The Annals of Probability
2009, Vol. 37, No. 6, 2480–2498
DOI: 10.1214/09-AOP470

A CHARACTERIZATION OF DIMENSION FREE CONCENTRATION IN TERMS OF TRANSPORTATION INEQUALITIES

BY NATHAEL GOZLAN

Université Paris-Est

The aim of this paper is to give a characterization of the dimension free concentration of measure phenomenon in terms of transportation-cost inequalities. We apply this theorem to give a new and very short proof of a result by Otto and Villani. Another application is to show that the Poincaré inequality is equivalent to a certain form of dimension free exponential concentration. The proofs of all these results rely on simple Large Deviations techniques.

1. Introduction. One says that a probability measure μ on $\mathbf{R}^k$ has the Gaussian dimension free concentration property if there are two nonnegative constants a and M such that for every positive integer n, the product measure μ^n verifies the following inequality:

$$\forall r \geq 0 \qquad \mu^n(A + rB_2) \geq 1 - Me^{-ar^2}, \tag{1.1}$$

for all measurable subset A of $(\mathbf{R}^k)^n$ with $\mu^n(A) \geq 1/2$ denoting by B_2 the Euclidean unit ball of $(\mathbf{R}^k)^n$. Here $A + rB_2 = \{x + y : x \in A, y \in rB_2\}$ is the Minkowski sum of A and rB_2.

The first example is of course the standard Gaussian measure on $\mathbf{R}$ for which the inequality (1.1) holds true with the sharp constants $M = 1/2$ and $a = 1/2$. Gaussian concentration is not the only possible behavior; for example, if $p \in [1, 2]$, the probability measure $d\mu_p(x) = Z_p^{-1} e^{-|x|^p}\, dx$ verifies a concentration inequality similar to (1.1) with r^2 replaced by $\min(r^p, r^2)$. In recent years many authors developed various functional approaches to the concentration of measure phenomenon. For example, the logarithmic-Sobolev inequality is well known to imply (1.1); this is the renowned Herbst argument (which is explained, for example, in Chapter 5 of Ledoux's book [28]). Among the many functional inequalities yielding concentration estimates, let us mention the following: Poincaré inequalities [8, 23], logarithmic-Sobolev inequalities [7, 27], modified logarithmic-Sobolev inequalities [4, 8, 10, 18], transportation-cost inequalities [6, 7, 9, 20, 30, 33–35], inf-convolution inequalities [26, 32], Beckner–Latała–Oleskiewicz inequalities [2, 3, 5, 25]. Several surveys and monographs are now available on this topic (see, for

Received October 2008; revised February 2009.
AMS 2000 subject classifications. Primary 60E15, 60F10, 26D10.
Key words and phrases. Concentration of measure, transportation-cost inequalities, Sanov's theorem, logarithmic-Sobolev inequalities.

instance, [1, 28] or [36, 37]). This large variety of tools and points of view raises the following natural question: is one of these functional inequalities equivalent to, say, (1.1)?

In this paper, one shows with a great generality that Talagrand's transportation-cost inequalities are equivalent to a dimension free concentration of measure. To state our main result, let us introduce some definitions and notation.

In all the sequel, $(\mathcal{X}, d)$ is a Polish space, $\mathrm{P}(\mathcal{X})$ is the set of all Borel probability measures on $\mathcal{X}$ and $\alpha : \mathbf{R}^+ \to \mathbf{R}^+$ is a convex function with $\alpha(0) = 0$. It will always be assumed that α verifies the following doubling property: there is some $K \geq 1$ such that

$$\forall t \geq 0 \qquad \alpha(2t) \leq K\alpha(t). \tag{1.2}$$

DEFINITION 1.1. Let $\mu \in \mathrm{P}(\mathcal{X})$; one says that μ verifies the dimension free concentration property $\mathbf{C}_\alpha(a)$ for some $a > 0$ if there is some $M > 0$ such that for all $n \geq 1$ and all measurable $A \subset \mathcal{X}^n$ with $\mu^n(A) \geq 1/2$, one has

$$\forall r \geq 0 \qquad \mu^n(A_\alpha^r) \geq 1 - Me^{-ar},$$

where A_α^r is the enlargement of A defined by

$$A_\alpha^r = \left\{ x \in \mathcal{X}^n \text{ such that } \inf_{y \in A} \sum_{i=1}^n \alpha(d(x_i, y_i)) \leq r \right\}.$$

Now let us define optimal transportation-costs and transportation-cost inequalities.

DEFINITION 1.2. Let $\nu_1, \nu_2 \in \mathrm{P}(\mathcal{X})$; the optimal transportation-cost $\mathcal{T}_\alpha(\nu_1, \nu_2)$ is defined by

$$\mathcal{T}_\alpha(\nu_1, \nu_2) = \inf_\pi \int \alpha(d(x, y))\, d\pi(x, y),$$

where π describes the set

$$P(\nu_1, \nu_2) = \{\pi \in \mathrm{P}(\mathcal{X} \times \mathcal{X}) \text{ s.t. } \pi(\cdot \times \mathcal{X}) = \nu_1 \text{ and } \pi(\mathcal{X} \times \cdot) = \nu_2\}.$$

One says that μ verifies the transportation-cost inequality $\mathbf{T}_\alpha(a)$ with $a > 0$ if

$$\forall \nu \in \mathrm{P}(\mathcal{X}) \qquad \mathcal{T}_\alpha(\nu, \mu) \leq a\, \mathrm{H}(\nu|\mu), \tag{1.3}$$

where $\mathrm{H}(\nu|\mu)$ is the relative entropy of ν with respect to μ defined by $\mathrm{H}(\nu|\mu) = \int \log(\frac{d\nu}{d\mu})\, d\nu$ if ν is absolutely continuous with respect to μ and $+\infty$ otherwise.

When $\alpha(x) = x^p$, with $p \geq 2$, one will write $\mathbf{C}_p(a)$, A_p^r, $\mathcal{T}_p(\nu, \mu)$ and $\mathbf{T}_p(a)$ instead of $\mathbf{C}_{x^p}(a)$, $A_{x^p}^r$, $\mathcal{T}_{x^p}(\nu, \mu)$ and $\mathbf{T}_{x^p}(a)$.

The idea of controlling an optimal transportation-cost by the relative entropy to obtain concentration first appeared in Marton's works [30, 31]. The inequality $\mathbf{T}_2$

was then introduced by Talagrand in [35], where it was proved to be fulfilled by Gaussian probability measures. In particular, the standard Gaussian measure on $\mathbf{R}$ verifies $\mathbf{T}_2(2)$ (the constant 2 is sharp). In recent years, many efforts have been made to find sufficient conditions for $\mathbf{T}_2$ and other transportation-cost inequalities (see Section 3.3), but the problem of finding a necessary and sufficient condition is still open.

The following theorem is the main result of this paper.

THEOREM 1.3. *Let μ be a probability measure on $\mathcal{X}$ and $a > 0$; μ verifies $\mathbf{T}_\alpha(a)$ if and only if μ verifies $\mathbf{C}_\alpha(b)$ for all $b \in (0, 1/a)$.*

Observe that the relation between the constants is sharp. In the important special case when $\alpha(x) = x^2$, the conclusion of the theorem is that the Gaussian dimension free concentration property (1.1) holds if and only if μ verifies Talagrand's inequality $\mathbf{T}_2(1/a)$. The fact that $\mathbf{T}_\alpha(a)$ implies $\mathbf{C}_\alpha(b)$ for some b is well known and follows from a nice and general argument of Marton. The proof of the converse is surprisingly easy and relies on a very simple Large Deviations argument. We think that this new result confirms the relevance of the Large Deviations point of view for functional inequalities initiated by Léonard and the author in [22] and pursued in [24] by Guillin et al. Moreover, Theorem 1.3 turns out to be a quite powerful tool. For example, the famous result by Otto and Villani stating that the logarithmic-Sobolev inequality (**LSI**) implies the $\mathbf{T}_2$ inequality (see [33, Theorem 1]) is a direct consequence of Theorem 1.3 for $\alpha(x) = x^2$ (see Theorem 4.1 and Corollary 4.2).

The paper is organized as follows. In Section 2 we give a brief account on the Large Deviations phenomenon entering the game. In Section 3 we prove in a general setting that transportation-cost inequalities and dimension free concentration inequalities are equivalent. In Section 4 we give a new proof of Otto and Villani's theorem, in an abstract metric space framework. In Section 5 we prove the equivalence between the Poincaré inequality and dimension free concentration of the exponential type.

2. Some preliminaries on large deviations. In what follows the set of all bounded continuous functions on the Polish space $\mathcal{X}$ is denoted by $C_b(\mathcal{X})$ and $\mathrm{P}(\mathcal{X})$ is equipped with the weak topology, that is, the smallest topology with respect to which all functionals $\nu \in \mathrm{P}(\mathcal{X}) \mapsto \int \varphi\, d\nu$, $\varphi \in C_b(\mathcal{X})$ are continuous.

Let μ be a probability measure on $\mathcal{X}$ and $(X_i)_i$ a sequence of independent and identically distributed random variables with law μ defined on some probability space $(\Omega, \mathbb{P})$. The empirical measure L_n is defined for all positive integer n by

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

where δ_x stands for the Dirac mass at point x.

According to Varadarajan's theorem (see, for instance, [17], Theorem 11.4.1), with probability 1 the sequence $(L_n)_n$ converges to μ in $\mathrm{P}(\mathcal{X})$ for the weak topology on $\mathrm{P}(\mathcal{X})$. In particular, if $O \subset \mathrm{P}(\mathcal{X})$ is some open set not containing μ, then $\mathbb{P}(L_n \in O) \to 0$, as n tends to $+\infty$. The celebrated Sanov's Theorem gives an estimation of the speed of convergence. Roughly speaking, it asserts that $\mathbb{P}(L_n \in O)$ behaves like $e^{-n\,\mathrm{H}(O|\mu)}$, where for all $A \subset \mathrm{P}(\mathcal{X})$, the quantity $\mathrm{H}(A|\mu)$ is defined by

$$\mathrm{H}(A|\mu) = \inf\{\mathrm{H}(\nu|\mu) : \nu \in A\}.$$

More precisely, we have the following:

THEOREM 2.1 (Sanov's theorem). *With the previous notation, for all $A \subset \mathrm{P}(\mathcal{X})$ measurable with respect to the Borel σ-field, one has*

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

where $\mathrm{int}(A)$ *and* $\mathrm{cl}(A)$ *denote respectively the interior and the closure of A.*

A proof of this famous result can be found in, for example, [15], Theorem 6.2.10. As in [22], the use of this Large Deviations theorem will be the key step in the proof of the main Theorem 1.3.

3. Concentration and transportation-cost inequalities.

A remarkable property of transportation-cost inequalities of the form (1.3) is that they tensorize well. More precisely, let us define a family of optimal transportation-costs on $\mathrm{P}(\mathcal{X}^n)$, $n \ge 1$ as follows. If ν_1, ν_2 are two probability measures on $\mathcal{X}^n$, the optimal transportation cost $\mathcal{T}_\alpha^{(n)}(\nu_1, \nu_2)$ is defined by

$$\mathcal{T}_\alpha^{(n)}(\nu_1, \nu_2) = \inf_{\pi \in P(\nu_1,\nu_2)} \int \sum_{i=1}^n \alpha(d(x_i, y_i))\, d\pi(x, y),$$

where $P(\nu_1, \nu_2)$ is the set of all probability measures π on $\mathcal{X}^n \times \mathcal{X}^n$ having ν_1 and ν_2 as marginal distributions. When $n = 1$, one will simply write $\mathcal{T}_\alpha(\nu_1, \nu_2)$ instead of $\mathcal{T}_\alpha^{(1)}(\nu_1, \nu_2)$.

With the notation above, one has the following well-known tensorization result.

PROPOSITION 3.1. *Suppose that μ verifies the inequality $\mathbf{T}_\alpha(a)$ for some $a > 0$, then for all positive integer n, the product measure μ^n verifies*

$$\forall \nu \in \mathrm{P}(\mathcal{X}^n) \qquad \mathcal{T}_\alpha^{(n)}(\nu, \mu^n) \le a\,\mathrm{H}(\nu|\mu^n).$$

Tensorization properties of transportation-cost inequalities were discovered by Marton (see, e.g., [30]). The interested reader can find a general result concerning these tensorization properties in [22], Theorem 5.

3.1. *From transportation-cost to concentration inequalities.* We recall below how dimension free concentration estimates can be deduced from a transportation-cost inequality. The material of this section comes mainly from the works of Marton and Talagrand [30, 31] and [35].

We will need the following lemma.

LEMMA 3.2. *Suppose that μ verifies the inequality $\mathbf{T}_\alpha(a)$ for some $a > 0$. For all $\lambda \in (0,1]$, define $\alpha_\lambda(x) = \alpha(x/\lambda)$, $x \geq 0$. Then for all λ, there is some $c > 0$ such that μ verifies the inequality $\mathbf{T}_{\alpha_\lambda}(c)$. Let c_λ be the optimal constant in the preceding inequality and define $a_\lambda = \max(a, c_\lambda)$. Then $a_\lambda \to a_1 = a$ when $\lambda \to 1$, and for all $n \geq 1$ and all $\lambda \in (0,1)$, μ^n verifies the following inequality*:

$$\begin{aligned} &\forall \nu_1, \nu_2 \in \mathrm{P}(\mathcal{X}^n) \\ &\mathcal{T}_\alpha^{(n)}(\nu_1, \nu_2) \leq \lambda a_\lambda \,\mathrm{H}(\nu_1|\mu^n) + (1-\lambda) a_{(1-\lambda)} \,\mathrm{H}(\nu_2|\mu^n). \end{aligned} \tag{3.1}$$

Now we can state the precise concentration result.

PROPOSITION 3.3. *If μ verifies the inequality $\mathbf{T}_\alpha(a)$ for some $a > 0$, then μ verifies the dimension free concentration inequality $\mathbf{C}_\alpha(b)$ for all $b \in (0, 1/a)$. More precisely, for all positive integer n and all measurable $A \subset \mathcal{X}^n$ with $\mu^n(A) \geq 1/2$, one has*

$$\forall \lambda \in (0,1), \forall r \geq 0 \qquad \mu^n(A_\alpha^r) \geq 1 - M_\lambda e^{-r/(\lambda a_\lambda)}, \tag{3.2}$$

where the numbers a_λ were defined in the preceding lemma and $M_\lambda = 2^{\frac{(1-\lambda)a_{(1-\lambda)}}{\lambda a_\lambda}}$.

In particular, if μ verifies $\mathbf{T}_p(a)$ for some $p \geq 2$, then μ verifies $\mathbf{C}_\alpha(1/a)$. More precisely, for all positive integer n and all measurable $A \subset \mathcal{X}^n$ with $\mu^n(A) \geq 1/2$, one has

$$\forall r \geq r_o := a \log(2) \qquad \mu^n(A_p^r) \geq 1 - e^{-a(r^{1/p} - r_o^{1/p})^p}.$$

Note that this concentration result is dimension free.

PROOF OF LEMMA 3.2. According to the doubling assumption (1.2) made on α, one sees that $K(\lambda) = \sup_{x>0} \alpha(x/\lambda)/\alpha(x) < +\infty$, for all $\lambda \in (0,1]$. Furthermore, it is not difficult to see that $K(\lambda) \to 1$ as $\lambda \to 1$. So, $\mathcal{T}_{\alpha_\lambda}(\nu, \mu) \leq K(\lambda)\mathcal{T}_\alpha(\nu, \mu)$ for all $\lambda \in (0,1]$. From this follows that μ verifies the inequality $\mathbf{T}_{\alpha_\lambda}(K(\lambda)a)$. So, $a \leq a_\lambda \leq K(\lambda)a$, which proves that $a_\lambda \to a$ when $\lambda \to 1$.

Let $\nu_1, \nu_2 \in \mathrm{P}(\mathcal{X}^n)$, then there exist X, Y, Z three random variables with values in $\mathcal{X}^n$ such that X has law ν_1, Y has law μ^n, Z has law ν_2, $\mathcal{T}_{\alpha_\lambda}^{(n)}(\nu_1, \mu^n) = \mathbb{E}[\sum_{i=1}^n \alpha_\lambda(d(X_i, Y_i))]$ and $\mathcal{T}_{\alpha_{1-\lambda}}^{(n)}(\nu_2, \mu^n) = \mathbb{E}[\sum_{i=1}^n \alpha_{1-\lambda}(d(Y_i, Z_i))]$ (see, for instance, the Gluing Lemma of [37], Chapter 1, page 23). Using the convexity of α, one gets

$$\begin{aligned}\mathcal{T}_\alpha^{(n)}(\nu_1,\nu_2) &\le \mathbb{E}\left[\sum_{i=1}^n \alpha(d(X_i,Z_i))\right] \le \mathbb{E}\left[\sum_{i=1}^n \alpha\big(d(X_i,Y_i)+d(Y_i,Z_i)\big)\right]\\ &\le \lambda\mathbb{E}\left[\sum_{i=1}^n \alpha\left(\frac{d(X_i,Y_i)}{\lambda}\right)\right] + (1-\lambda)\mathbb{E}\left[\sum_{i=1}^n \alpha\left(\frac{d(Y_i,Z_i)}{1-\lambda}\right)\right] \quad (3.3)\\ &= \lambda\mathcal{T}_{\alpha_\lambda}^{(n)}(\nu_1,\mu^n) + (1-\lambda)\mathcal{T}_{\alpha_{1-\lambda}}^{(n)}(\nu_2,\mu^n).\end{aligned}$$

According to Proposition 3.1, the inequalities $\mathbf{T}_{\alpha_\lambda}$ and $\mathbf{T}_{\alpha_{1-\lambda}}$ tensorize and so applying (3.3), one sees that μ^n verifies (3.1). □

PROOF OF PROPOSITION 3.3. According to Lemma 3.2, the following inequality holds:

$$\forall \nu_1,\nu_2 \in \mathrm{P}(\mathcal{X}^n) \qquad \mathcal{T}_\alpha^{(n)}(\nu_1,\nu_2) \le \lambda a_\lambda \mathrm{H}(\nu_1|\mu^n) + (1-\lambda)a_{(1-\lambda)} \mathrm{H}(\nu_2|\mu^n).$$

Take $A \subset \mathcal{X}^n$ with $\mu^n(A) \ge \frac{1}{2}$ and define $B = \mathcal{X}^n \setminus A_\alpha^r$, $d\nu_1(x) = \mathbb{1}_B(x)\,d\mu^n(x)/\mu^n(B)$ and $d\nu_2(x) = \mathbb{1}_A(x)\,d\mu^n(x)/\mu^n(A)$. Then $\mathcal{T}_\alpha^{(n)}(\nu_1,\nu_2) \ge r$ and $\mathrm{H}(\nu_1|\mu^n) = -\log\mu^n(B)$ and $\mathrm{H}(\nu_2|\mu^n) = -\log\mu^n(A) \le \log(2)$ and the inequality (3.2) follows immediately. Since $\lambda a_\lambda \to a$ when $\lambda \to 1$, one concludes that μ verifies the concentration property $\mathbf{C}_\alpha(b)$ for all $b \in (0, 1/a)$.

Now suppose that μ verifies $\mathbf{T}_p(a)$. Due to the homogeneity of order p, one can take $a_\lambda = \frac{a}{\lambda^p}$. So, according to what precedes, one has

$$\forall \lambda \in (0,1), \forall r \ge 0 \qquad \mu^n(A_p^r) \ge 1 - \exp\left(\log(2)\left(\frac{\lambda}{1-\lambda}\right)^{p-1} - \frac{r}{a}\lambda^{p-1}\right),$$

for all $A \subset \mathcal{X}^n$ with $\mu^n(A) \ge 1/2$. For a fixed $r \ge r_o = a\log(2)$, the choice $\lambda = 1 - (\frac{a\log(2)}{r})^{1/p}$ maximizes the right-hand side of the above inequality and gives the expected result. □

3.2. *From concentration to transportation-cost inequalities*. Now we are going to establish the converse of Proposition 3.3. The following theorem is the main result of this work.

THEOREM 3.4. *If μ verifies the dimension free concentration inequality $\mathbf{C}_\alpha(b)$ for some $b > 0$, then μ verifies $\mathbf{T}_\alpha(1/b)$.*

PROOF OF THEOREM 1.3. According to Proposition 3.3, the inequality $\mathbf{T}_\alpha(a)$ implies the inequalities $\mathbf{C}_\alpha(b)$ for all $b \in (0, 1/a)$. Conversely, suppose that the inequality $\mathbf{C}_\alpha(b)$ holds for all $b < 1/a$, then Theorem 3.4 implies that μ verifies the inequality $\mathbf{T}_\alpha(1/b)$ for all $b < 1/a$. Letting b go to $1/a$ gives the result. □

PROOF OF THEOREM 3.4. For every positive integer n, and $x \in \mathcal{X}^n$, define $L_n^x = n^{-1} \sum_{i=1}^n \delta_{x_i}$. Recall that for all $u > 0$, the function α_u is defined by $\alpha_u(x) = \alpha(x/u)$.

Let $\lambda \in (0, 1)$; applying inequality (3.3) with $n = 1$ and λd instead of d yields the following Triangle inequality:

$$\forall \nu, \nu' \in \mathrm{P}(\mathcal{X}) \qquad \mathcal{T}_{\alpha_{1/\lambda}}(\nu, \mu) \le \lambda \mathcal{T}_{\alpha}(\nu, \nu') + (1-\lambda)\mathcal{T}_{\alpha_{(1-\lambda)/\lambda}}(\nu', \mu). \tag{3.4}$$

Consider the subset A of $\mathcal{X}^n$ defined by

$$A = \{x \in \mathcal{X}^n \text{ such that } \mathcal{T}_{\alpha_{(1-\lambda)/\lambda}}(L_n^x, \mu) \le m_n\},$$

where m_n is a median of $x \mapsto \mathcal{T}_{\alpha_{(1-\lambda)/\lambda}}(L_n^x, \mu)$ under μ^n [so that $\mu^n(A) \ge 1/2$]. Let us show that $A_\alpha^r \subset \{x \in \mathcal{X}^n$ such that $\mathcal{T}_{\alpha_{1/\lambda}}(L_n^x, \mu) \le \lambda r/n + (1-\lambda) m_n\}$. Namely, if $x \in A_\alpha^r$, then there exists x' such that $\sum_{i=1}^n \alpha(d(x_i, x_i')) \le r$. According to the convexity property of $\mathcal{T}_\alpha$ (see, e.g., [37], Theorem 4.8), one has

$$\mathcal{T}_\alpha(L_n^x, L_n^{x'}) \le \frac{1}{n}\sum_{i=1}^n \mathcal{T}_\alpha(\delta_{x_i}, \delta_{x_i'}) = \frac{1}{n}\sum_{i=1}^n \alpha(d(x_i, x_i'))$$

and so $\mathcal{T}_\alpha(L_n^x, L_n^{x'}) \le \frac{r}{n}$. Now, applying inequality (3.4), with $\nu = L_n^x$ and $\nu' = L_n^{x'}$ gives $\mathcal{T}_{\alpha_{1/\lambda}}(L_n^x, L_n^{x'}) \le \lambda r/n + (1-\lambda) m_n$, which proves the claim.

Define $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, where X_i is a sequence of independent and identically distributed random variables with law μ. Applying the concentration property $\mathbf{C}_\alpha(b)$ to A gives with probabilistic notation

$$\forall r \ge 0 \qquad \mathbb{P}\big(\mathcal{T}_{\alpha_{1/\lambda}}(L_n, \mu) > \lambda r/n + (1-\lambda) m_n\big) \le M e^{-rb}.$$

Equivalently, for all $u \ge (1-\lambda) m_n$,

$$\mathbb{P}\big(\mathcal{T}_{\alpha_{1/\lambda}}(L_n, \mu) > u\big) \le M \exp\left(-\frac{nb}{\lambda}\big(u - (1-\lambda) m_n\big)\right).$$

For all $s \in (0, 1]$, the optimal transportation-cost $\mathcal{T}_{\alpha_s}$ is lower semi-continuous with respect to the weak topology on $\mathrm{P}(\mathcal{X})$ (this fact is classical; it is a consequence of, e.g., [37], Lemma 4.3). Consequently, the set $O_t^s := \{\nu \in \mathrm{P}(\mathcal{X})$ such that $\mathcal{T}_{\alpha_s}(\nu, \mu) > t\}$ is open for all $s \in (0, 1]$ and $t > 0$. Since $\mu \notin O_t^s$, Varadarajan's theorem implies that $\mathbb{P}(L_n \in O_t^s) \to 0$ when $n \to +\infty$. With $s = (1-\lambda)/\lambda$, this fact easily implies that $m_n \to 0$ when $n \to +\infty$. Consequently, taking the "$\limsup_{n\to+\infty} 1/n \log$" in the preceding inequality yields

$$\forall u \ge 0 \qquad \limsup_{n \to +\infty} \frac{1}{n} \log \mathbb{P}\big(\mathcal{T}_{\alpha_{1/\lambda}}(L_n, \mu) > u\big) \le -\frac{ub}{\lambda}. \tag{3.5}$$

On the other hand, since $O_u^{1/\lambda}$ is open, Sanov's theorem implies that

$$-\inf\{\mathrm{H}(\nu|\mu) : \nu \in \mathrm{P}(\mathcal{X}) \text{ such that } \mathcal{T}_{\alpha_{1/\lambda}}(\nu, \mu) > u\}$$
$$\le \liminf_{n \to +\infty} \frac{1}{n} \log \mathbb{P}\big(\mathcal{T}_{\alpha_{1/\lambda}}(\nu, \mu) > u\big).$$

This together with the upper bound (3.5) yields

$$\inf\{\mathrm{H}(\nu|\mu) : \nu \in \mathrm{P}(\mathcal{X}) \text{ such that } \mathcal{T}_{\alpha_{1/\lambda}}(\nu,\mu) > u\} \geq \frac{ub}{\lambda}.$$

In other words,

$$\forall \nu \in \mathrm{P}(\mathcal{X}) \qquad \mathcal{T}_{\alpha_{1/\lambda}}(\nu,\mu) \leq \frac{\lambda}{b}\mathrm{H}(\nu|\mu).$$

The number $\lambda \in (0,1)$ is arbitrary; letting $\lambda \to 1$ gives the result. □

3.3. *Remarks.*

3.3.1. *Sufficient conditions for transportation-cost inequalities.* Necessary and sufficient conditions for transportation-cost inequalities are not known, even in the case of the real line. Nevertheless, several concrete criteria have been discovered recently. Let us recall some of them. In [20], Theorem 5, the author proved the following result:

THEOREM 3.5. *Let μ be a symmetric probability measure on $\mathbf{R}$ of the form $d\mu(x) = e^{-V(x)}dx$, with V a smooth function such that $V''(x)/(V'(x)^2) \to 0$ when $x \to +\infty$. Let $p \geq 1$; if V is such that* $\limsup_{x\to+\infty} x^{p-1}/V'(x) < +\infty$, *then μ verifies the inequality $\mathbf{T}_{\alpha_p}(C)$ for some constant $C > 0$, where $\alpha_p(u) = u^2$ if $|u| \leq 1$ and $\alpha_p(u) = |u|^p$ if $|u| \geq 1$.*

The case $p = 2$ was first established by Cattiaux and Guillin in [12] with a completely different proof. Other cost functions α can be considered in place of the α_p (see [20], Theorem 5). Furthermore, if μ satisfies Cheeger's inequality on $\mathbf{R}$, then a necessary and sufficient condition is known for the transportation-cost inequality associated to α (see [20], Theorem 2).

On $\mathbf{R}^k$, a relatively weak sufficient condition for $\mathbf{T}_2$ (and other transportation-cost inequalities) has been established by the author in [21] (Theorem 4.7 and Corollary 4.12). Define $\omega^{(k)} : \mathbf{R}^k \to \mathbf{R}^k : (x_1, \dots x_k) \mapsto (\omega(x_1), \dots, \omega(x_k))$, where $\omega(u) = \varepsilon(u)\max(|u|, u^2)$ with $\varepsilon(u) = 1$ when u is nonnegative and -1 otherwise. If the image of μ under the map $\omega^{(k)}$ verifies the Poincaré inequality, then μ satisfies $\mathbf{T}_2$. It can be shown that this condition is strictly weaker than the condition μ verifies **LSI** (see [21], Theorem 5.11).

Other sufficient conditions were obtained by Bobkov and Ledoux in [9] with an approach based on the Prekopa–Leindler inequality, or in [13] by Cordero–Erausquin, Gangbo and Houdré with an optimal transportation method.

3.3.2. *Concentration on a fixed space.* Let μ be a probability measure on a Polish space $(\mathcal{X}, d)$. If $A \subset \mathcal{X}$, and $r \geq 0$, let A^r be defined by $A^r = \{x \in \mathcal{X} \text{ s.t. } \inf_{y \in A} d(x, y) \leq r\}$. Let us say that μ verifies the concentration property $\mathbf{c}_2(a)$ for some $a > 0$, if there is some $M > 0$ such that for all measurable $A \subset \mathcal{X}$ with $\mu(A) \geq 1/2$

$$\forall r \geq 0 \qquad \mu(A^r) \geq 1 - M \exp(-ar^2).$$

Of course, the concentration property $\mathbf{c}_2(a)$ is much weaker than $\mathbf{C}_2(a)$. This concentration property can also be characterized in terms of transportation-cost inequalities involving the optimal transportation cost $\mathcal{T}(\nu, \mu) = \inf_{\pi \in P(\nu,\mu)} \int d(x, y)\, d\pi(x, y)$. More precisely, one has the following theorem due to Djellout, Guillin and Wu [16].

THEOREM 3.6. *With the notation above, the three following properties are equivalent*:

(1) *The probability measure* μ *verifies* $\mathbf{c}_2(a)$ *for some* $a > 0$.
(2) *The probability measure* μ *verifies the following transportation-cost inequality*:

$$\forall \nu \in \mathrm{P}(\mathcal{X}) \qquad (\mathcal{T}(\nu, \mu))^2 \leq b\, \mathrm{H}(\nu|\mu)$$

for some $b > 0$.
(3) *There is some* $c > 0$ *such that* $\int e^{c d(x,y)^2}\, d\mu(x)\, d\mu(y) < +\infty$.

This theorem has been generalized to other types of concentration by Bolley and Villani in [11] and by the author in [19]. One can relate the constants a, b, c to each other, but the link is far from being optimal. In particular, the integrability condition (3) behaves very badly with respect to tensorization. In comparison, Theorem 1.3 is a much deeper result.

3.3.3. *The* (τ) *property.* Transportation-cost inequalities are closely related to the so-called (τ) property introduced by Maurey in [32]. If $c(x, y)$ is a nonnegative function defined on some product space $\mathcal{X} \times \mathcal{X}$ and μ is a probability measure on $\mathcal{X}$, one says that (μ, c) has the (τ) property if for all nonnegative f on $\mathcal{X}$,

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

where $Q_c f(x) = \inf_{y \in \mathcal{X}}\{f(y) + c(x, y)\}$. The recent paper by Latała and Wojtaszczyk [26] provides an excellent introduction together with a lot of new results concerning this class of inequalities.

The (τ) property is in fact a sort of dual version of the transportation-cost inequality. This was first observed by Bobkov and Götze in [7]. In the case of $\mathbf{T}_2$, one can show that if μ verifies $\mathbf{T}_2(a)$ on the Euclidean space $(\mathbf{R}^k, |\cdot|_2)$, then $(\mu, (2a)^{-1}|x - y|_2^2)$ has the (τ) property and, conversely, if $(\mu, a^{-1}|x - y|_2^2)$ has the (τ) property, then μ verifies $\mathbf{T}_2(a)$. A general statement can be found in [21, Proposition 4.17].

4. Otto and Villani's theorem revisited. Our aim is now to recover and extend a theorem by Otto and Villani stating that the logarithmic-Sobolev inequality is stronger than Talagrand's $\mathbf{T}_2$ inequality.

Let us recall that a probability measure μ on $\mathcal{X}$ verifies the logarithmic-Sobolev inequality with constant $C > 0$ [$\mathbf{LSI}(C)$ for short] if

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

for all locally Lipschitz f, where the entropy functional is defined by

$$\mathrm{Ent}_\mu(f) = \int f \log f \, d\mu - \int f \, d\mu \log\left(\int f \, d\mu\right), \qquad f \geq 0,$$

and the length of the gradient is defined by

$$|\nabla f|(x) = \limsup_{y \to x} \frac{|f(x) - f(y)|}{d(x, y)} \tag{4.1}$$

[when x is an isolated point, we put $|\nabla f|(x) = 0$].

In [33, Theorem 1], Otto and Villani proved that if a probability measure μ on a Riemannian manifold M satisfies the inequality $\mathbf{LSI}(C)$, then it also satisfies the inequality $\mathbf{T}_2(C)$. Their proof was rather involved and uses partial differential equations, optimal transportation results and fine observations relating relative entropy and Fisher information. A simpler proof, as well as a generalization, was proposed by Bobkov, Gentil and Ledoux in [6]. It makes use of the dual formulation of transportation-cost inequalities discovered by Bobkov and Götze in [7] and relies on hypercontractivity properties of the Hamilton–Jacobi semigroup put in light in the same paper [6]. Otto and Villani's result was successfully generalized by Wang on paths spaces in [38]. More recently, Lott and Villani showed that implication $\mathbf{LSI} \Rightarrow \mathbf{T}_2$ remains true on a length space provided the measure μ satisfies a doubling condition and a local Poincaré inequality (see [29], Theorem 1.8). The Hamilton–Jacobi approach is explained in greater detail in Chapter 22 of Villani's book (see [37], Theorem 22.28). In particular, the interested reader will find there a completely self contained presentation of the properties of the Hamilton–Jacobi semigroup (see [37], Theorem 22.46).

The converse implication $\mathbf{T}_2 \Rightarrow \mathbf{LSI}$ is sometimes true. For example, it is the case when μ is a Log-concave probability measure (see [33], Corollary 3.1). However, in the general case, $\mathbf{T}_2$ and $\mathbf{LSI}$ are not equivalent. In [12], Cattiaux and Guillin give an example of a probability measure verifying $\mathbf{T}_2$ and not $\mathbf{LSI}$.

On "regular" spaces (say, $\mathbf{R}^k$) the implication $\mathbf{LSI} \Rightarrow \mathbf{T}_2$ is completely straightforward. Namely, according to the tensorization property of $\mathbf{LSI}$ and the celebrated Herbst argument, $\mathbf{LSI}(C)$ implies the concentration property $\mathbf{C}_\alpha(1/C)$; since this latter is equivalent to $\mathbf{T}_2(C)$, the proof is completed. When dealing with abstract metric space, a subtle differentiability question arises [see (4.2)]. This is discussed in Theorems 4.1, 4.5 and 4.6.

In the sequel, one will denote $W_2(\nu,\mu)=\sqrt{\mathcal{T}_2(\nu,\mu)}$. This quantity defines a metric on the set $\mathrm{P}_2(\mathcal{X})=\{\nu\in\mathrm{P}(\mathcal{X})$ such that $\int d(x,x_o)^2\,d\nu(x)<+\infty$ for some $x_o\in\mathcal{X}\}$, which is called the (quadratic) Wasserstein metric (see [37], Chapter 6).

THEOREM 4.1. *Let μ be a probability measure on some Polish space $\mathcal{X}$ and suppose that for all positive integer n the function F_n defined on $\mathcal{X}^n$ by $F_n(x)=W_2(L_n^x,\mu)$ verifies*

$$\sum_{i=1}^n |\nabla_i F_n|^2(x)\le 1/n \qquad \text{for } \mu^n \text{ almost every } x\in\mathcal{X}^n, \tag{4.2}$$

where $|\nabla_i F_n(x)|=\limsup_{y_i\to x_i}\frac{|F_n(x_1,\dots,x_{i-1},y_i,x_{i+1},\dots,x_n)-F_n(x)|}{d(y_i,x_i)}$.

If μ verifies the inequality $\mathbf{LSI}(C)$, then μ verifies the inequality $\mathbf{T}_2(C)$.

PROOF. Since μ verifies the $\mathbf{LSI}(C)$ inequality, then according to the additive property of the logarithmic-Sobolev inequality, one can conclude that the product measure μ^n verifies

$$\operatorname{Ent}_{\mu^n}(f^2)\le C\int\sum_{i=1}^n|\nabla_i f|^2(x)\,d\mu^n(x). \tag{4.3}$$

Apply this inequality to $f=e^{sF_n/2}$, with $s\in\mathbf{R}^+$. It is easy to show that $|\nabla_i e^{sF_n/2}|=\frac{s}{2}e^{sF_n/2}|\nabla_i F_n|$, thus, using condition (4.2), one sees that the right-hand side of (4.3) is less than $C\frac{s^2}{4n}\int e^{sF_n}\,d\mu^n$. Letting $Z(s)=\int e^{sF_n}\,d\mu^n$, one gets the differential inequality:

$$\frac{Z'(s)}{sZ(s)}-\frac{\log Z(s)}{s^2}\le\frac{C}{4n}.$$

Integrating this yields

$$\forall s\in\mathbf{R}^+\qquad Z(s)=\int e^{sF_n}\,d\mu^n\le e^{s\int F_n\,d\mu^n+Cs^2/(4n)}.$$

This implies that

$$\mathbb{P}\big(W_2(L_n,\mu)\ge t+\mathbb{E}[W_2(L_n,\mu)]\big)\le e^{-nt^2/C}.$$

Let $Y_n=W_2(L_n,\mu)$; let us show that $\mathbb{E}[Y_n]\to 0$ as $n\to+\infty$. Arguing as in the proof of Theorem 3.4, one sees that $\mathbb{P}(Y_n>t)\to 0$ as $n\to+\infty$ for all $t>0$. It is easy to show that $K:=\sup_{n\ge 1}\mathbb{E}[Y_n^2]<+\infty$. For all $\varepsilon>0$, one has $\mathbb{E}[Y_n]\le\varepsilon+\mathbb{E}[Y_n\mathbb{1}_{Y_n>\varepsilon}]\le\varepsilon+K^{1/2}\mathbb{P}(Y_n>\varepsilon)^{1/2}$ and so $\limsup_{n\to+\infty}\mathbb{E}[Y_n]\le\varepsilon$, and since $\varepsilon>0$ is arbitrary, this proves the claim.

As a consequence, for all $u>0$,

$$\limsup_{n\to+\infty}\frac{1}{n}\log\mathbb{P}\big(\mathcal{T}_2(L_n,\mu)\ge u\big)\le -u/C.$$

Applying Sanov's theorem as in proof of Theorem 3.4, one concludes that the inequality $\mathbf{T}_2(C)$ holds. □

COROLLARY 4.2. *Suppose that $\mathcal{X}$ is the Euclidean space $\mathbf{R}^k$ and μ is absolutely continuous with respect to the Lebesgue measure; then condition (4.2) holds true and so*

$$(\mu \textit{ verifies } \mathbf{LSI}(C)) \Rightarrow (\mu \textit{ verifies } \mathbf{T}_2(C)).$$

PROOF. The map $x \mapsto W_2(L_n^x, \mu)$ is $1/\sqrt{n}$-Lipschitz for the Euclidean distance. Indeed, if $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ are in $(\mathbf{R}^k)^n$, then, thanks to the triangle inequality,

$$\left|W_2(L_n^x, \mu) - W_2(L_n^y, \mu)\right| \le W_2(L_n^x, L_n^y).$$

According to the convexity property of $\mathcal{T}_2(\cdot, \cdot)$ (see, e.g., [37], Theorem 4.8), one has

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

which proves the claim. According to Rademacher's Theorem, F_n is almost everywhere differentiable on $(\mathbf{R}^k)^n$ with respect to the Lebesgue measure. It is then easy to show that condition (4.2) is fulfilled when μ is absolutely continuous with respect to the Lebesgue measure. □

REMARK 4.3. The same result holds true when $\mathcal{X}$ is a Riemannian manifold.

The following proposition shows that a weaker form of condition (4.2) is always true. If f is a locally Lipschitz map on $(\mathcal{X}, d)$, one defines its subgradient norm $|\nabla^- f|$ by

$$|\nabla^- f|(x) = \limsup_{y \to x} \frac{[f(y) - f(x)]_+}{d(x, y)},$$

with $[a]_+ = \max(a, 0)$. Observe that $|\nabla^- f| \le |\nabla f|$.

PROPOSITION 4.4. *If $(\mathcal{X}, d)$ is a Polish space, then the following inequality holds*:

$$\forall x \in \mathcal{X}^n \qquad \sum_{i=1}^n |\nabla_i^- F_n|^2(x) \le 1/n. \tag{4.4}$$

The following proof uses an argument which I learned from Paul–Marie Samson.

PROOF OF PROPOSITION 4.4. Let us show how to compute $|\nabla_1^- F_n|$. Let $a = (a_1, \ldots, a_n) \in \mathcal{X}^n$ and $z \in \mathcal{X} \setminus \{a_1, \ldots, a_n\}$ and define $za = (z, a_2, \ldots, a_n)$, then

$$|\nabla_1^- F_n|(a) = \frac{1}{2F_n(a)} \limsup_{z \to a_1} \frac{[\mathcal{T}_2(L_n^{za}, \mu) - \mathcal{T}_2(L_n^a, \mu)]_+}{d(z, a_1)}.$$

Let $\pi \in P(L_n^a, \mu)$ be an optimal coupling (see [37], Theorem 4.1, for the existence); it is not difficult to see that one can write $\pi(dx, dy) = p(x, dy) L_n^a(dx)$, with $p(x, dy) = \frac{\pi(\{x\} \times dy)}{L_n^a(x)}$, for all x belonging to the support of L_n. Put $\nu_i(dy) = p(a_i, dy)$, for all $1 \le i \le n$. The probability measures ν_i are such that $n^{-1} \sum_i \nu_i = \mu$. Define $\tilde{p}$ as follows: $\tilde{p}(z, dy) = \nu_1$ and $\tilde{p}(a_i, dy) = \nu_i$ for all $i \ge 2$. Then $\tilde{\pi} = \tilde{p}(x, dy) L_n^{za}(dy)$ belongs to $P(L_n^{za}, \mu)$ (but is not necessary optimal). One has

$$\begin{aligned}
\mathcal{T}_2(L_n^{za}, \mu) - \mathcal{T}_2(L_n^a, \mu) &\le \int d(x, y)^2 \, d\tilde{\pi}(x, y) - \int d(x, y)^2 \, d\pi(x, y) \\
&= \frac{1}{n} \sum_{i=1}^n \int d((za)_i, y)^2 \, d\nu_i(y) - \frac{1}{n} \sum_{i=1}^n \int d(a_i, y)^2 \, d\nu_i(y) \\
&= \frac{1}{n} \int d(z, y)^2 - d(a_1, y)^2 \, d\nu_1(y) \\
&\le \frac{1}{n} d(z, a_1) \int d(z, y) + d(a_1, y) \, d\nu_1(y).
\end{aligned}$$

Since the function $x \mapsto [x]_+$ is nondecreasing, one has

$$\frac{[\mathcal{T}_2(L_n^{za}, \mu) - \mathcal{T}_2(L_n^a, \mu)]_+}{d(z, a_1)} \le \frac{1}{n} \int d(z, y) + d(a_1, y) \, d\nu_1(y).$$

Letting $z \to a_1$ yields $|\nabla_1^- F_n(a)|^2 \le \frac{\int d(a_1, y)^2 \, d\nu_1(y)}{n^2 \mathcal{T}_2(L_n^a, \mu)}$. Doing the same computations for the other derivatives (with the same optimal coupling π), one gets $|\nabla_i^- F_n(a)|^2 \le \frac{\int d(a_i, y)^2 \, d\nu_i(y)}{n^2 \mathcal{T}_2(L_n^a, \mu)}$. Summing these inequalities gives $\sum_i |\nabla_i^- F_n|^2(a) \le 1/n$ for all $a \in \mathcal{X}^n$, which achieves the proof. □

With this proposition in hand, we can recover and extend a recent result of Lott and Villani. Following [29], one says that a probability measure μ on $\mathcal{X}$ verifies the inequality $\mathbf{LSI}^+(C)$ if

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

holds true for all locally Lipschitz f, where the subgradient norm $|\nabla^- f|$ was defined above. Since $|\nabla^- f| \le |\nabla f|$, the inequality $\mathbf{LSI}^+$ is stronger than $\mathbf{LSI}$; more precisely, $\mathbf{LSI}^+(C) \Rightarrow \mathbf{LSI}(C)$.

THEOREM 4.5. *Let μ be a probability measure on some Polish space $\mathcal{X}$; if μ verifies the inequality* $\mathbf{LSI}^+(C)$, *then μ verifies* $\mathbf{T}_2(C)$.

This result was first obtained by Lott and Villani using the Hamilton–Jacobi method. This approach forced them to make many assumptions on $\mathcal{X}$ and μ. In particular, in [29], Theorem 1.8, $\mathcal{X}$ was supposed to be a compact length space and a doubling condition was imposed on μ. The result above shows that the implication $\mathbf{LSI}^+ \Rightarrow \mathbf{T}_2$ is in fact always true.

PROOF. The inequality $\mathbf{LSI}^+$ tensorizes, so μ^n verifies

$$\operatorname{Ent}_{\mu^n}(f^2) \le C \int \sum_{i=1}^n |\nabla_i^- f|^2 \, d\mu^n.$$

Take $f = e^{sF_n/2}$, $s \in \mathbf{R}^+$ with $F_n(x) = W_2(L_n^x, \mu)$. It is easy to check that $|\nabla_i^- e^{sF_n/2}| = \frac{s}{2} e^{sF_n/2} |\nabla_i^- F_n|$ (note that the function $x \mapsto e^{sx}$ is nondecreasing). According to (4.4), $\sum_i |\nabla_i^- F_n|^2 \le 1/n$ and so letting $Z(s) = \int e^{sF_n} \, d\mu^n$, one has $\frac{Z'(s)}{sZ(s)} - \frac{\log Z(s)}{s^2} \le \frac{C}{4n}$, for all $s \ge 0$. One concludes as in the proof of Theorem 4.1. □

Our next result shows that condition (4.2) holds when the Monge–Kantorovich problem of transporting μ on a probability measure with finite support admits a unique deterministic solution. To state this result, let us recall some definitions.

Let $\nu, \mu \in \mathrm{P}_2(\mathcal{X})$; recall that $P(\nu, \mu) = \{\pi \in \mathrm{P}(\mathcal{X}^2) \text{ s.t. } \pi(\cdot \times \mathcal{X}) = \nu$ and $\pi(\mathcal{X} \times \cdot) = \mu\}$. One says that $\pi^* \in P(\nu, \mu)$ is an optimal coupling of (ν, μ) if $\mathcal{T}_2(\nu, \mu) = \int d^2(x, y) \, d\pi^*(x, y)$. If $F : \mathcal{X} \to \mathcal{Y}$ is a measurable map, one denotes by $F\sharp\mu$ the image of μ under F. By definition, $F\sharp\mu(A) = \mu(F^{-1}(A))$, for all measurable $A \subset \mathcal{Y}$. A coupling $\pi \in P(\nu, \mu)$ is said to be deterministic if there is a measurable map $H : \mathcal{X} \to \mathcal{X}$ such that $\pi = F\sharp\mu$, with $F(x) = (H(x), x)$, $x \in \mathcal{X}$.

One will say that a probability measure μ on $\mathcal{X}$ is well transportable on finite probability measures if for all probability measure ν on $\mathcal{X}$ with finite support, there is a unique optimal coupling between ν and μ and this coupling is deterministic.

THEOREM 4.6. *If μ is well transportable on finite probability measures then condition* (4.2) *holds true and so*

$$(\mu \text{ verifies } \mathbf{LSI}(C)) \Rightarrow (\mu \text{ verifies } \mathbf{T}_2(C)).$$

The following result is due to Cuesta–Albertos and Tuero–Díaz (see [14], Theorem 3):

THEOREM 4.7 (Cuesta–Albertos and Tuero–Díaz). *If a probability measure μ verifies the following continuity condition*:

$$\forall k \in \mathbf{R}, \forall u \ne v \in \mathcal{X} \qquad \mu\{x \in \mathcal{X} \text{ s.t. } d^2(x, u) - d^2(x, v) = k\} = 0, \tag{4.5}$$

then it is well transportable on finite probability measures.

REMARK 4.8. If μ is a probability measure on the Euclidean space $\mathbf{R}^k$, the condition (4.5) amounts to say that μ does not charge hyperplanes.

PROOF. For all $a = (a_1, \dots, a_n) \in \mathcal{X}^n$, let π_a^* be the unique optimal coupling between L_n^a and μ. By assumption, it is deterministic so there is a map $H_a : \mathcal{X} \to \mathcal{X}$ such that π_a^* is the image of μ under the map $x \mapsto (H_a(x), x)$. Define $E_a^i = \{x \in \mathcal{X}$ s.t. $H_a(x) = a_i\}$.

If μ is well transportable on finite probability measures then $\mu\{x\} = 0$ for all $x \in \mathcal{X}$. We leave the verification of this fact to the reader. As a consequence, $\mu^n(\forall i \neq j, x_i \neq x_j) = 1$ and so it is enough to verify the condition $\sum_i |\nabla_i F_n|^2(a) \leq 1/n$ in the particular case where $a = (a_1, \dots, a_n)$ with $a_i \neq a_j$ for all $i \neq j$.

Let us show how to compute $|\nabla_1 F_n(a)|$. Let $a = (a_1, \dots, a_n) \in \mathcal{X}^n$ with $a_i \neq a_j$ and $z \in \mathcal{X} \setminus \{a_1, \dots, a_n\}$ and define $za = (z, a_2, \dots, a_n)$, then

$$|\nabla_1 F_n|(a) = \frac{1}{2F_n(a)} \limsup_{z \to a_1} \frac{|\mathcal{T}_2(L_n^{za}, \mu) - \mathcal{T}_2(L_n^a, \mu)|}{d(z, a_1)}.$$

Repeating the proof of Proposition 4.4, we see that

$$\frac{\mathcal{T}_2(L_n^{za}, \mu) - \mathcal{T}_2(L_n^a, \mu)}{d(z, a_1)} \leq \int \bigl(d(z, y) + d(a_1, y)\bigr)\mathbb{1}_{E_a^1}(y)\, d\mu(y) := A(z).$$

Exchanging the roles a and za, we obtain

$$\frac{\mathcal{T}_2(L_n^a, \mu) - \mathcal{T}_2(L_n^{za}, \mu)}{d(z, a_1)} \leq \int \bigl(d(z, y) + d(a_1, y)\bigr)\mathbb{1}_{E_{za}^1}(y)\, d\mu(y) := B(z).$$

So

$$\frac{|\mathcal{T}_2(L_n^{za}, \mu) - \mathcal{T}_2(L_n^a, \mu)|}{d(z, a_1)} \leq \max(A(z), B(z)).$$

It is clear that $A(z) \to 2\int d(a_1, y)\mathbb{1}_{E_a^1(y)}\, d\mu(y)$, when $z \to a_1$. Let us see that $B(z)$ goes to the same quantity. According to [37], Corollary 5.23, H_{za} converges to H_a in probability when z goes to a_1. More precisely, for all $\varepsilon > 0$, $\mu(x : d(H_{za}(x), H_a(x)) \geq \varepsilon) \to 0$ when $z \to a_1$. In particular, $\mu(E_{za}^1 \setminus E_a^1) \to 0$ and $\mu(E_a^1 \setminus E_{za}^1) \to 0$. From this follows easily that $|A(z) - B(z)| \to 0$ when $z \to a_1$, and so $B(z) \to 2\int d(a_1, y)\mathbb{1}_{E_a^1(y)}\, d\mu(y)$, when $z \to a_1$.

According to what precedes,

$$|\nabla_1 F_n(a)|^2(a) \leq \frac{(\int d(a_1, y)\mathbb{1}_{E_a^1}(y)\, d\mu(y))^2}{\mathcal{T}_2(L_n^a, \mu)} \leq \frac{\int d(a_1, y)^2 \mathbb{1}_{E_a^1}(y)\, d\mu(y)}{n\mathcal{T}_2(L_n^a, \mu)},$$

since $\mu(E_a^1) = 1/n$. Similar inequalities hold for the other derivatives; summing these inequalities gives the desired result since $\mathcal{T}_2(L_n^a, \mu) = \sum_i \int d(a_i, y)^2 \times \mathbb{1}_{E_a^i}(y)\, d\mu(y)$. □

5. Poincaré inequality and exponential concentration. In this section $\mathcal{X} = \mathbf{R}^k$ and $d(x, y) = |x - y|_2$ is the usual Euclidean distance.

Let us recall that a probability measure μ on $\mathcal{X}$ satisfies the Poincaré inequality with constant $a > 0$ if

$$\mathrm{Var}_\mu(f) \leq a \int |\nabla f|_2^2 \, d\mu$$

for all locally Lipschitz f.

The following theorem proves the equivalence between the Poincaré inequality, the dimension free exponential concentration and the corresponding transportation-cost inequality.

THEOREM 5.1. *Let μ be a probability measure on $\mathbf{R}^k$. Define $\alpha_1(x) = \min(x^2, 2x - 1)$; the following propositions are equivalent*:

(1) *The probability measure μ verifies the Poincaré inequality with a constant $a > 0$.*
(2) *The probability measure μ verifies the concentration property $\mathbf{C}_{\alpha_1}(b)$ for some $b > 0$.*
(3) *The probability measure μ verifies the inequality $\mathbf{T}_{\alpha_1}(c)$ for some $c > 0$.*

More precisely:

- (1) *implies* (2) *with $b = \kappa \max(a, \sqrt{a})^{-1}$, κ being a universal constant.*
- (2) *implies* (3) *with $c = 1/b$.*
- (3) *implies* (1) *with $a = c/2$.*

The equivalence of (1) and (3) was first obtained by Bobkov, Gentil and Ledoux in [6], Corollary 5.1, with the Hamilton–Jacobi approach (see also [37], Theorem 22.25) and the fact that (1) implies (2) is the main result of [8]. The equivalence of (1) and (2) [or (2) and (3)] seems to be new.

Before stating the proof, let us mention an interesting open question related to the Poincaré inequality and exponential concentration. Since the work by Gromov and Milman (see [23]), it is well known that under the Poincaré inequality, the following dimension free concentration inequality holds:

$$\begin{gathered} \forall n \geq 1, \forall A \subset (\mathbf{R}^k)^n \text{ with } \mu^n(A) \geq 1/2, \forall r \geq 0 \\ \mu^n(A^r) \geq 1 - Me^{-ar}, \end{gathered} \tag{5.1}$$

where M and a are positive constants independent of n and where the enlargement is performed with respect to the *Euclidean metric* d_2 on $(\mathbf{R}^k)^n$. Note that (5.1) is weaker than the concentration property $\mathbf{C}_{\alpha_1}$. We do not know if (5.1) is equivalent to the Poincaré inequality or if it is related to some transportation-cost inequality.

PROOF. According to (a careful reading of) [8, Corollary 3.2], (1) implies (2) with $b = \kappa \max(a, \sqrt{a})^{-1}$, where κ is a universal constant. According to Theorem 3.4, (2) implies (3) (with $c = 1/b$). It remains to prove that (3) implies (1). This last point is classical; let us simply sketch the proof. The transportation-cost inequality is equivalent to the following property: for all bounded f on $\mathbf{R}^k$,

$$\int e^{Qf}\,d\mu \le e^{\int f\,d\mu},$$

where $Qf(x) = \inf_{y\in\mathbf{R}^k}\{f(y) + c^{-1}\alpha_1(|x-y|_2)\}$ (see, e.g., [22], Corollary 1). Let f be a smooth function and apply the preceding inequality to tf. When t goes to 0, it can be shown that

$$Q(tf)(x) - tf(x) = -\frac{ct^2}{4}|\nabla f|_2^2(x) + o(t^2),$$

so $\int e^{Q(tf)}\,d\mu = 1 + t\int f\,d\mu + \frac{t^2}{2}\int f^2\,d\mu - \frac{ct^2}{4}\int |\nabla f|_2^2\,d\mu + o(t^2)$. On the other hand, $e^{t\int f\,d\mu} = 1 + t\int f\,d\mu + \frac{t^2}{2}(\int f\,d\mu)^2$. One concludes that $\mathrm{Var}_\mu(f) \le \frac{c}{2}\int |\nabla f|^2\,d\mu$, which achieves the proof. □

Acknowledgments. I want to warmly acknowledge Franck Barthe, Patrick Cattiaux, Arnaud Guillin, Michel Ledoux, Paul–Marie Samson and Cédric Villani for their valuable comments about this work.

REFERENCES

[1] ANÉ, C., BLACHÈRE, S., CHAFAÏ, D., FOUGÈRES, P., GENTIL, I., MALRIEU, F., ROBERTO, C. and SCHEFFER, G. (2000). *Sur les Inégalités de Sobolev Logarithmiques*. *Panoramas et Synthèses* [*Panoramas and Syntheses*] **10**. Société Mathématique de France, Paris.

[2] BARTHE, F., CATTIAUX, P. and ROBERTO, C. (2006). Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry. *Rev. Mat. Iberoam.* **22** 993–1067. MR2320410

[3] BARTHE, F. and ROBERTO, C. (2003). Sobolev inequalities for probability measures on the real line. *Studia Math.* **159** 481–497. MR2052235

[4] BARTHE, F. and ROBERTO, C. (2008). Modified logarithmic Sobolev inequalities on $\mathbb{R}$. *Potential Anal.* **29** 167–193. MR2430612

[5] BECKNER, W. (1989). A generalized Poincaré inequality for Gaussian measures. *Proc. Amer. Math. Soc.* **105** 397–400. MR954373

[6] BOBKOV, S. G., GENTIL, I. and LEDOUX, M. (2001). Hypercontractivity of Hamilton–Jacobi equations. *J. Math. Pures Appl.* **80** 669–696. MR1846020

[7] BOBKOV, S. G. and GÖTZE, F. (1999). Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** 1–28. MR1682772

[8] BOBKOV, S. G. and LEDOUX, M. (1997). Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields* **107** 383–400. MR1440138

[9] BOBKOV, S. G. and LEDOUX, M. (2000). From Brunn–Minkowski to Brascamp–Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.* **10** 1028–1052. MR1800062

[10] BOBKOV, S. G. and ZEGARLINSKI, B. (2005). Entropy bounds and isoperimetry. *Mem. Amer. Math. Soc.* **176** 829. MR2146071

[11] BOLLEY, F. and VILLANI, C. (2005). Weighted Csiszár–Kullback–Pinsker inequalities and applications to transportation inequalities. *Ann. Fac. Sci. Toulouse Math.* (6) **14** 331–352. MR2172583

[12] CATTIAUX, P. and GUILLIN, A. (2006). On quadratic transportation cost inequalities. *J. Math. Pures Appl.* **86** 341–361. MR2257848

[13] CORDERO-ERAUSQUIN, D., GANGBO, W. and HOUDRÉ, C. (2004). Inequalities for generalized entropy and optimal transportation. In *Recent Advances in the Theory and Applications of Mass Transport. Contemp. Math.* **353** 73–94. Amer. Math. Soc., Providence, RI. MR2079071

[14] CUESTA-ALBERTOS, J. A. and TUERO-DÍAZ, A. (1993). A characterization for the solution of the Monge–Kantorovich mass transference problem. *Statist. Probab. Lett.* **16** 147–152. MR1212046

[15] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. Springer, New York. MR1619036

[16] DJELLOUT, H., GUILLIN, A. and WU, L. (2004). Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.* **32** 2702–2732. MR2078555

[17] DUDLEY, R. M. (1989). *Real Analysis and Probability*. Wadsworth & Brooks/Cole, Pacific Grove, CA. MR982264

[18] GENTIL, I., GUILLIN, A. and MICLO, L. (2005). Modified logarithmic Sobolev inequalities and transportation inequalities. *Probab. Theory Related Fields* **133** 409–436. MR2198019

[19] GOZLAN, N. (2006). Integral criteria for transportation-cost inequalities. *Electron. Comm. Probab.* **11** 64–77 (electronic). MR2231734

[20] GOZLAN, N. (2007). Characterization of Talagrand's like transportation-cost inequalities on the real line. *J. Funct. Anal.* **250** 400–425. MR2352486

[21] GOZLAN, N. (2008). Poincaré inequalities and dimension free concentration of measure. Preprint.

[22] GOZLAN, N. and LÉONARD, C. (2007). A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields* **139** 235–283. MR2322697

[23] GROMOV, M. and MILMAN, V. D. (1983). A topological application of the isoperimetric inequality. *Amer. J. Math.* **105** 843–854. MR708367

[24] GUILLIN, A., LÉONARD, C., WU, L. M. and YIAO, N. (2009). Transportation-information inequalities for Markov processes. *Probab. Theory Related Fields* **144** 669–695. MR2496446

[25] LATAŁA, R. and OLESZKIEWICZ, K. (2000). Between Sobolev and Poincaré. In *Geometric Aspects of Functional Analysis. Lecture Notes in Math.* **1745** 147–168. Springer, Berlin. MR1796718

[26] LATAŁA, R. and WOJTASZCZYK, J. O. (2008). On the infimum convolution inequality. *Studia Math.* **189** 147–187. MR2449135

[27] LEDOUX, M. (1995/97). On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.* **1** 63–87 (electronic). MR1399224

[28] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347

[29] LOTT, J. and VILLANI, C. (2007). Hamilton–Jacobi semigroup on length spaces and applications. *J. Math. Pures Appl.* **88** 219–229. MR2355455

[30] MARTON, K. (1986). A simple proof of the blowing-up lemma. *IEEE Trans. Inform. Theory* **32** 445–446. MR838213

[31] MARTON, K. (1996). Bounding $\overline{d}$-distance by informational divergence: A method to prove measure concentration. *Ann. Probab.* **24** 857–866. MR1404531

[32] MAUREY, B. (1991). Some deviation inequalities. *Geom. Funct. Anal.* **1** 188–197. MR1097258

[33] OTTO, F. and VILLANI, C. (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** 361–400. MR1760620

[34] SAMSON, P.-M. (2000). Concentration of measure inequalities for Markov chains and Φ-mixing processes. *Ann. Probab.* **28** 416–461. MR1756011

[35] TALAGRAND, M. (1996). Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.* **6** 587–600. MR1392331

[36] VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483

[37] VILLANI, C. (2009). *Optimal Transport*. Springer, Berlin. MR2459454

[38] WANG, F.-Y. (2004). Probability distance inequalities on Riemannian manifolds and path spaces. *J. Funct. Anal.* **206** 167–190. MR2024350

LABORATOIRE D'ANALYSE ET
DE MATHÉMATIQUES APPLIQUÉES (UMR CNRS 8050)
UNIVERSITÉ PARIS-EST
5 BD DESCARTES
77454 MARNE LA VALLÉE CEDEX 2
FRANCE
E-MAIL: nathael.gozlan@univ-mlv.fr