{T}_2$ (with the constant $2/N$ or $1/N$). Therefore, using the dimension-free tensorization property of $\mathbf{T}_2$, it is easy to check that $P_{N,\sigma}$ verifies the transport inequality $\mathcal{T}_c \leq 2N^{-1}H$ on $\mathcal{H}_N$, where the cost function c is defined by

$$c(A,B) := \sum_{i=1}^{N} |A(i,i) - B(i,i)|^2 + 2\sum_{i<j} |A(i,j) - B(i,j)|^2 = \|A-B\|_F^2.$$

As a conclusion, for all $N \geq 1$, the inequality $\mathcal{T}_2(P_{N,\nu}, P_{N,\sigma}) \leq 2N^{-1}H\left(P_{N,\nu}|P_{N,\sigma}\right)$ holds. Using Step 2, we get

$$\mathcal{T}_2(\nu_N, \sigma_N) \leq \frac{2}{N^2} H\left(P_{N,\nu}|P_{N,\sigma}\right), \ N \geq 1.$$

Fourth step. The last step is devoted to the computation of the limit of $N^{-2}H\left(P_{N,\nu}|P_{N,\sigma}\right)$ when N goes to ∞. We have

$$\begin{aligned}\frac{H\left(P_{N,\nu}|P_{N,\sigma}\right)}{N^2} &= \frac{1}{N^2}\log Z_N(Q_\sigma) - \frac{1}{N^2}\log Z_N(Q_\nu) + \frac{1}{N}\int \operatorname{Tr}\left(Q_\nu(A) - \frac{1}{2}A^2\right) dP_{N,\nu} \\ &= \frac{1}{N^2}\log Z_N(Q_\sigma) - \frac{1}{N^2}\log Z_N(Q_\nu) + \int Q_\nu(x) - \frac{x^2}{2}\, d\nu_N(x)\end{aligned}$$

Using Theorem 12.3 and the convergence of ν_N to ν, it is not difficult to see that the right-hand side tends to $\widetilde{\Sigma}(\nu|\sigma)$ when N goes to ∞ (observe that the function $x \mapsto Q_\nu(x) - x^2/2$ is continuous and has a compact support). Since $\mathcal{T}_2$ is lower semicontinuous (this is a direct consequence of the Kantorovich dual equality, see Theorem 2.2), $\mathcal{T}_2(\nu,\sigma) \leq \liminf_{N\to\infty} \mathcal{T}_2(\nu_N, \sigma_N)$, which completes the proof. □

Remark 12.5.
(1) It is possible to adapt the preceding proof to show that probability measures μ_Q with $Q'' \geq \rho$, with $\rho > 0$ verify the transport inequality $\mathcal{T}_2(\nu, \mu_Q) \leq \frac{2}{\rho}\widetilde{\Sigma}(\nu|\mu_Q)$, for all $\nu \in \mathrm{P}(\mathbb{R})$, see [60].
(2) The random matrix approximation method can be applied to obtain a free analogue of the logarithmic Sobolev inequality, see [14]. It has been shown by Ledoux in [69] that a free analogue of Otto-Villani theorem holds. Ledoux and Popescu have also obtained in [70] a free HWI inequality.

13. Optimal transport is a tool for proving other functional inequalities

We already saw in Section 7 that the logarithmic Sobolev inequality can be derived by means of the quadratic optimal transport. It has been discovered by Barthe, Cordero-Erausquin, McCann, Nazaret and Villani, among others, that this is also true for other well-known functional inequalities such as Prékopa-Leindler, Brascamp-Lieb and Sobolev inequalities, see [5, 6, 29, 30, 41, 83].

In this section, we do an excursion a step away from transport inequalities and visit Brunn-Minkowski and Prékopa-Leindler inequalities. We are going to sketch their proofs. Our main tool will be the Brenier map which was described at Theorem 2.9. For a concise and enlightening discussion on this topic, it is worth reading Villani's exposition in [103, Ch. 6].

The Prékopa-Leindler inequality. It is a functional version of Brunn-Minkowski inequality which has been proved several times and named after the papers by Prékopa [92] and Leindler [71].

Theorem 13.1 (Prékopa-Leindler inequality)**.** *Let f, g, h be three nonnegative integrable functions on $\mathbb{R}^k$ and $0 \le \lambda \le 1$ be such that for all $x, y \in \mathbb{R}^k$,*

$$h((1-\lambda)x + \lambda y) \ge f(x)^{1-\lambda} g(y)^{\lambda}.$$

Then,

$$\int_{\mathbb{R}^k} h(x)\,dx \ge \left(\int_{\mathbb{R}^k} f(x)\,dx \right)^{1-\lambda} \left(\int_{\mathbb{R}^k} g(x)\,dx \right)^{\lambda}.$$

The next proof comes from Barthe's PhD thesis [5].

Proof. Without loss of generality, assume that f, g and h are probability densities. Pick another probability density p on $\mathbb{R}^k$, for instance the indicator function of the unit cube $[0,1]^d$. By Theorem 2.9, there exist two Brenier maps $\nabla\Phi_1$ and $\nabla\Phi_2$ which transport p onto f and p onto g, respectively. Since Φ_1 is a convex function, it admits an Alexandrov Hessian (defined almost everywhere) $\nabla_A^2\phi_1$ which is nonnegative definite. Similarly, for Φ_2 and $\nabla_A^2\phi_2$. The change of variable formula leads us to the Monge-Ampère equations

$$f(\nabla\phi_1(x))\,\det(\nabla_A^2\phi_1(x)) = 1, \qquad g(\nabla\phi_2(x))\,\det(\nabla_A^2\phi_2(x)) = 1$$

for almost all $x \in [0,1]^d$. Defining $\phi = (1-\lambda)\phi_1 + \lambda\phi_2$, one obtains

$$\begin{aligned}
&\int_{\mathbb{R}^k} h(y)\,dy \\
\ge\; & \int_{[0,1]^d} h(\nabla\phi(x))\,\det(\nabla_A^2\phi(x))\,dx \\
\overset{\text{(i)}}{\ge}\; & \int_{[0,1]^d} h\Big((1-\lambda)\nabla\phi_1(x) + \lambda\nabla\phi_2(x)\Big) \Big[\det(\nabla_A^2\phi_1(x))\Big]^{1-\lambda} \Big[\det(\nabla_A^2\phi_2(x))\Big]^{\lambda} dx \\
\overset{\text{(ii)}}{\ge}\; & \int_{[0,1]^d} f(\nabla\phi_1(x))^{1-\lambda} g(\nabla\phi_2(x))^{\lambda} \Big[\det(\nabla_A^2\phi_1(x))\Big]^{1-\lambda} \Big[\det(\nabla_A^2\phi_2(x))\Big]^{\lambda} dx \\
\overset{\text{(iii)}}{=}\; & \int_{[0,1]^d} 1\,dx = 1
\end{aligned}$$

where inequality (i) follows from the claim below, inequality (ii) uses the assumption on f, g and h and the equality (iii) is a direct consequence of the above Monge-Ampère equations.

Claim. The function $S \in \mathcal{S}_+ \mapsto \log\det(S) \in [-\infty, \infty)$ is a concave function on the convex cone $\mathcal{S}_+$ of nonnegative definite symmetric matrices. □

The decisive trick of this proof is to take advantage of the concavity of $\log\det$, once it is noticed that the Hessian of the *convex* function ϕ, which gives rise to the Brenier map $\nabla\phi$, belongs to $\mathcal{S}_+$.

As a corollary, one obtains the celebrated Brunn-Minkowski inequality.

Corollary 13.2 (Brunn-Minkowski inequality). *For all A, B compact subsets of $\mathbb{R}^k$,*

$$\mathrm{vol}^{1/d}(A+B) \geq \mathrm{vol}^{1/d}(A) + \mathrm{vol}^{1/d}(B)$$

where $\mathrm{vol}^{1/d}(A) := \left(\int_A dx\right)^{1/d}$ *and* $A + B := \{a + b; a \in A, b \in B\}$.

Proof. For any $0 \leq \lambda \leq 1$, the functions $f = \mathbf{1}_A$, $g = \mathbf{1}_B$ and $h = \mathbf{1}_{[(1-\lambda)A+\lambda B]}$ satisfy Theorem 13.1's assumptions. Therefore, we have $\int h \geq (\int f)^{1-\lambda}(\int g)^\lambda$ which is $\mathrm{vol}((1-\lambda)A + \lambda B) \geq \mathrm{vol}(A)^{1-\lambda}\mathrm{vol}(B)^\lambda$. It follows that $\mathrm{vol}(A + B) = \mathrm{vol}((1-\lambda)\frac{A}{1-\lambda} + \lambda\frac{B}{\lambda}) \geq \mathrm{vol}(\frac{A}{1-\lambda})^{1-\lambda}\mathrm{vol}(\frac{B}{\lambda})^\lambda$ which is equivalent to $\mathrm{vol}^{1/d}(A+B) \geq \left(\frac{\mathrm{vol}^{1/d}(A)}{1-\lambda}\right)^{1-\lambda}\left(\frac{\mathrm{vol}^{1/d}(B)}{\lambda}\right)^{\lambda}$. It remains to optimize in λ. □

Appendix A. Tensorization of transport costs

During the proof of the tensorization property of transport-entropy inequalities at Proposition 1.8, we made use of the following tensorization property of transport costs. A detailed proof of this property in the literature being unknown to the authors, we find it useful to present it here.

Proposition A.1. *We assume that the cost functions c_1 and c_2 are lower semicontinous on the products of polish spaces $\mathcal{X}_1 \times \mathcal{Y}_1$ and $\mathcal{X}_2 \times \mathcal{Y}_2$, respectively. Then, for all $\nu \in \mathrm{P}(\mathcal{Y}_1 \times \mathcal{Y}_2)$, $\mu_1 \in \mathrm{P}(\mathcal{X}_1)$ and $\mu_2 \in \mathrm{P}(\mathcal{X}_2)$, we have*

$$\mathcal{T}_{c_1 \oplus c_2}(\nu, \mu_1 \otimes \mu_2) \leq \mathcal{T}_{c_1}(\nu_1, \mu_1) + \int_{\mathcal{Y}_1} \mathcal{T}_{c_2}(\nu_2^{y_1}, \mu_2)\, d\nu_1(y_1) \tag{64}$$

where ν disintegrates as follows: $d\nu(y_1, y_2) = d\nu_1(y_1) d\nu_2^{y_1}(y_2)$.

Proof. One first faces a nightmare of notation. It might be helpful to introduce random variables and see $\pi \in \mathrm{P}(\mathcal{X} \times \mathcal{Y}) = \mathrm{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}_1 \times \mathcal{Y}_2)$ as the law of (X_1, X_2, Y_1, Y_2). One denotes $\pi_1 = \mathcal{L}(X_1, Y_1)$, $\pi_2^{x_1,y_1}\mathcal{L}(X_2, Y_2 \mid X_1 = x_1, Y_1 = y_1)$, $\pi_{X_2}^{x_1,y_1} = \mathcal{L}(X_2 \mid X_1 = x_1, Y_1 = y_1)$, $\pi_{Y_2}^{x_1,y_1} = \mathcal{L}(Y_2 \mid X_1 = x_1, Y_1 = y_1)$, $\pi_X = \mathcal{L}(X_1, X_2)$, $\pi_Y = \mathcal{L}(Y_1, Y_2)$ and so on.

Let us denote $\Pi(\nu, \mu)$ the set of all $\pi \in \mathrm{P}(\mathcal{X} \times \mathcal{Y})$ such that $\pi_X = \nu$ and $\pi_Y = \mu$, $\Pi_1(\nu_1, \mu_1)$ the set of all $\eta \in \mathrm{P}(\mathcal{X}_1 \times \mathcal{Y}_1)$ such that $\eta_{X_1} = \nu_1$ and $\eta_{Y_1} = \mu_1$ and $\Pi_2(\nu_2, \mu_2)$ the set of all $\eta \in \mathrm{P}(\mathcal{X}_2 \times \mathcal{Y}_2)$ such that $\eta_{X_2} = \nu_2$ and $\eta_{Y_2} = \mu_2$.

We only consider couplings π such that under the law π

- $\mathcal{L}(X_1, X_2) = \nu$,
- $\mathcal{L}(Y_1, Y_2) = \mu$,
- Y_1 and X_2 are independent conditionally on X_1 and
- X_1 and Y_2 are independent conditionally on Y_1.

By the definition of the optimal cost, optimizing over this collection of couplings leads us to

$$\mathcal{T}_c(\nu, \mu) \leq \inf_{\pi_1, \pi_2^\diamond} \int c_1 \oplus c_2(x_1, y_1, x_2, y_2)\, d\pi_1(x_1, y_1) d\pi_2^{x_1,y_1}(x_2, y_2)$$

where the infimum is taken over all $\pi_1 \in \Pi_1(\nu_1, \mu_1)$ and all Markov kernels $\pi_2^\diamond = (\pi_2^{x_1,y_1}; x_1 \in \mathcal{X}_1, y_1 \in \mathcal{Y}_1)$ such that $\pi_2^{x_1,y_1} \in \Pi_2(\nu_{X_2}^{x_1}, \mu_{Y_2}^{y_1})$ for π_1-almost every (x_1, y_1). As μ is a tensor

product: $\mu=\mu_1\otimes\mu_2$, we have $\mu_{Y_2}^{y_1}=\mu_2$, π_1-a.e. so that $\pi_2^{x_1,y_1}\in\Pi_2(\nu_{X_2}^{x_1},\mu_2)$ for π_1-almost every (x_1,y_1).
We obtain

$$\begin{aligned}
\mathcal{T}_c(\nu,\mu) &\le \inf_{\pi_1,\pi_2^\diamond}\int c_1\oplus c_2(x_1,y_1,x_2,y_2)\,d\pi_1(x_1,y_1)d\pi_2^{x_1,y_1}(x_2,y_2)\\
&= \inf_{\pi_1}\left[\int_{\mathcal{X}_1\times\mathcal{Y}_1} c_1\,d\pi_1+\inf_{\pi_2^\diamond}\int_{\mathcal{X}_1\times\mathcal{Y}_1}\int_{\mathcal{X}_2\times\mathcal{Y}_2} c_2(x_2,y_2)d\pi_2^{x_1,y_1}(x_2,y_2)\,d\pi_1(x_1,y_1)\right]\\
&\overset{(a)}{=} \inf_{\pi_1}\left[\int_{\mathcal{X}_1\times\mathcal{Y}_1} c_1\,d\pi_1+\int_{\mathcal{X}_1\times\mathcal{Y}_1}\left(\inf_{\pi_2^\diamond}\int_{\mathcal{X}_2\times\mathcal{Y}_2} c_2(x_2,y_2)d\pi_2^{x_1,y_1}(x_2,y_2)\right)\,d\pi_1(x_1,y_1)\right]\\
&\overset{(b)}{=} \inf_{\pi_1}\left[\int_{\mathcal{X}_1\times\mathcal{Y}_1} c_1\,d\pi_1+\int_{\mathcal{X}_1\times\mathcal{Y}_1}\mathcal{T}_{c_2}(\nu_{X_2}^{x_1},\mu_2)\,d\pi_1(x_1,y_1)\right]\\
&= \inf_{\pi_1}\left\{\int_{\mathcal{X}_1\times\mathcal{Y}_1} c_1\,d\pi_1\right\}+\int_{\mathcal{X}_1}\mathcal{T}_{c_2}(\nu_{X_2}^{x_1},\mu_2)\,d\nu_1(x_1)\\
&= \mathcal{T}_{c_1}(\nu_1,\mu_1)+\int_{\mathcal{X}_1}\mathcal{T}_{c_2}(\nu_{X_2}^{x_1},\mu_2)\,d\nu_1(x_1)
\end{aligned}$$

which is the desired result.
Equality (a) is not that obvious. First of all, one is allowed to commute $\inf_{\pi_2^\diamond}$ and $\int_{\mathcal{X}_1\times\mathcal{Y}_1}$ since $\pi_2^\diamond$ lives in a rich enough family for being able to optimize separately for each (x_1,y_1). But also, one must check that after commuting, the integrand $\inf_{\pi_2^\diamond}\int_{\mathcal{X}_2\times\mathcal{Y}_2} c_2(x_2,y_2)d\pi_2^{x_1,y_1}(x_2,y_2)$ is measurable as a function of (x_1,y_1). But for each fixed (x_1,y_1), this integrand is the optimal transport cost $\mathcal{T}_{c_2}(\nu_{X_2}^{x_1},\mu_2)$ (this is the content of equality (b)). Now, with the Kantorovich dual equality (14), one sees that $\mathcal{T}_c$ is a lower semicontinuous function as the supremum of a family of continuous functions. A fortiori, $\mathcal{T}_{c_2}$ is measurable on $\mathrm{P}(\mathcal{X}_2)\times\mathrm{P}(\mathcal{Y}_2)$ and $(x_1,y_1)\mapsto\mathcal{T}_{c_2}(\nu_{X_2}^{x_1},\mu_2)$ is also measurable as a composition of measurable functions (use the polish assumption for the existence of measurable Markov kernels). This completes the proof of the proposition. □

Let us have a look at another tensorization result which appears in [57]. On the polish product space $\mathcal{X}^{(n)}:=\prod_{i=1}^n\mathcal{X}_i$, consider the cost function

$$\oplus_i c_i(x,y):=\sum_{i=1}^n c(x_i,y_i),\ \forall x,y\in\mathcal{X}^{(n)}$$

where for each index i, c_i is lower semicontinuous on $\mathcal{X}_i^2$. Let $(X_i)_{1\le i\le n}$ be the canonical process on $\mathcal{X}^{(n)}=\prod_{i=1}^n\mathcal{X}_i$. For each i, $\widetilde{X}_i=(X_j)_{1\le j\le n; j\ne i}$ is the configuration without its value at index i. Given a probability measure ν on $\mathcal{X}^{(n)}$,

$$\nu_i^{\widetilde{x}_i}=\nu(X_i\in\cdot|\widetilde{X}_i=\widetilde{x}_i)$$

denotes the regular conditional distribution of X_i knowing that $\widetilde{X}_i=\widetilde{x}_i$ under ν and

$$\nu_i=\nu(X_i\in\cdot)$$

denotes the i-th marginal of ν.

Proposition A.2. *Let $\mu = \bigotimes_{i=1}^n \mu_i$ be a product probability measure on $\mathcal{X}^{(n)}$. For all $\nu \in \mathrm{P}(\mathcal{X}^{(n)})$,*

$$\mathcal{T}_{\oplus c_i}(\nu,\mu) \le \int_{\mathcal{X}^{(n)}} \left(\sum_{i=1}^n \mathcal{T}_{c_i}(\nu_i^{\widetilde{x}_i}, \mu_i) \right) d\nu(x).$$

Proof. • *A first sketch.* Let $(W_i)_{1\le i\le n} = (U_i,V_i)_{1\le i\le n}$ be a sequence of random variables taking their values in $\prod_{i=1}^n \mathcal{X}_i^2$ which is defined on some probability space $(\Omega,\mathbb{P})$ so that it realizes $\mathcal{T}_{\oplus c_i}(\nu,\mu)$. This means that the law of $U=(U_i)_{1\le i\le n}$ is ν, the law of $V=(V_i)_{1\le i\le n}$ is μ and $\mathbb{E}\sum_i c_i(U_i,V_i) = \mathcal{T}_{\oplus c_i}(\nu,\mu)$.

Let i be a fixed index. There exists a couple of random variables $\widehat{W}_i := (\widehat{U}_i, \widehat{V}_i)$ such that its conditional law given $(\widetilde{U}_i, \widetilde{V}_i) = \widetilde{W}_i := (W_j)_{j\ne i}$ is a coupling of $\nu_i^{\widetilde{U}_i}$ and $\mu_i^{\widetilde{V}_i}$, and $\mathbb{P}$-a.s., $\mathbb{E}[c_i(\widehat{U}_i,\widehat{V}_i)|\widetilde{W}_i] = \mathcal{T}_{c_i}\left(\nu_i^{\widetilde{U}_i}, \mu_i^{\widetilde{V}_i}\right)$. This implies

$$\mathbb{E}c_i(\widehat{U}_i,\widehat{V}_i) = \mathbb{E}\mathcal{T}_{c_i}\left(\nu_i^{\widetilde{U}_i}, \mu_i^{\widetilde{V}_i}\right). \tag{65}$$

Clearly, $\left[(\widetilde{U}_i,\widehat{U}_i);(\widetilde{V}_i,\widehat{V}_i)\right]$ is a coupling of (ν,μ). The optimality of W gives us $\mathbb{E}\sum_j c_j(U_j,V_j) \le \mathbb{E}\left(\sum_{j\ne i} c_j(U_j,V_j) + c_i(\widehat{U}_i,\widehat{V}_i)\right)$ which boils down to

$$\mathbb{E}c_i(U_i,V_i) \le \mathbb{E}c_i(\widehat{U}_i,\widehat{V}_i) = \mathbb{E}\mathcal{T}_{c_i}\left(\nu_i^{\widetilde{U}_i}, \mu_i^{\widetilde{V}_i}\right) \tag{66}$$

where the equality is (65). Summing over all the indices i, we see that

$$\mathcal{T}_{\oplus c_i}(\nu,\mu) = \mathbb{E}\sum_{i=1}^n \ c_i(U_i,V_i) \le \mathbb{E}\sum_{i=1}^n \mathcal{T}_{c_i}\left(\nu_i^{\widetilde{U}_i}, \mu_i^{\widetilde{V}_i}\right).$$

As μ is a product measure, we have $\mu_i^{\widetilde{V}_i} = \mu_i$, $\mathbb{P}$-almost surely and we obtain

$$\mathcal{T}_{\oplus c_i}(\nu,\mu) \le \int_{\mathcal{X}^{(n)}} \sum_{i=1}^n \mathcal{T}_{c_i}(\nu_i^{\widetilde{x}_i},\mu_i)\, d\nu(x)$$

which is the announced result.

• *Completion of the proof.* This first part is an incomplete proof, since one faces a measurability problem when constructing the conditional optimal coupling $\widehat{W}_i := (\widehat{U}_i,\widehat{V}_i)$. This measurability is needed to take the expectation in (65). More precisely, it is true that for each value $\widetilde{w}_i$ of $\widetilde{W}_i$, there exists a coupling $\widehat{W}_i(\tilde{w}_i)$ of $\nu_i^{\widetilde{u}_i}$ and $\mu_i^{\widetilde{v}_i}$. But the dependence in $\widetilde{w}_i$ must be Borel measurable for $\widehat{W}_i = \widehat{W}_i(\widetilde{W}_i)$ to be a random variable.
One way to circumvent this problem is to proceed as in Proposition A.1. The important features of this proof are:

(1) The Markov kernels $\nu_i^{\widetilde{u}_i}$ and $\mu_i^{\widetilde{v}_i}$ are built with conditional independence properties as in Proposition A.1's proof. More precisely
  - $\widetilde{V}_i$ and U_i are independent conditionally on $\widetilde{U}_i$ and
  - $\widetilde{U}_i$ and V_i are independent conditionally on $\widetilde{V}_i$.

  These kernels admit measurable versions since the state space is polish.

(2) The measurability is not required at the level of the optimal coupling but only through the optimal cost.

This leads us to (66). We omit the details of the proof which is a variation on Proposition A.1's proof with another "nightmare of notation". □

Proposition A.2 differs from Marton's original result [78] which requires an ordering of the indices.

Appendix B. Variational representations of the relative entropy

At Section 3, we took great advantage of the variational representations of $\mathcal{T}_c$ and the relative entropy. Here, we give a proof of the variational representation formulae (24) and (25) of the relative entropy.

Proposition B.1. *For all $\nu \in \mathrm{P}(\mathcal{X})$,*

$$\begin{aligned} H(\nu|\mu) &= \sup\left\{\int u\,d\nu - \log\int e^u\,d\mu; u \in \mathcal{C}_b(\mathcal{X})\right\}. \\ &= \sup\left\{\int u\,d\nu - \log\int e^u\,d\mu; u \in \mathcal{B}_b(\mathcal{X})\right\} \end{aligned} \tag{67}$$

and for all $\nu \in \mathrm{P}(\mathcal{X})$ such that $\nu \ll \mu$,

$$H(\nu|\mu) = \sup\left\{\int u\,d\nu - \log\int e^u\,d\mu; u: \textit{measurable}, \int e^u\,d\mu < \infty, \int u_-\,d\nu < \infty\right\} \tag{68}$$

where $u_- = (-u) \vee 0$ and $\int u\,d\nu \in (-\infty, \infty]$ is well-defined for all u such that $\int u_-\,d\nu < \infty$.

Proof. Once we have (68), (67) follows by standard approximation arguments.
The proof of (68) relies on Fenchel inequality for the convex function $h(t) = t\log t - t + 1$:

$$st \le (t\log t - t + 1) + (e^s - 1)$$

for all $s \in [-\infty, \infty)$, $t \in [0, \infty)$, with the conventions $0\log 0 = 0$, $e^{-\infty} = 0$ and $-\infty \times 0 = 0$ which are legitimated by limiting procedures. The equality is attained when $t = e^s$.

Taking $s = u(x)$, $t = \frac{d\nu}{d\mu}(x)$ and integrating with respect to μ leads us to

$$\int u\,d\nu \le H(\nu|\mu) + \int (e^u - 1)\,d\mu,$$

whose terms are meaningful with values in $(-\infty, \infty]$, provided that $\int u_-\,d\nu < \infty$. Formally, the case of equality corresponds to $\frac{d\nu}{d\mu} = e^u$. With the monotone convergence theorem, one sees that it is approached by the sequence $u_n = \log(\frac{d\nu}{d\mu} \vee e^{-n})$, as n tends to infinity. This gives us $H(\nu|\mu) = \sup\left\{\int u\,d\nu - \int(e^u - 1)\,d\mu; u: \int e^u\,d\mu < \infty, \inf u > -\infty\right\}$, which in turn implies that

$$H(\nu|\mu) = \sup\left\{\int u\,d\nu - \int(e^u - 1)\,d\mu; u: \int e^u\,d\mu < \infty, \int u_-\,d\nu < \infty\right\},$$

since the integral $\int \log(d\nu/d\mu)\,d\nu = \int h(d\nu/d\mu)\,d\mu \in [0, \infty]$ is well-defined.

Now, we take advantage of the unit mass of $\nu \in \mathrm{P}(\mathcal{X})$:

$$\int (u+b)\,d\nu - \int (e^{(u+b)}-1)\,d\mu = \int u\,d\nu - e^b \int e^u\,d\mu + b + 1, \quad \forall b \in \mathbb{R},$$

and we use the easy identity $\log a = \inf_{b\in\mathbb{R}}\{ae^b - b - 1\}$ to obtain

$$\sup_{b\in\mathbb{R}} \left\{ \int (u+b)\,d\nu - \int (e^{(u+b)}-1)\,d\mu \right\} = \int u\,d\nu - \log \int e^u\,d\mu.$$

Whence,

$$\begin{aligned}
& \sup\left\{ \int u\,d\nu - \int (e^u - 1)\,d\mu; u : \int e^u\,d\mu < \infty, \int u_-\,d\nu < \infty \right\} \\
= \; & \sup\left\{ \int (u+b)\,d\nu - \int (e^{(u+b)}-1)\,d\mu; b\in\mathbb{R}, u : \int e^u\,d\mu < \infty, \int u_-\,d\nu < \infty \right\} \\
= \; & \sup\left\{ \int u\,d\nu - \log\int e^u\,d\mu; u : \int e^u\,d\mu < \infty, \int u_-\,d\nu < \infty \right\}.
\end{aligned}$$

This completes the proof of (68). □

References

[1] G. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[2] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000.

[3] D. Bakry, F. Barthe, P. Cattiaux, and A. Guillin. A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case. *Electron. Commun. Probab.*, 13:60–66, 2008.

[4] D. Bakry and M. Emery. Diffusions hypercontractives. In *Séminaire de probabilités*, volume 1123 of *Lecture Notes in Mathematics*, pages 177–206. Springer-Verlag, 1985.

[5] F. Barthe. *Inégalités fonctionnelles et géométriques obtenues par transport de mesures*. PhD thesis, Univ. Marne-la-Vallée, 1997.

[6] F. Barthe. On a reverse form of the brascamp-lieb inequality. *Invent. Math.*, 134(2):335–361, 1998.

[7] F. Barthe, P. Cattiaux, and C. Roberto. Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry. *Rev. Mat. Iberoamericana*, 22(3):993–1067, 2006.

[8] F. Barthe and A. V. Kolesnikov. Mass transport and variants of the logarithmic Sobolev inequality. *J. Geom. Anal.*, 18(4):921–979, 2008.

[9] F. Barthe and C. Roberto. Sobolev inequalities for probability measures on the real line. *Studia Math.*, 159(3):481–497, 2003.

[10] F. Barthe and C. Roberto. Modified logarithmic Sobolev inequalities on $\mathbb{R}$. *Potential Anal.*, 29(2):167–193, 2008.

[11] W. Beckner. A generalized Poincaré inequality for Gaussian measures. *Proc. Amer. Math. Soc.*, 105:397–400, 1989.

[12] M. Beiglböck and W. Schachermayer. Duality for Borel measurable cost functions. To appear in Trans. Amer. Math. Soc.

[13] G. Ben Arous and A. Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probab. Theory Related Fields*, 108(4):517–542, 1997.

[14] P. Biane. Logarithmic Sobolev inequalities, matrix models and free entropy. *Acta Math. Sin. (Engl. Ser.)*, 19(3):497–506, 2003. International Workshop on Operator Algebra and Operator Theory (Linfen, 2001).

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

[16] S. Bobkov, I. Gentil, and M. Ledoux. Hypercontractivity of Hamilton-Jacobi equations. *J. Math. Pures Appl. (9)*, 80(7):669–696, 2001.

[17] S. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999.

[18] S. Bobkov and C. Houdré. Isoperimetric constants for product probability measures. *Ann. Probab.*, 25(1):184–205, 1997.

[19] S. Bobkov and M. Ledoux. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields*, 107(3):383–400, 1997.

[20] S. Bobkov and M. Ledoux. From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.*, 10(5):1028–1052, 2000.

[21] S. Bobkov and B. Zegarlinski. Entropy bounds and isoperimetry. *Mem. Amer. Math. Soc.*, 176(829):x+69, 2005.

[22] F. Bolley and C. Villani. Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Ann. Fac. Sci. Toulouse Math. (6)*, 14(3):331–352, 2005.

[23] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44:375–417, 1991.

[24] E. Carlen. Superadditivity of Fisher's information and logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 101(1):194–211, 1991.

[25] P. Cattiaux and A. Guillin. Talagrand's like quadratic transportation cost inequalities. *J. Math. Pures Appl.*, 86(9):341–361, 2006.

[26] P. Cattiaux, A. Guillin, and L-M. Wu. A note on Talagrand transportation inequality and logarithmic Sobolev inequality. to appear in Probab. Theory Related Fields, 2008.

[27] D. Cordero Erausquin. Some applications of mass transport to Gaussian-type inequalities. *Arch. Ration. Mech. Anal.*, 161(3):257–269, 2002.

[28] D. Cordero Erausquin, W. Gangbo, and C. Houdré. Inequalities for generalized entropy and optimal transportation. In *Recent advances in the theory and applications of mass transport*, volume 353 of *Contemp. Math.*, pages 73–94. Amer. Math. Soc., Providence, RI, 2004.

[29] D. Cordero Erausquin, R. McCann, and M. Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.*, 146(2):219–257, 2001.

[30] D. Cordero Erausquin, R. McCann, and M. Schmuckenschläger. Prékopa-Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport. *Ann. Fac. Sci. Toulouse Math. (6)*, 15(4):613–635, 2006.

[31] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Stud. Sci. Math. Hung.*, 2:299–318, 1967.

[32] A. Dembo. Information inequalities and concentration of measure. *Ann. Probab.*, 25(2):927–939, 1997.

[33] A. Dembo and O. Zeitouni. Transportation approach to some concentration inequalities in product spaces. *Electron. Comm. Probab.*, 1:no. 9, 83–90 (electronic), 1996.

[34] A. Dembo and O. Zeitouni. *Large deviations techniques and applications. Second edition.* Applications of Mathematics 38. Springer Verlag, 1998.

[35] J-D. Deuschel and D. Stroock. *Large Deviations*, volume 137 of *Pure and Applied Mathematics*. Academic Press, 1989.

[36] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.*, 32(3B):2702–2732, 2004.

[37] R. Dobrushin. The description of a random field by means of conditional probabilities and condition of itd regularity. *Theor. Probab. Appl.*, 13:197–224, 1968.

[38] R. Dobrushin. Prescribing a system of random variables by conditional distributions. *Theor. Probab. Appl.*, 15:458–486, 1970.

[39] M. Donsker and S. Varadhan. Asymptotic evaluations of certain Markov process expectations for large time, I. *Comm. Pure Appl. Math.*, 28:1–47, 1975.

[40] R. Dudley. *Real Analysis and Probability*, volume 74 of *Cambridge studies in advanced mathematics*. Cambridge University Press, 2002.

[41] D. Cordero Erausquin, B. Nazaret, and C. Villani. A mass-transportation approach to sharp Sobolev inequalities and Gagliardo-Nirenberg inequalities. *Adv. Math.*, 182(2):307–332, 2004.

[42] D. Feyel and A.S. Üstünel. Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space. *Probab. Theory Related Fields*, 128(3):347–385, 2004.

[43] H. Föllmer. *Random fields and diffusion processes, in Ecole d'Eté de Probabilités de Saint-Flour XV-XVII-1985-87*, volume 1362 of *Lecture Notes in Mathematics*. Springer, Berlin, 1988.

[44] W. Gangbo and R. McCann. The geometry of optimal transportation. *Acta Math.*, 177:113–161, 1996.

[45] F. Gao, A. Guillin, and L. Wu. Bernstein type's concentration inequalities for symmetric Markov processes. Preprint, 2009.

[46] F. Gao and L. Wu. Transportation-information inequalities for Gibbs measures. Preprint, 2007.

[47] I. Gentil, A. Guillin, and L. Miclo. Modified logarithmic Sobolev inequalities and transportation inequalities. *Probab. Theory Related Fields*, 133(3):409–436, 2005.

[48] I. Gentil and F. Malrieu. Equation de Hamilton-Jacobi et inégalité entropiques généralisées. *C. R. Acad. Sci. Paris*, 335:437–440, 2002.

[49] N. Gozlan. Integral criteria for transportation cost inequalities. *Electron. Comm. Probab.*, 11:64–77, 2006.

[50] N. Gozlan. Characterization of Talagrand's like transportation-cost inequalities on the real line. *J. Funct. Anal.*, 250(2):400–425, 2007.

[51] N. Gozlan. Poincaré inequality and dimension free concentration of measure. To appear in Ann. Inst. H. Poincaré Probab. Statist., 2008.

[52] N. Gozlan. A characterization of dimension free concentration in terms of transport inequalities. *Ann. Probab.*, 37(6):2480–2498, 2009.

[53] N. Gozlan and C. Léonard. A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields*, 139(1-2):235–283, 2007.

[54] N. Gozlan, C. Roberto, and P-M. Samson. A refined version of Otto and Villani theorem. In preparation, 2009.

[55] M. Gromov and V.D. Milman. A topological application of the isoperimetric inequality. *Amer. J. Math.*, 105:843–854, 1983.

[56] A. Guillin, C. Léonard, F. Wang, and L. Wu. Transportation-information inequalities for markov processes (II) : relations with other functional inequalities. Preprint, http://arxiv.org/abs/0902.2101, 2008.

[57] A. Guillin, C. Léonard, L. Wu, and N. Yao. Transportation-information inequalities for Markov processes. *Probab. Th. Rel. Fields*, 144:669–695, 2009.

[58] A. Guionnet. *Large random matrices: lectures on macroscopic asymptotics*, volume 1957 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2009. Lectures from the 36th Probability Summer School held in Saint-Flour, 2006.

[59] A. Guionnet and B. Zegarlinski. *Lectures on logarithmic Sobolev inequalities*, volume 1801 of *Lecture Notes in Mathematics*. Springer, 2003. Séminaire de probabilités XXXVI.

[60] F. Hiai, D. Petz, and Y. Ueda. Free transportation cost inequalities via random matrix approximation. *Probab. Theory Related Fields*, 130(2):199–221, 2004.

[61] R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1994. Corrected reprint of the 1991 original.

[62] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete Comput. Geom.*, 13(3-4):541–559, 1995.

[63] M. Knott and C. Smith. On the optimal mapping of distributions. *J. Optim. Theory Appl.*, 43(1):39–49, 1984.

[64] S. Kullback. A lower bound for discrimination information in terms of variation. *IEEE Trans. Inform. Theory*, 4:126–127, 1967.

[65] R. Latała and K. Oleszkiewicz. Between Sobolev and Poincaré. In *Geometric aspects of functional analysis*, volume 1745 of *Lecture Notes in Math.*, pages 147–168. Springer, Berlin, 2000.

[66] R. Latała and J. Wojtaszczyk. On the infimum convolution inequality. *Studia Math.*, 189(2):147–187, 2008.

[67] M. Ledoux. On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.*, 1:63–87, 1996.

[68] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[69] M. Ledoux. A (one-dimensional) free Brunn-Minkowski inequality. *C. R. Math. Acad. Sci. Paris*, 340(4):301–304, 2005.
[70] M. Ledoux and I. Popescu. Mass transportation proofs of free functional inequalities, and free Poincaré inequalities. *J. Funct. Anal.*, 257(4):1175–1221, 2009.
[71] L. Leindler. On a certain converse of Hölder's inequality. In *Linear operators and approximations (Proc. Conf. Oberwolfach, 1971)*, volume 20 of *International Series Numer. Math.*, pages 182–184, Basel, 1972. Birkhaüser.
[72] C. Léonard. A saddle-point approach to the Monge-Kantorovich optimal transport problem. *ESAIM-COCV (Control, Optimization and Calculus of Variations)*. To appear. Preprint downloadable from `www.cmap.polytechnique.fr/~leonard`.
[73] J. Lott and C. Villani. Hamilton-Jacobi semigroup on length spaces and applications. *J. Math. Pures Appl. (9)*, 88(3):219–229, 2007.
[74] Y. Ma, S. Shen, X. Wang, and L. Wu. Transportation inequalities: From Poisson to Gibbs measures. To appear in Stoch. Proc. Appl.
[75] F. Martinelli. *Lectures on Glauber dynamics for discrete spin models*, volume 1717 of *Lecture Notes in Mathematics*. Springer, 1999. École d'été de Saint-Flour (1997).
[76] K. Marton. A simple proof of the blowing-up lemma. *IEEE Trans. Inform. Theory*, 32(3):445–446, 1986.
[77] K. Marton. Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration. *Ann. Probab.*, 24(2):857–866, 1996.
[78] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6(3):556–571, 1996.
[79] K. Marton. Measure concentration for a class of random processes. *Probab. Theory Related Fields*, 110(3):427–439, 1998.
[80] K. Marton. Measure concentration for Euclidean distance in the case of dependent random variables. *Ann. Probab.*, 32(3B):2526–2544, 2004.
[81] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.
[82] B. Maurey. Some deviation inequalities. *Geom. Funct. Anal.*, 1(2):188–197, 1991.
[83] R. McCann. *A convexity theory for interacting gases and equilibrium crystals*. PhD thesis, Princeton Univ., 1994.
[84] R. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80:309–323, 1995.
[85] E. Milman. Isoperimetric and concentration inequalities - part i: Equivalence under curvature lower bound. http://arxiv.org/abs/0902.1560, 2009.
[86] E. Milman. On the role of convexity in functional and isoperimetric inequalities. *Proc. Lond. Math. Soc. (3)*, 99(1):32–66, 2009.
[87] E. Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.*, 177(1):1–43, 2009.
[88] B. Muckenhoupt. Hardy's inequality with weights. *Studia Math.*, 44:31–38, 1972.
[89] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.
[90] M. Pinsker. *Information and information stability of random variables and processes*. Holden-Day, San Francisco, 1964.
[91] I. Popescu. Talagrand inequality for the semicircular law and energy of the eigenvalues of beta ensembles. *Math. Res. Lett.*, 14(6):1023–1032, 2007.
[92] A. Prékopa. On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)*, 34:335–343, 1973.
[93] G. Royer. *Une initiation aux inégalités de Sobolev logarithmiques*, volume 5 of *Cours spécialisés*. Société Mathématique de France, 1999.
[94] P-M. Samson. Concentration of measure inequalities for Markov chains and Φ-mixing processes. *Ann. Probab.*, 28(1):416–461, 2000.
[95] P-M. Samson. Concentration inequalities for convex functions on product spaces. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 33–52. Birkhäuser, Basel, 2003.

[96] P-M Samson. Infimum-convolution description of concentration properties of product probability measures, with applications. *Ann. Inst. H. Poincaré Probab. Statist.*, 43(3):321–338, 2007.
[97] D. Stroock and B. Zegarlinski. The equivalence between the logarithmic Sobolev inequality and the Dobrushin-Shlosman mixing condition. *Comm. Math. Physics*, 144:303–323, 1992.
[98] D. Stroock and B. Zegarlinski. The logarithmic Sobolev inequality for discrete spin systems on the lattice. *Comm. Math. Physics*, 149:175–193, 1992.
[99] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'I.H.E.S.*, 81:73–203, 1995.
[100] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.
[101] M. Talagrand. A new look at independence. *Ann. Probab.*, 24(1):1–34, 1996.
[102] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3):587–600, 1996.
[103] C. Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics 58. American Mathematical Society, Providence RI, 2003.
[104] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
[105] F-Y. Wang. Logarithmic Sobolev inequalities on noncompact Riemannian manifolds. *Probab. Theory Related Fields*, 109(3):417–424, 1997.
[106] F-Y. Wang. Functional inequalities for empty essential spectrum. *J. Funct. Anal.*, 170(1):219–245, 2000.
[107] F-Y. Wang. Probability distance inequalities on Riemannian manifolds and path spaces. *J. Funct. Anal.*, 206(1):167–190, 2004.
[108] F-Y. Wang. A generalization of Poincaré and log-Sobolev inequalities. *Potential Anal.*, 22(1):1–15, 2005.
[109] F-Y. Wang. From super Poincaré to weighted log-Sobolev and entropy-cost inequalities. *J. Math. Pures Appl. (9)*, 90(3):270–285, 2008.
[110] F-Y. Wang. Generalized transportation-cost inequalities and applications. *Potential Anal.*, 28(4):321–334, 2008.
[111] L. Wu. Uniformly integrable operators and large deviations for Markov processes. *J. Funct. Anal*, 172:301–376, 2000.
[112] L. Wu. Poincaré and transportation inequalities for Gibbs measures under the Dobrushin uniqueness condition. *Ann. Probab.*, 34(5):1960–1989, 2006.
[113] L. Wu. Transportation inequalities for stochastic differential equations of pure jumps. Preprint, 2009.
[114] B. Zegarlinski. Dobrushin uniqueness theorem and logarithmic Sobolev inequalities. *J. Funct. Anal.*, 105:77–111, 1992.

Laboratoire d'analyse et mathématiques appliquées, UMR CNRS 8050. Université Paris Est. 5 bd Descartes, 77454 Marne la Vallée, France

E-mail address: `nathael.gozlan@univ-mlv.fr`

Modal-X. Université Paris Ouest. Bât. G, 200 av. de la République. 92001 Nanterre, France

E-mail address: `christian.leonard@u-paris10.fr`