Instructional Conference on
Combinatorial Aspects of Mathematical Analysis
Edinburgh, 25 March-5 April 2002

and

Summer School on
Singular Phenomena and Scaling in Mathematical Models
Bonn, 10-13 June, 2003

MEASURE CONCENTRATION, TRANSPORTATION COST, AND FUNCTIONAL INEQUALITIES

M. Ledoux
University of Toulouse, France

Abstract. — *In these lectures, we present a triple description of the concentration of measure phenomenon, geometric (through Brunn-Minkoswki inequalities), measure-theoretic (through transportation cost inequalities) and functional (through logarithmic Sobolev inequalities), and investigate the relationships between these various viewpoints. Special emphasis is put on optimal mass transportation and the dual hypercontractive bounds on solutions of Hamilton-Jacobi equations that offer a unified treatment of these various aspects.*

These notes survey recent developments around the concentration of measure phenomenon through various descriptions, geometric, measure theoretic and functional. These descriptions aim to analyze measure concentration for both product and (strictly) log-concave measures, with a special emphasis on dimension free bounds. Inequalities independent of the number of variables are indeed a key information in the study of a number of models in probability theory and statistical mechanics, with a view towards infinite dimensional analysis. To this task, we review the geometric tool of Brunn-Minkowski inequalities, transportation cost inequalities, and functional logarithmic Sobolev inequalities and semigroup methods. Connections are developed on the basis of optimal mass transportation and dual hypercontractive bounds on solutions of Hamilton-Jacobi equations, providing a synthetic view of these recent developments. Results and methods are only outlined in the simplest and basic setting. References to recent PDE extensions are briefly discussed in the last part of the notes. These notes only collect a few basic results on the topics of these lectures, and only aim to give a flavour of the subject. We refer to [Le], [B-G-L], [O-V1], [CE], [CE-G-H], [Vi]... for further material, proofs and detailed references.

1. The concentration of measure phenomenon

The concentration of measure phenomenon was put forward in the early seventies by V. Milman [Mi1], [Mi3] in the asymptotic geometry of Banach spaces and the proof of the famous Dvoretzky theorem on spherical sections of convex bodies. Of isoperimetric inspiration, it is of powerful interest in applications, in various areas such as geometry, functional analysis and infinite dimensional integration, discrete mathematics and complexity theory, and probability theory. General references, from various viewpoints, are [Bal], [Grom], [Le], [MD], [Mi2], [Mi4], [M-S], [Sc], [St], [Ta1]...

1.1 Introduction

To introduce to the concept of measure concentration, we first briefly discuss a few examples.

A first illustration is suggested by the example of the standard n-sphere $\mathbb{S}^n$ in $\mathbb{R}^{n+1}$ when dimension n is large. By a standard computation, uniform measure σ^n on $\mathbb{S}^n$ is almost concentrated when the dimension n is large around the (every) equator. Actually, the isoperimetric inequality on $\mathbb{S}^n$ expresses that spherical caps (geodesic balls) minimize the boundary measure at fixed volume. In its integrated form, given a Borel set A on $\mathbb{S}^n$ with the same measure as a spherical cap B, then for every $r > 0$,

$$\sigma^n(A_r) \geq \sigma^n(B_r)$$

where $A_r = \{x \in \mathbb{S}^n; d(x, A) < r\}$ is the (open) neighborhood of order r for the geodesic distance on $\mathbb{S}^n$. One main feature of concentration with respect to isoperimetry is to analyze this inequality for the non-infinitesimal values of $r > 0$. The explicit evaluation of the measure of spherical caps then implies that given any measurable set A with, say, $\sigma^n(A) \geq \frac{1}{2}$, for every $r > 0$,

$$\sigma^n(A_r) \geq 1 - \mathrm{e}^{-(n-1)r^2/2}. \tag{1.1}$$

Therefore, almost all points on $\mathbb{S}^n$ are within (geodesic) distance $\frac{1}{\sqrt{n}}$ from A which is of particular interest when the dimension n is large. From a "tomographic" point of view, the visual diameter of $\mathbb{S}^n$ (for σ^n) is of the order of $\frac{1}{\sqrt{n}}$ as $n \to \infty$ which is in contrast with the diameter of $\mathbb{S}^n$ as metric space.

This example is a first, and main, instance of the concentration of measure phenomenon for which nice patterns develop as the dimension is large. It furthermore suggests the introduction of a concentration function in order to evaluate the decay in (1.1). Setting

$$\alpha_{\sigma^n}(r) = \sup\big\{1 - \sigma^n(A_r); A \subset \mathbb{S}^n, \sigma^n(A) \geq \tfrac{1}{2}\big\}, \quad r > 0,$$

the bound (1.1) amounts to say that

$$\alpha_{\sigma^n}(r) \leq \mathrm{e}^{-(n-1)r^2/2}, \quad r > 0. \tag{1.2}$$

Note that $r > 0$ in (1.2) actually ranges up to the diameter π of $\mathbb{S}^n$ and that (1.2) is thus mainly of interest when n is large.

By rescaling of the metric, the preceding results apply similarly to uniform measure σ_R^n on the n-sphere $\mathbb{S}_R^n$ of radius $R > 0$. In particular,

$$\alpha_{\sigma_R^n}(r) \leq \mathrm{e}^{-(n-1)r^2/2R^2}, \quad r > 0. \tag{1.3}$$

Properly normalized, uniform measures on high dimensional spheres approximate Gaussian distributions. More precisely, the measures $\sigma^n_{\sqrt{n}}$ converge when n tends to infinity to the canonical Gaussian measure on $\mathbb{R}^{\mathbb{N}}$. The isoperimetric inequality on spheres may then be transferred to an isoperimetric inequality for Gaussian measures. Precisely, if $\gamma = \gamma^k$ is the canonical Gaussian measure on $\mathbb{R}^k$ with density $(2\pi)^{-k/2}\mathrm{e}^{-|x|^2/2}$ with respect to Lebesgue measure, and if A is a Borel set in $\mathbb{R}^k$ with $\gamma(A) = \Phi(a)$ for some $a \in [-\infty, +\infty]$ where $\Phi(t) = (2\pi)^{-1/2}\int_{-\infty}^t \mathrm{e}^{-x^2/2}dx$ is the distribution function of the standard normal distribution on the line, then for every $r > 0$,

$$\gamma(A_r) \geq \Phi(a + r).$$

Here A_r denotes the r-neigborhood of A with respect to the standard Euclidean metric on $\mathbb{R}^k$. Throughout these notes, $\mathbb{R}^k$ (or subsets of $\mathbb{R}^k$) will be equipped here with the standard Euclidean structure and the metric $|x - y|$, $x, y \in \mathbb{R}^k$, induced by the norm $|x| = \big(\sum_{i=1}^k x_i^2\big)^{1/2}$, $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$. The scalar product will be denoted $x \cdot y = \sum_{i=1}^k x_i y_i$, $x = (x_1, \ldots, x_k)$, $y = (y_1, \ldots, y_k) \in \mathbb{R}^k$. Defining similarly the concentration function for γ as

$$\alpha_\gamma(r) = \sup\big\{1 - \gamma(A_r); A \subset \mathbb{R}^k, \gamma(A) \geq \tfrac{1}{2}\big\}$$

we get in particular since $\Phi(0) = \frac{1}{2}$ and $1 - \Phi(r) \leq \mathrm{e}^{-r^2/2}$, $r > 0$, that

$$\alpha_\gamma(r) \leq \mathrm{e}^{-r^2/2}, \quad r > 0. \tag{1.4}$$

One may also think of (1.3) in the limit as $n \to \infty$ with $R = \sqrt{n}$. One may again interpret (1.4) by saying that given a set A with $\gamma(A) \geq \frac{1}{2}$, almost all points in $\mathbb{R}^k$ are within distance 5 or 10 say from the set A whereas of course $\mathbb{R}^k$ is unbounded. We have thus here a second instance of measure concentration with the particular feature that the concentration function of (1.4) does not depend on the dimension of the underlying state space $\mathbb{R}^k$ for the product measure $\gamma = \gamma^k$. This example will be the reference one in these lectures.

Our third example will be discrete. Consider the n-dimensional discrete cube $X = \{0,1\}^n$ and equip X with the normalized Hamming metric

$$d(x,y) = \frac{1}{n}\,\mathrm{Card}\big(\{1 \leq i \leq n : x_i \neq y_i\}\big),$$

$x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \{0,1\}^n$. Let $\mu = \mu^n$ be uniform (product) measure on $\{0,1\}^n$ defined by $\mu(A) = 2^{-n}\,\mathrm{Card}\,(A)$ for every subset A of X. Identifying the extremal sets A in X for which the infimum $\inf\{\mu(A_r); \mu(A) \geq \frac{1}{2}\}$ is attained may be used to show here that

$$\alpha_\mu(r) \leq \mathrm{e}^{-2nr^2}, \quad r > 0,$$

where the concentration function α_μ for μ on $\{0,1\}^n$ equipped with the Hamming metric is defined as above.

A further, non-product, example is given by the symmetric group Π^n of permutations of $\{1, \ldots, n\}$ equipped with the normalized metric

$$d(\sigma,\pi) = \frac{1}{n}\,\mathrm{Card}\big(\{1 \leq i \leq n; \sigma(i) \neq \pi(i)\}\big)$$

and uniform measure μ (assigning mass $(n!)^{-1}$ to each permutation). Then

$$\alpha_\mu(r) \leq \mathrm{e}^{-nr^2/32}, \quad r > 0.$$

1.2 Concentration functions and deviation inequalities

Motivated by these examples, we introduce and formalize the concept of concentration function of a probability measure on, say, a metric space. The preceding concentration examples indeed rely on two main ingredients, a (probability) measure and a notion of (isoperimetric) enlargement with respect to which concentration is evaluated.

Let thus (X,d) be a metric space equipped with a probability measure μ on the Borel sets of (X,d). In other words, (X,d,μ) is a metric measure space in the sense of M. Gromov [Grom]). The concentration function $\alpha_{(X,d,\mu)}$ (denoted more simply α_μ when the underlying metric space (X,d), is implicit) of μ is defined as

$$\alpha_{(X,d,\mu)}(r) = \sup\big\{1 - \mu(A_r); A \subset X, \mu(A) \geq \tfrac{1}{2}\big\}, \quad r > 0. \tag{1.5}$$

Here $A_r = \{x \in X; d(x, A) < r\}$ is the (open) r-neighborhood of A (with respect to d). A concentration function is less than or equal to $\frac{1}{2}$, and decreases to 0 as $r \to \infty$. When (X, d) is bounded, the enlargements $r > 0$ in (1.5) actually range up to the diameter of (X, d), the concentration function being 0 when r is larger than the diameter. By definition of the concentration function $\alpha_\mu = \alpha_{(X,d,\mu)}$, given a set A with measure $\mu(A) \geq \frac{1}{2}$, the set of points which are within distance $r > 0$ from a point in A has measure larger than or equal to $1 - \alpha_\mu(r)$.

The idea of the concentration of measure phenomenon is that, in a number of basic examples, $\alpha_{(X,d,\mu)}(r)$ decreases rapidly as r, or the dimension of X, is large. In particular, we say that μ has normal concentration on (X, d) if there are constants $C, c > 0$ such that, for every $r > 0$,

$$\alpha_{(X,d,\mu)}(r) \leq C\,\mathrm{e}^{-cr^2}. \tag{1.6}$$

Important examples share this normal concentration, and we will mostly be concerned with this property throughout these notes. In particular, as we have seen with (1.2), the normalized invariant measure σ^n on the standard n-sphere $\mathbb{S}^n$, $n \geq 2$, satisfies a normal concentration with $c = (n-1)/2$ and $C = 1$ that thus yields strong concentration in high dimension. By (1.4), the canonical Gaussian measure on Euclidean space satisfies this concentration (with constants $c = \frac{1}{2}$ and $C = 1$ independent of the dimension). Concentration on the cube $\{0, 1\}^n$ and the symmetric group Π^n also belong to this family. While these examples follow from stronger isoperimetric inequalities, the approximate property of measure concentration may be investigated in settings far away from isoperimetry as demonstrated by the tools discussed in these notes. See also [Ta1] for further examples.

One important technical aspect of the preceding definition is that it may be expressed equivalently in terms of deviation and concentration inequalities for Lipschitz functions. If μ is a probability measure on the Borel sets of (X, d), and if F is a measurable real-valued function on (X, d), we say m_F is a median of F for μ if

$$\mu\big(\{F \leq m_F\}\big) \geq \tfrac{1}{2} \quad \text{and} \quad \mu\big(\{F \geq m_F\}\big) \geq \tfrac{1}{2}\,.$$

A median m_F may not be unique. A real-valued function F on (X, d) is said to be Lipschitz if

$$\|F\|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|F(x) - F(y)|}{d(x, y)} < \infty.$$

We say that F is 1-Lipschitz if $\|F\|_{\mathrm{Lip}} \leq 1$.

If F is Lipschitz on (X, d) and if $A = \{F \leq m\}$, for every $r > 0$ we have the inclusion $A_r \subset \{F < m + r\|F\|_{\mathrm{Lip}}\}$. Therefore, if $m = m_F$ is a median of F for μ, we get that for every $r > 0$,

$$\mu\big(\{F \geq m_F + r\}\big) \leq \alpha_\mu\big(r/\|F\|_{\mathrm{Lip}}\big). \tag{1.7}$$

We speak of (1.7) as a deviation inequality. The same argument with $-F$ yields

$$\mu\big(\{F \leq m_F - r\}\big) \leq \alpha_\mu\big(r/\|F\|_{\mathrm{Lip}}\big).$$

Therefore, together with (1.7), we deduce that for every $r > 0$,

$$\mu\big(\big\{|F - m_F| \geq r\big\}\big) \leq 2\alpha_\mu\big(r/\|F\|_{\mathrm{Lip}}\big). \tag{1.8}$$

This inequality describes a concentration inequality of F around its median (one of them) with rate α_μ. According to the relative size of α_μ and $\|F\|_{\mathrm{Lip}}$, the Lipschitz function F "concentrates" around one constant value on a portion of the space of big measure. For example, by (1.2), a Lipschitz function on a high-dimensional sphere will be almost constant on almost all the space. Moreover, it should be emphasized that m_F and $\|F\|_{\mathrm{Lip}}$ might be of rather different scales, an observation of fundamental importance in applications. On the other hand, concentration usually does not yield any particular kind of information on the size of the Lipschitz functions themselves (in particular of m_F). By homogeneity, it is enough to consider the preceding deviation and concentration inequalities for 1-Lipschitz functions.

The deviation or concentration inequalities on Lipschitz functions (1.7) and (1.8) are actually equivalent to the corresponding statement on sets. Let A be a Borel set in (X, d) with $\mu(A) \geq \frac{1}{2}$. Set $F(x) = d(x, A)$, $x \in X$. Clearly $\|F\|_{\mathrm{Lip}} \leq 1$ while

$$\mu\big(\{F = 0\}\big) \geq \mu(A) \geq \frac{1}{2}\,.$$

Hence, since $F \geq 0$, 0 is a median of F for μ, and thus, by (1.7), for every $r > 0$,

$$1 - \mu(A_r) = \mu\big(\{F \geq r\}\big) \leq \alpha_\mu(r).$$

We may summarize these conclusions in the following statement.

Proposition 1.1. *Let (X, d, μ) be a metric measure space, and let $\alpha_\mu = \alpha_{(X,d,\mu)}$ be its concentration function. Then, for every $r > 0$,*

$$\alpha_\mu(r) = \sup \mu\big(\{F \geq m_F + r\}\big)$$

where the supremum runs over all 1-Lipschitz functions $F : X \to \mathbb{R}$.

The notion of observable diameter emphasized by M. Gromov [Grom] is dual to the one of concentration function, and may be used as a further description of measure concentration. It describes the diameter of a metric space (X, d) viewed through a given probability measure μ on the Borel sets of (X, d). Fix $\kappa > 0$ to be thought of as small. define first the partial diameter $\mathrm{PartDiam}_\mu(X, d)$ of (X, d) with respect to μ as the infimal D such that there exists a subset A of X with diameter less than or equal to D and measure $\mu(A) \geq 1-\kappa$. This diameter is clearly monotone for the Lipschitz ordering: if $\varphi : (X,d) \to (Y,\delta)$ is 1-Lipschitz, and if μ_φ is the pushed forward measure μ by φ, then $\mathrm{PartDiam}_{\mu_\varphi}(Y,\delta) \leq \mathrm{PartDiam}_\mu(X,d)$ (for all $\kappa > 0$). What is not obvious is that the partial diameter may dramatically decrease under all 1-Lipschitz maps from X to a certain Y, that we always take to be $\mathbb{R}$ below. We then define the observable diameter $\mathrm{ObsDiam}_\mu(X,d)$ of (X,d) with respect to μ as the supremum of $\mathrm{PartDiam}_{\mu_F}(\mathbb{R})$ over all image measures μ_F of μ by a 1-Lipschiz map $F : X \to \mathbb{R}$. Following [Grom2], we think of μ as a state on the configuration space (X,d) and a Lipschitz map $F : X \to \mathbb{R}$ is interpreted as an observable giving the tomographic image μ_F on $\mathbb{R}$. We watch μ_F and can only distinguish a part of its support of measure $1-\kappa$.

On the basis of (1.8), it is not difficult to see (cf. [Le]) that

$$\mathrm{ObsDiam}_\mu(X,d) \leq 2\alpha_\mu^{-1}\Big(\frac{\kappa}{2}\Big)$$

(where α_μ^{-1} is the generalized inverse function of α_μ. As an example, if μ has normal concentration $\alpha_\mu(r) \leq C\,\mathrm{e}^{-cr^2}$, $r > 0$, on (X,d), then

$$\mathrm{ObsDiam}_\mu(X,d) \leq 2\sqrt{\frac{1}{c}\,\log\frac{2C}{\kappa}}\,.$$

The important parameter in this bound is the rate c in the exponential decay of the concentration function, the value of $C > 0$ being usually a numerical constant that simply modifies the numerical value of κ by a factor. For example, by (1.2),

$$\mathrm{ObsDiam}_{\sigma^n}(\mathbb{S}^n) = O\Big(\frac{1}{\sqrt{n}}\Big)$$

as n is large, which is of course in strong contrast with the diameter of $\mathbb{S}^n$ itself as a metric space. Similarly, the observable diameter of Euclidean space with respect to Gaussian measures is bounded.

Inequality (1.8) describes a concentration property of the Lipschitz function F around some median value m_F. Often, measure concentration is expressed by tail inequalities around the mean of Lipschitz functions. The following statement easily shows that these also entail measure concentration on sets. For simplicity, we only state it for normal concentration. The numerical constants are not sharp.

Proposition 1.2. *Let μ be a Borel probability measure on a metric space (X,d). Assume that for any bounded 1-Lipschitz function F on (X,d),*

$$\mu\big(\{F \geq \textstyle\int F d\mu + r\}\big) \leq C\,\mathrm{e}^{-cr^2/2} \tag{1.9}$$

for every $r > 0$. Then

$$\alpha_{(X,d,\mu)}(r) \leq C\,\mathrm{e}^{-cr^2/8}, \quad r > 0.$$

Moreover, any 1-Lipschitz function F is integrable with respect to μ and satisfies (1.9).

Proof. Let A with $\mu(A) \geq \frac{1}{2}$ and fix $r > 0$. Consider $F(x) = \min(d(x,A), r)$, $x \in X$. Clearly $\|F\|_{\mathrm{Lip}} \leq 1$, while

$$\int F d\mu \leq \big(1 - \mu(A)\big) r \leq \frac{r}{2}\,.$$

By the hypothesis,

$$\begin{aligned} 1 - \mu(A_r) &= \mu\big(\{F \geq r\}\big) \\ &\leq \mu\big(\big\{F \geq \textstyle\int F d\mu + \frac{r}{2}\big\}\big) \\ &\leq C\,\mathrm{e}^{-cr^2/8}. \end{aligned}$$

The first claim follows.

Let now F be a 1-Lipschitz function on (X,d). For every $n \geq 0$, $F_n = \min(|F|, n)$ is again 1-Lipschitz and bounded. Applying (1.9) to $-F_n$, for every $r > 0$,

$$\mu\big(\{F_n \leq \textstyle\int F_n d\mu - r\}\big) \leq C\,\mathrm{e}^{-cr^2/2}. \tag{1.10}$$

Choose m such that $\mu(\{|F| \leq m\}) \geq \frac{1}{2}$ and r_0 such that $C\mathrm{e}^{-cr_0^2/8} < \frac{1}{2}$. Since for every n, $\mu(\{F_n \leq m\}) \geq \frac{1}{2}$, intersecting with (1.10) for $r = r_0$, we get that, independently of n,

$$\textstyle\int F_n d\mu \leq m + r_0$$

and thus $\int |F| d\mu < \infty$ by monotone convergence. Apply then (1.9) to the 1-Lipschitz function $\min(\max(F, -n), n)$ and let $n \to \infty$. Proposition 1.2 is established. $\square$

A convenient tool to achieve normal concentration is Laplace transforms. If (X,d,μ) is a metric measure space, define the the Laplace functional of μ on (X,d) as

$$\mathrm{E}_{(X,d,\mu)}(\lambda) = \sup \int \mathrm{e}^{\lambda F} d\mu, \quad \lambda \in \mathbb{R}, \tag{1.11}$$

where the supremum runs over all (bounded) mean zero 1-Lipschitz functions F on (X,d). We often write more simply $\mathrm{E}_\mu = \mathrm{E}_{(X,d,\mu)}$.

The following elementary proposition, that follows from Chebyshev's inequality, gives a simple criterion on the Laplace functional E_μ for normal concentration.

Proposition 1.3. *Let μ be a probability measure on the Borel sets of a metric space (X,d). If, for some constant $C > 0$,*

$$\mathrm{E}_{(X,d,\mu)}(\lambda) \leq \mathrm{e}^{C\lambda^2/2}, \quad \lambda \in \mathbb{R}, \tag{1.12}$$

then, every 1-Lipschitz function $F : X \to \mathbb{R}$ is integrable and for every $r \geq 0$,

$$\mu\big(\{F \geq \textstyle\int F d\mu + r\}\big) \leq \mathrm{e}^{-r^2/2C}.$$

In particular, (X,d,μ) has normal concentration

$$\alpha_{(X,d,\mu)}(r) \le \mathrm{e}^{-r^2/8C}, \quad r > 0.$$

The Laplace functional $\mathrm{E}_{(X,d,\mu)}$ is a convenient tool to handle concentration in product spaces with respect to the ℓ^1-metric. Namely, if (X,d) and (Y,δ) are two metric spaces, we equip the product space $X \times Y$ with the metric

$$d(x,x') + \delta(y,y'), \quad x,x' \in X,\ y,y' \in Y.$$

Then

$$\mathrm{E}_{(X\times Y, d+\delta, \mu\otimes\nu)} \le \mathrm{E}_{(X,d,\mu)}\, \mathrm{E}_{(Y,\delta,\nu)}.$$

In particular, if (1.12) holds for metric measure spaces (X_i, d_i, μ_i), $i = 1, \ldots, n$, it holds with constant Cn for $X_1 \times \cdots \times X_n$ and $\mu_1 \otimes \cdots \otimes \mu_n$ with respect to the ℓ^1-metric $\sum_{i=1}^n d_i$. In this form, measure concentration behaves ackwardly with respect to products, and stability for the ℓ^1-products, as on the discrete cube, induces dimensional factors. The basic example of the Gaussian measure suggests that tensorization should be well-behaved with respect to the classical ℓ^2-metric. This is why we investigate here various descriptions of measure concentration that will be well-suited with respect to Euclidean tensorization and motivated by dimension free bounds. We refer to [Ta1], [Le] for further aspects on concentration for product measures.

2. A triple description of measure concentration

In this section, we present three approaches to measure concentration, geometric (through Brunn-Minkoswki inequalities), measure-theoretic (through transportation cost inequalities) and functional (through logarithmic Sobolev inequalities). All of them are well-behaved with respect to Euclidean products, are satisfied by families of log-concave measures, and yield dimension free concentration bounds. Mass transportation will be used as a first unifying tool to establish the basic inequalities.

2.1 Brunn-Minkowski inequalities

Our first description will be geometric. Brunn-Minkoswki inequalities may indeed be used to provide simple but powerful concentration results.

The classical Brunn-Minkowski inequality indicates that for all bounded Borel measurable sets A, B in $\mathbb{R}^n$,

$$\mathrm{vol}_n(A+B)^{1/n} \ge \mathrm{vol}_n(A)^{1/n} + \mathrm{vol}_n(B)^{1/n} \tag{2.1}$$

where $A+B = \{x+y; x \in A, y \in B\}$ is the Minkowski sum of A and B and where we recal that $\mathrm{vol}_n(\cdot)$ denotes the volume element in $\mathbb{R}^n$. In its equivalent (dimension free) multiplicative form, for every $\theta \in [0,1]$,

$$\mathrm{vol}_n\big(\theta A + (1-\theta)B\big) \geq \mathrm{vol}_n(A)^\theta \mathrm{vol}_n(B)^{1-\theta}. \tag{2.2}$$

Indeed, under (2.1),

$$\begin{aligned}\mathrm{vol}_n\big(\theta A + (1-\theta)B\big)^{1/n} &\geq \theta\, \mathrm{vol}_n(A)^{1/n} + (1-\theta)\mathrm{vol}_n(B)^{1/n} \\ &\geq \Big(\mathrm{vol}_n(A)^\theta \mathrm{vol}_n(B)^{1-\theta}\Big)^{1/n}\end{aligned}$$

by the arithmetic-geometric mean inequality. Conversely, if $A' = \mathrm{vol}_n(A)^{-1/n}A$ and $B' = \mathrm{vol}_n(B)^{-1/n}B$, then (2.2) implies that $\mathrm{vol}_n(\theta A' + (1-\theta)B') \geq 1$ for every $\theta \in [0,1]$. Since

$$\theta A' + (1-\theta)B' = \frac{A+B}{\mathrm{vol}_n(A)^{1/n} + \mathrm{vol}_n(B)^{1/n}}$$

for

$$\theta = \frac{\mathrm{vol}_n(A)^{1/n}}{\mathrm{vol}_n(A)^{1/n} + \mathrm{vol}_n(B)^{1/n}}\,,$$

(2.1) immediately follows by homogeneity.

The Brunn-Minkowski inequality may be used to produce a simple proof of the classical isoperimetric inequality in $\mathbb{R}^n$ by taking B the ball with center the origin and radius $r > 0$: then (2.1) shows that

$$\mathrm{vol}_n(A_r)^{1/n} = \mathrm{vol}_n\big(A+B\big)^{1/n} \geq \mathrm{vol}_n(A)^{1/n} + v(r)^{1/n}$$

where $v(r)$ is the volume of the Euclidean ball of radius $r > 0$. Whenever $\mathrm{vol}_n(A) = \mathrm{vol}_n(D)$ for some ball D with radius $s > 0$, since $v^{1/n}$ is linear,

$$\mathrm{vol}_n(A)^{1/n} + v(r)^{1/n} = v(s+r)^{1/n}$$

so that

$$\mathrm{vol}_n(A_r) \geq \mathrm{vol}_n(D_r)$$

which amounts to isoperimetry.

The multiplicative form of the Brunn-Minskowski inequality admits a functional version. We refer to [DG], [Bar] for the historical aspects of this result.

Theorem 2.1. *Let $\theta \in [0,1]$ and let u, v, w be non-negative measurable functions on $\mathbb{R}^n$ such that for all $x, y \in \mathbb{R}^n$,*

$$w\big(\theta x + (1-\theta)y\big) \geq u(x)^\theta v(y)^{1-\theta}. \tag{2.3}$$

Then

$$\int w dx \geq \left(\int u dx\right)^{\theta}\left(\int v dx\right)^{1-\theta}. \tag{2.4}$$

Applied to the characteristic functions of bounded measurable sets A and B in $\mathbb{R}^n$, it yields the multiplicative form (2.2) of the geometric Brunn-Minkowski inequality.

For the sake of completeness, we present a proof of the functional Brunn-Minkowski theorem using mass transportation that we learned from F. Barthe (cf. [Bar]). Mass transportation is used in [CE-MC-S] to provide the suitable extension of Theorem 2.1 in a Riemannian context using geodesics and curvature.

Proof. We start with $n = 1$ and then perform induction on dimension. By homogeneity, we may assume that $\int u dx = \int v dx = 1$, and by approximation that u and v are continuous with strictly positive values. Set $d\mu = u dx$, $d\nu = v dx$, and define $T : \mathbb{R} \to \mathbb{R}$ by

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

Then T is increasing and differentiable, and ν is the image measure of μ by T. By the change of variable formula,

$$v\big(T(x)\big)T'(x) = u(x), \quad x \in \mathbb{R}.$$

Set then $z(x) = \theta x + (1-\theta)T(x)$ so that $z'(x) = \theta + (1-\theta)T'(x)$. By the arithmetic-geometric mean inequality, for every x,

$$z'(x) \geq \big(T'(x)\big)^{1-\theta}. \tag{2.5}$$

Now, since z is injective, by the hypothesis (2.3) and (2.5),

$$\begin{aligned}
\int w dx &= \int w\big(z(x)\big)z'(x)dx \\
&\geq \int u(x)^{\theta} v\big(T(x)\big)^{1-\theta}\big(T'(x)\big)^{1-\theta} dx \\
&= \int u dx = 1.
\end{aligned}$$

This proves the case $n = 1$. It is then easy to deduce the general case by induction on n. Suppose $n > 1$ and assume the Brunn-Minkowski theorem holds in $\mathbb{R}^{n-1}$. Let u, v, w be non-negative measurable functions on $\mathbb{R}^n$ satisfying (2.3) for some $\theta \in [0,1]$. Let $q \in \mathbb{R}$ be fixed and define $u_q : \mathbb{R}^{n-1} \to [0, \infty)$ by $u_q(x) = u(x, q)$ and similarly for v_q and w_q. Clearly, if $q = \theta q_0 + (1-\theta)q_1$, $q_0, q_1 \in \mathbb{R}$,

$$w_q\big(\theta x + (1-\theta)y\big) \geq u_{q_0}(x)^{\theta} v_{q_1}(y)^{1-\theta}$$

for all $x, y \in \mathbb{R}^{n-1}$. Therefore, by the induction hypothesis,

$$\int_{\mathbb{R}^{n-1}} w_q dx \geq \left(\int_{\mathbb{R}^{n-1}} u_{q_0} dx\right)^{\theta} \left(\int_{\mathbb{R}^{n-1}} v_{q_1} dx\right)^{1-\theta}.$$

Finally, applying the one-dimensional case shows that

$$\int w dx = \int \left(\int_{\mathbb{R}^{n-1}} w_q dx\right) dq \geq \left(\int u dx\right)^{\theta} \left(\int v dx\right)^{1-\theta}$$

which is the desired result. Theorem 2.1 is established. □

The transportation argument for dimension one may actually be extended to higher dimension by the introduction of monotone mass transportation (see below) to give a direct proof of the theorem (cf. [Bar]).

As announced, Brunn-Minkowski inequalities may be used to produce concentration type inequalities, including the basic example of uniform measures on spheres (see [Le]) and Gaussian measures. The argument goes back to [Mau].

Assume that μ is a probability measure on $\mathbb{R}^n$ with smooth strictly positive density e^{-V} with respect to Lebesgue measure, where V is strictly convex in the sense that for some $c > 0$ and all $x, y \in \mathbb{R}^n$,

$$V(x) + V(y) - 2V\left(\frac{x+y}{2}\right) \geq \frac{c}{4}\,|x-y|^2. \tag{2.6}$$

The typical example is of course the canonical Gaussian measure γ on $\mathbb{R}^n$ for which $c = 1$. Given a bounded measurable function f on $\mathbb{R}^n$, apply then the functional Brunn-Minkowski Theorem 2.1 to

$$u(x) = \mathrm{e}^{Qf(x)-V(x)}, \quad v(y) = \mathrm{e}^{-f(y)-V(y)}, \quad w(z) = \mathrm{e}^{-V(z)},$$

where we define Qf as the infimum-convolution

$$Qf(x) = \inf_{y \in \mathbb{R}^n} \left[f(y) + \frac{c}{4}\,|x-y|^2\right], \quad x \in \mathbb{R}^n.$$

By definition of Qf and the convexity hypothesis (2.6) on V, condition (2.3) is satisfied with $\theta = \frac{1}{2}$ so that

$$1 \geq \int \mathrm{e}^{Qf} d\mu \int \mathrm{e}^{-f} d\mu. \tag{2.7}$$

Given now a measurable set A in X, apply this result to the function f that is equal to 0 on A and $+\infty$ outside. Then $Qf(x) = \frac{c}{4}\, d(x, A)^2$ where $d(x, A)$ is the Euclidean distance from the point x to the set A and thus

$$\int \mathrm{e}^{\frac{c}{4}\, d(\cdot, A)^2} d\mu \leq \frac{1}{\mu(A)}\,.$$

Hence, for every $r > 0$,

$$1 - \mu(A_r) \leq \frac{1}{\mu(A)} \, \mathrm{e}^{-cr^2/4}.$$

Theorem 2.2. *Let $d\mu = \mathrm{e}^{-V} dx$ where V satisfies (2.6). Then,*

$$\alpha_\mu(r) \leq 2 \, \mathrm{e}^{-cr^2/4}, \quad r > 0.$$

In particular

$$\alpha_\gamma(r) \leq 2 \, \mathrm{e}^{-r^2/4}, \quad r > 0,$$

for the canonical Gaussian measure γ on $\mathbb{R}^n$.

2.2 Transportation cost inequalities

We next turn to our second description of measure concentration that will be measure-theoretic. Let us start with the classical Pinsker-Csizsar-Kullback inequality (cf. [Pi], [R-R]) that indicates that whenever μ and ν are two probability measures, then

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \sqrt{\frac{1}{2} \, \mathrm{H}\big(\nu \,|\, \mu\big)} \, . \tag{2.8}$$

Here $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation distance, whereas $\mathrm{H}(\nu \,|\, \mu)$ is the relative entropy of ν with respect to μ defined by

$$\mathrm{H}\big(\nu \,|\, \mu\big) = \int \log \frac{d\nu}{d\mu} \, d\nu$$

whenever ν is absolutely continuous with respect to μ with Radon-Nikodym derivative $\frac{d\nu}{d\mu}$, and $+\infty$ if not. Inequalities such as (2.8) have been often considered in information theory.

That such an inequality is related to concentration properties was emphasized by K. Marton [Mar1], [Mar2], and may be shown in the following way. Given a metric space (X, d) and two Borel probability measures μ and ν on X, consider the Wasserstein distance between μ and ν

$$\mathrm{W}_1(\mu, \nu) = \inf \int\!\!\int d(x, y) d\pi(x, y)$$

where the infimum runs over all probability measures π on the product space $X \times X$ with marginals μ and ν having a finite first moment (for d). By the Monge-Kantorovitch theorem (see below), the total variation distance corresponds to the trivial metric on X. Given μ, consider then the inequality

$$\mathrm{W}_1(\mu, \nu) \leq \sqrt{2C \, \mathrm{H}\big(\nu \,|\, \mu\big)} \tag{2.9}$$

for some $C > 0$ and every ν. Let A and B be Borel sets with $\mu(A), \mu(B) > 0$, and consider the conditional probabilities $\mu_A = \mu(\cdot\,|A)$ and $\mu_B = \mu(\cdot\,|B)$. By the triangle inequality for W_1 and (2.9),

$$\begin{aligned} \mathrm{W}_1(\mu_A, \mu_B) &\leq \mathrm{W}_1(\mu, \mu_A) + \mathrm{W}_1(\mu, \mu_B) \\ &\leq \sqrt{2C\,\mathrm{H}\big(\mu_A\,|\,\mu\big)} + \sqrt{2C\,\mathrm{H}\big(\mu_B\,|\,\mu\big)} \\ &= \sqrt{2C \log \frac{1}{\mu(A)}} + \sqrt{2C \log \frac{1}{\mu(B)}}\,. \end{aligned} \tag{2.10}$$

Now, all measures with marginals μ_A and μ_B must be supported on $A \times B$, so that, by definition of W_1,

$$\mathrm{W}_1(\mu_A, \mu_B) \geq d(A, B) = \inf\big\{d(x, y); x \in A, y \in B\big\}.$$

Then (2.10) implies a concentration inequality. Indeed, given A and B in X such that $d(A, B) \geq r > 0$, we get

$$r \leq \sqrt{2C \log \frac{1}{\mu(A)}} + \sqrt{2C \log \frac{1}{1 - \mu(A_r)}} \tag{2.11}$$

where we recall that $A_r = \{x \in X, ; d(x, A) < r\}$. Inequality (2.11) appears as a symmetric form of concentration. If, say, $\mu(A) \geq \frac{1}{2}$,

$$r \leq \sqrt{2C \log 2} + \sqrt{2C \log \frac{1}{1 - \mu(A_r)}}$$

so that, whenever $r \geq 2\sqrt{2C \log 2}$ for example,

$$1 - \mu(A_r) \leq \mathrm{e}^{-r^2/8C}.$$

As for Laplace transforms, while suitable to ℓ^1-metrics, the W_1 Wasserstein distance does not behave dimensional free with respect to Euclidean products. Motivated again by the example of Gaussian measures, it is more fruitful in order to reach dimension free concentration properties to think in terms of a quadratic cost. To this task, let us restrict ourselves to the case of $\mathbb{R}^n$ with the Euclidean norm $|\cdot|$. Given a probability measure μ on the Borel sets of $\mathbb{R}^n$, say that it satisfies a quadratic transportation cost inequality whenever there exists a constant $C > 0$ such that for all probability measures ν,

$$\mathrm{W}_2(\mu, \nu) \leq \sqrt{C\,\mathrm{H}\big(\nu\,|\,\mu\big)}. \tag{2.12}$$

Here W_2 is the Wasserstein distance with quadratic cost

$$\mathrm{W}_2(\mu, \nu) = \inf\bigg(\int\!\!\int \frac{1}{2}\,|x - y|^2 d\pi(x, y)\bigg)^{1/2}$$

where the infimum is running over all probability measures π on $\mathbb{R}^n \times \mathbb{R}^n$ with respective marginals μ and ν. (The infimum in W_2 is finite as soon as μ and ν have finite second moment which we shall always assume.) It is clear by Jensen's inequality that the quadratic transportation cost inequality is stronger than the W_1 transportation cost inequality (with $d(x,y) = |x-y|$, $x, y \in \mathbb{R}^n$).

It has been shown by M. Talagrand [Ta2] that the canonical Gaussian measure γ on $\mathbb{R}^n$ satisfies (2.12) with $C = 1$. We outline a proof of it relying, as for the Brunn-Minkowski theorem, on mass transportation. We first consider the one-dimensional case $n = 1$. Let $f \geq 0$ such that $\int f d\gamma = 1$, and set $d\nu = f d\gamma$. For simplicity, assume that $f > 0$ everywhere. As in the proof of Theorem 2.1, let $T : \mathbb{R} \to \mathbb{R}$ be defined by

$$\nu\big((-\infty, T(x)]\big) = \gamma\big((-\infty, x]\big), \quad x \in \mathbb{R},$$

so that ν is the image measure of γ by the increasing map T. By the change of variables formula,

$$f\big(T(x)\big)T'(x)\, \mathrm{e}^{-T(x)^2/2} = \mathrm{e}^{-x^2/2}, \quad x \in \mathbb{R}.$$

Hence, taking logarithms, for every x,

$$\log f\big(T(x)\big) + \log T'(x) - \frac{1}{2}\, T(x)^2 = -\frac{1}{2}\, x^2.$$

Integrating with respect to γ, and using that $T * \gamma = \nu$, we get that

$$\int \log f d\nu = \frac{1}{2} \int \big[T(x)^2 - x^2\big] d\gamma - \int \log T' d\gamma.$$

Integrating by parts,

$$\int x(T - x) d\gamma = \int (T' - 1) d\gamma$$

so that

$$\begin{aligned} \int \log f d\nu &= \frac{1}{2} \int \big|x - T(x)\big|^2 d\gamma + \int [T' - 1 - \log T'] d\gamma \\ &\geq \frac{1}{2} \int \big|x - T(x)\big|^2 d\gamma \end{aligned}$$

where we used that $\alpha - 1 - \log \alpha \geq 0$, $\alpha \geq 0$. Now, since ν is the image of γ by T, the image measure π of γ under the map $x \mapsto (x, T(x))$ has marginals γ and ν respectively so that

$$\frac{1}{2} \int \big|x - T(x)\big|^2 d\gamma = \iint \frac{1}{2}\, |x - y|^2 d\pi \geq \mathrm{W}_2(\gamma, \nu)^2.$$

The result follows in this case. The n-dimensional case follows by tensorization from the fundamental product property of entropy (together with additivity of the Euclidean cost) as outlined in [Ta2]. The argument is summarized in the next simple proposition (see [Le] for a proof).

Proposition 2.3. *Let $P = \mu_1 \otimes \cdots \otimes \mu_n$ be a product probability measure on the Borel sets of $\mathbb{R}^n$. Assume that each μ_i, $i = 1, \ldots, n$, satisfies a quadratic transportation cost inequality*

$$\mathrm{W}_2(\mu_i, \nu_i) \le \sqrt{C_i \, \mathrm{H}\big(\nu_i \,|\, \mu_i\big)}$$

for every ν_i on $\mathbb{R}$. Then

$$\mathrm{W}_2(P, R) \le \sqrt{\max_{1 \le i \le n} C_i \, \mathrm{H}\big(R \,|\, P\big)}$$

for every probability measure R on $\mathbb{R}^n$.

The n-dimensional case may actually be proved alternatively following the one-dimensional case ([Bl], [CE]) by appropriate monotone transportation in the form of the Brenier-McCann [Br], [MC] map. Given two probability measures μ and ν on $\mathbb{R}^n$, a map $T : \mathbb{R}^n \to \mathbb{R}^n$ (defined μ-almost everywhere) is said to push μ forward to ν (or to transport μ onto ν) if ν is the image measure of μ under T. In other words, for every bounded non-negative Borel functions $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\int \varphi(y) d\nu(y) = \int \varphi\big(T(x)\big) d\mu(x).$$

If μ and ν have finite second moments, a map T pushing μ forward to ν is said to be optimal with respect to the Wasserstein distance W_2 if

$$\mathrm{W}_2(\mu, \nu)^2 = \frac{1}{2} \int \big|x - T(x)\big|^2 d\mu(x).$$

A fundamental result of Y. Brenier [Br], and R. McCann [MC] (cf. [Vi]), is that when μ is absolutely continuous with respect to Lebesgue measure, there exists a convex function ϕ such that $T = \nabla\phi$ pushes μ towards ν in the optimal sense.

Take $\mu = \gamma$ the canonical Gaussian measure on $\mathbb{R}^n$, and assume that $d\nu = f d\gamma$ for some $f \ge 0$ with $\int f d\gamma = 1$. Whenever licit, the change of variables formula in the transport from γ to ν leads to the so-called Monge-Ampère equation

$$f\big(T(x)\big) \det\big(\mathrm{Hess}\, \phi(x)\big) \mathrm{e}^{-|T(x)|^2/2} = \mathrm{e}^{-|x|^2/2}, \quad x \in \mathbb{R}^n,$$

where $\mathrm{Hess}\, \phi$ is the Hessian of ϕ (we ignore here and below the fact that $T = \nabla\phi$ may exist only almost everywhere). Proceeding then exactly as in the one-dimensional case, using that

$$\log\det(\mathrm{Hess}\, \phi(x)) \le \Delta\phi - n = \Delta\Big(\phi - \frac{|x|^2}{2}\Big), \tag{2.13}$$

shows that $\mathrm{W}_2(\gamma, \nu) \le \sqrt{\mathrm{H}(\nu \,|\, \gamma)}$. The argument easily extends to probability measures $d\mu = \mathrm{e}^{-V} dx$ with a strictly convex potential V (in the sense of (2.6)) so to yield the following theorem.

Theorem 2.4. *Let $d\mu = \mathrm{e}^{-V} dx$ where, for some $c > 0$, $\mathrm{Hess}\, V(x) \geq c\, \mathrm{Id}$ uniformly in $x \in \mathbb{R}^n$. Then, for every probability measure ν on $\mathbb{R}^n$,*

$$\mathrm{W}_2(\mu, \nu) \leq \sqrt{\frac{1}{c}\, \mathrm{H}\big(\nu \,|\, \mu\big)}\,.$$

Extensions to Riemannian manifolds are pointed out in [O-V1], [B-G-L]. More general transportation cost inequalities on probabilistic path spaces are investigated in [Wa].

2.3 Logarithmic Sobolev inequalities

Our third description will be functional through logarithmic Sobolev inequalities.

Given a probability measure μ, for every non-negative measurable function f define its entropy as

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

if $\int f \log(1+f) d\mu < \infty$, $+\infty$ if not. In other words, if $\int f d\mu = 1$,

$$\mathrm{Ent}_\mu(f) = \int \log f d\nu = \mathrm{H}\big(\nu \,|\, \mu\big)$$

where $d\nu = f d\mu$. Note that $\mathrm{Ent}_\mu(f) \geq 0$ by Jensen's inequality and that entropy is homogeneous of degree 1.

We introduce the concept of logarithmic Sobolev inequality. To avoid technical questions, let us consider only the case of the Euclidean space $\mathbb{R}^n$. A probability measure μ on the Borel sets of $\mathbb{R}^n$ is said to satisfy a logarithmic Sobolev inequality if for some constant $C > 0$ and all smooth enough functions f on $\mathbb{R}^n$,

$$\mathrm{Ent}_\mu(f^2) \leq 2C \int |\nabla f|^2 d\mu. \tag{2.14}$$

Here ∇f denotes the usual gradient of f and $|\nabla f|$ its Euclidean length. By smooth we understand above and below enough regularity so that the various terms in (2.14) make sense. Changing f^2 into $f > 0$ such that $\int f d\mu = 1$, (2.14) may be written equivalently as

$$\mathrm{H}(\nu \,|\, \mu) \leq \frac{C}{2} \int \frac{|\nabla f|^2}{f} d\mu = \frac{C}{2}\, I(\nu) \tag{2.15}$$

for all ν where $I(\nu)$ is the Fisher information of $d\nu = f d\mu$.

Logarithmic Sobolev inequalities are parts of the family of classical Sobolev inequalities. In terms of Sobolev embeddings, under a logarithmic Sobolev inequality, functions in H^1 do not belong necessarily to some L^p-space with $p > 2$, but to the

Orlicz space $\mathrm{L}^2 \log \mathrm{L}$. This embedding is optimal for the basic example of Gaussian measures. On the other hand, no constant depending on the dimension arises in the logarithmic Sobolev inequality for Gaussian measures. This is one fundamental aspect of the infinite dimensional character of logarithmic Sobolev inequalities that will be exploited here toward dimension free concentration.

The canonical Gaussian measure γ on $\mathbb{R}^n$ satisfies (2.14) with $C = 1$. Numerous proofs of this result may be found in the literature (see [An] and the references therein). We present below two proofs, one relying on semigroup tools, and one based on mass transportation in the spirit of the one for the transportation cost inequality (2.12) (and the Brunn-Minkowski theorem) due to D. Cordero-Erausquin [CE] (he actually presents there a statement including both (2.12) and (2.14) as well as more general inequalities - see below). In particular, the latter shows the central role played by mass transportation in this context.

Theorem 2.5. *For every smooth enough function f on $\mathbb{R}^n$,*

$$\mathrm{Ent}_\gamma(f^2) \le 2\int |\nabla f|^2 d\gamma.$$

First Proof. This proof is based on semigroup methods and diffusion operators. Let the second-order differential operator $\mathrm{L} = \Delta - x \cdot \nabla$ on $\mathbb{R}^n$ with associated semigroup $(P_t)_{t\ge 0}$ called the Hermite or Ornstein-Uhlenbeck semigroup (cf. [Bak1]). The operator L satisfies the integration by parts formula with respect to γ expressed by

$$\int f(-\mathrm{L}g)d\gamma = \int \nabla f \cdot \nabla g \, d\gamma$$

for every smooth functions f, g. The Ornstein-Uhlenbeck semigroup $(P_t)_{t\ge 0}$ admits an explicit integral representation as

$$P_t f(x) = \int f\big(\mathrm{e}^{-t}x + (1-\mathrm{e}^{-2t})^{1/2}y\big)d\gamma(y), \quad t \ge 0,\ x \in \mathbb{R}^n. \tag{2.16}$$

Let f be smooth and non-negative on $\mathbb{R}^n$. To be more precise, we take f smooth and such that $\varepsilon \le f \le 1/\varepsilon$ for some $\varepsilon > 0$ that we take to 0 at the end of the argument. Since $P_0 f = f$ and $\lim_{t\to\infty} P_t f = \int f d\gamma$, write

$$\mathrm{Ent}_\gamma(f) = -\int_0^\infty \frac{d}{dt}\bigg(\int P_t f \log P_t f d\gamma\bigg)dt.$$

Derivation along the semigroup yields the basic relation between entropy and Fisher information. Namely, by the chain rule formula and integration by parts for L,

$$\begin{aligned}\frac{d}{dt}\int P_t f \log P_t f d\gamma &= \int \mathrm{L}P_t f \log P_t f d\gamma + \int \mathrm{L}P_t f d\gamma \\ &= -\int \frac{|\nabla P_t f|^2}{P_t f}\, d\gamma\end{aligned}$$

since γ is invariant under the action of P_t and thus $\int \mathrm{L} P_t f d\gamma = 0$. Now, by the integral representation (2.16), for every $t \geq 0$,

$$\nabla P_t f = \mathrm{e}^{-t} P_t(\nabla f)$$

and thus

$$\big|\nabla P_t f\big| \leq \mathrm{e}^{-t} P_t\big(|\nabla f|\big).$$

By the Cauchy-Schwarz inequality for P_t,

$$P_t\big(|\nabla f|\big)^2 \leq P_t f \, P_t\Big(\frac{|\nabla f|^2}{f}\Big).$$

Summarizing,

$$\mathrm{Ent}_\gamma(f) \leq \int_0^\infty \mathrm{e}^{-2t} \bigg(\int P_t\Big(\frac{|\nabla f|^2}{f}\Big) d\mu\bigg) dt = \frac{1}{2} \int \frac{|\nabla f|^2}{f} \, d\mu$$

by invariance. By the change of f into f^2, the inequality is established.

Second Proof. As announced, the second proof is based on mass transportation and follows [CE]. With respect to the transportation proof of (2.12), we now transport $d\nu = f d\gamma$ ($f \geq 0$, $\int f d\gamma = 1$) towards γ in the optimal sense. The corresponding Monge-Ampère equation then reads

$$f(x)\,\mathrm{e}^{-|x|^2/2} = \mathrm{e}^{-|T(x)|^2/2} \det\big(\mathrm{Hess}\,\phi(x)\big), \quad x \in \mathbb{R}^n.$$

Taking logarithms, and using (2.13),

$$\begin{aligned} \log f(x) - \frac{1}{2}\,|x|^2 &= -\frac{1}{2}\,\big|T(x)\big|^2 + \log\det\big(\mathrm{Hess}\,\phi(x)\big) \\ &\leq -\frac{1}{2}\,\big|T(x)\big|^2 + \Delta\Big(\phi - \frac{|x|^2}{2}\Big). \end{aligned}$$

Hence,

$$\begin{aligned} \log f(x) &\leq -\frac{1}{2}\,|\nabla\phi - x|^2 - x\cdot(\nabla\phi - x) + \Delta\Big(\phi - \frac{|x|^2}{2}\Big) \\ &\leq -\frac{1}{2}\,|\nabla\phi - x|^2 + \mathrm{L}\Big(\phi - \frac{|x|^2}{2}\Big) \end{aligned}$$

where we recall the Ornstein-Uhlenbeck operator $\mathrm{L} = \Delta - x\cdot\nabla$. Integrate now with respect to $d\nu = f d\gamma$ to get

$$\int f \log f d\gamma \leq -\frac{1}{2}\int f|\nabla\phi - x|^2 d\gamma + \int f\,\mathrm{L}\Big(\phi - \frac{|x|^2}{2}\Big) d\gamma.$$

By the integration by parts formula for L,

$$\int f\,\mathrm{L}\Big(\phi - \frac{|x|^2}{2}\Big) d\gamma = -\int \nabla f\cdot\nabla(\phi - x) d\gamma$$

so that

$$\int f\log f d\gamma \le -\frac{1}{2}\int f|\nabla\phi - x|^2 d\gamma - \int \nabla f\cdot\nabla(\phi - x)d\gamma$$
$$\le \frac{1}{2}\int \frac{|\nabla f|^2}{f}\,d\gamma \quad \big(= \frac{1}{2}\,I(\nu)\big),$$

where we used that for vectors a, b in $\mathbb{R}^n$, $-\frac{1}{2}|b|^2 - a\cdot b \le \frac{1}{2}|a|^2$. The proof is complete. $\square$

The preceding proofs work similarly for strictly log-concave probability measures $d\mu = \mathrm{e}^{-V}dx$ in $\mathbb{R}^n$ (cf. [Bak1], [CE]).

Theorem 2.6. *Let $d\mu = \mathrm{e}^{-V}dx$ where, for some $c > 0$, $\mathrm{Hess}\,V(x) \ge c\,\mathrm{Id}$ uniformly in $x \in \mathbb{R}^n$. Then for all smooth functions f on $\mathbb{R}^n$,*

$$\mathrm{Ent}_\mu(f^2) \le \frac{2}{c}\int |\nabla f|^2 d\mu.$$

It should be observed that L. Caffarelli [Ca] (see [Vi]) recently showed that whenever μ is a probability measure on $\mathbb{R}^n$ such that $d\mu = \mathrm{e}^{-V}dx$ and, for some $c > 0$, $\mathrm{Hess}\,V(x) \ge c\,\mathrm{Id}$ uniformly in $x \in \mathbb{R}^n$, then the Brenier-McCann transportation map [Br], [MC] from the canonical Gaussian measure γ to μ is Lipschitz with Lipschitz norm $c^{-1/2}$. In particular, Theorems 2.2, 2.4 and 2.6 all follow from their corresponding statements for Gaussian measure. This justifies limiting ourselves to the Gaussian model.

As for the quadratic transportation cost inequalities, the fundamental tensorization property of entropy shows that logarithmic Sobolev inequalities are stable by products for the Euclidean structure (cf. [Le]).

Proposition 2.7. *Let μ_i on $\mathbb{R}$ satisfy the logarithmic Sobolev inequality*

$$\mathrm{Ent}_{\mu_i}(f^2) \le 2C_i\int f'^2 d\mu_i,$$

$i = 1, \ldots, n$. Then the product measure $P = \mu_1\otimes\cdots\otimes\mu_n$ on $\mathbb{R}^n$ satisfies the logarithmic Sobolev inequality on $\mathbb{R}^n$

$$\mathrm{Ent}_P(f^2) \le 2\max_{1\le i\le n} C_i\int |\nabla f|^2 dP.$$

We now come to the application of logarithmic Sobolev inequalities to measure concentration through the Herbst argument. To start with, let us consider a probability measure μ on the Borel sets of $\mathbb{R}^n$. Assume that μ satisfies the logarithmic Sobolev inequality (2.14). We will show that the Laplace functional E_μ of μ (cf. (1.11)) satisfies

$$\mathrm{E}_\mu(\lambda) \le \mathrm{e}^{C\lambda^2/2}, \quad \lambda \in \mathbb{R}. \tag{2.17}$$

By Proposition 1.3, μ has normal concentration $\alpha_\mu(r) \leq \mathrm{e}^{-r^2/8C}$, $r > 0$.

Let thus F be a smooth bounded 1-Lipschitz function on $\mathbb{R}^n$ such that $\int F d\mu = 0$. In particular, since F is assumed to be regular enough, we can have that $|\nabla F| \leq 1$ at every point. We apply the logarithmic Sobolev inequality (2.14) to $f^2 = \mathrm{e}^{\lambda F}$ for every $\lambda \in \mathbb{R}$. We have

$$\int |\nabla f|^2 d\mu = \frac{\lambda^2}{4} \int |\nabla F|^2 \mathrm{e}^{\lambda F} d\mu \leq \frac{\lambda^2}{4} \int \mathrm{e}^{\lambda F} d\mu.$$

Setting $\Lambda(\lambda) = \int \mathrm{e}^{\lambda F} d\mu$, $\lambda \in \mathbb{R}$, by the definition of entropy,

$$\lambda \Lambda'(\lambda) - \Lambda(\lambda) \log \Lambda(\lambda) \leq \frac{C}{2}\, \lambda^2 \Lambda(\lambda)$$

for every λ. If $K(\lambda) = \frac{1}{\lambda} \log \Lambda(\lambda)$ (with $K(0) = \Lambda'(0)/\Lambda(0) = \int F d\mu = 0$), $\lambda \in \mathbb{R}$, then $K'(\lambda) \leq \frac{C}{2}$ for every λ. Therefore, $K(\lambda) \leq \frac{C\lambda}{2}$ for every λ, that is

$$\int \mathrm{e}^{\lambda F} d\mu \leq \mathrm{e}^{C\lambda^2/2}.$$

Replacing F by a smooth convolution, the latter extends to all mean zero Lipschitz functions and hence (2.17) is established.

3. Relations between the three descriptions

In this third section, we investigate the connections between the preceding descriptions of measure concentration by means of hypercontractive bounds. As discussed in the preceding proofs, optimal mass transportation provides a first unified approach to most of the geometric, measure-theoretic and functional inequalities developed so far. Through the Monge-Kantorovich theorem, hypercontractive bounds will actually appear as dual to mass transportation in this framework. They will allow a (almost) complete circle of implications between the various descriptions presented in the previous section. We start with a simple result between transportation cost inequalities and Laplace bounds that will allow the further developments. The main references for this section are [B-G], [O-V1], [B-G-L], [CE], [CE-G-H], [Vi]...

3.1 Transportation and Laplace bounds

In a first step, we observe that the transportation cost inequality (2.9) is actually equivalent to the normal Laplace bound (2.17) deduced in the preceding chapter from logarithmic Sobolev inequalities, and may be thought of as its dual version. We indeed have the following result due to S. Bobkov and F. Götze [B-G]. We recall the Laplace functional

$$\mathrm{E}_\mu(\lambda) = \sup \int \mathrm{e}^{\lambda F} d\mu, \quad \lambda \in \mathbb{R},$$

where the supremum runs over all 1-Lipschitz mean zero functions $F : X \to \mathbb{R}$, of a probability measure μ on (X, d).

Proposition 3.1. *Let μ be a Borel probability measure on a metric space (X, d). Then*

$$\mathrm{W}_1(\mu, \nu) \leq \sqrt{2C\,\mathrm{H}\big(\nu\,|\,\mu\big)} \tag{3.1}$$

for some $C > 0$ and all ν if and only if

$$\mathrm{E}_\mu(\lambda) \leq \mathrm{e}^{C\lambda^2/2}, \quad \lambda \in \mathbb{R}. \tag{3.2}$$

As we have seen at the end of Section 1, the functional description (3.2) (and thus equivalently (3.1)) of measure concentration is not well adapted as it is to dimension free bounds. This observation motivates the investigation of stronger transportation and logarithmic Sobolev inequalities that do tensorize with dimension free bounds with respect to the ℓ^2-metric (as in the fundamental example of the canonical Gaussian product measure on $\mathbb{R}^n$).

Proof of Proposition 3.1. By the Monge-Kantorovitch-Rubinstein dual characterization (cf. [Du], [R-R], [Vi]) of the Wasserstein distance,

$$\mathrm{W}_1(\mu, \nu) = \sup \bigg[\int g d\nu - \int f d\mu\bigg]$$

where the supremum is running over all bounded measurable functions f and g such that

$$g(x) \leq f(y) + d(x, y)$$

for every $x, y \in X$. Under (3.1),

$$\int g d\nu - \int f d\mu \leq \sqrt{2C\,\mathrm{Ent}_\mu\Big(\frac{d\nu}{d\mu}\Big)}\,,$$

or, equivalently, for every $\lambda > 0$,

$$\int g d\nu - \int f d\mu \leq \frac{C\lambda}{2} + \frac{1}{\lambda}\,\mathrm{Ent}_\mu\Big(\frac{d\nu}{d\mu}\Big).$$

Set $\phi = \frac{d\nu}{d\mu}$. The preceding indicates that

$$\int \psi\phi d\mu \leq \mathrm{Ent}_\mu(\phi)$$

where $\psi = \lambda g - \lambda \int f d\mu - C\lambda^2/2$. Since this inequality holds for every choice of ϕ (i.e. ν), applying it to $\phi = \mathrm{e}^\psi / \int \mathrm{e}^\psi d\mu$ yields that $\log \int \mathrm{e}^\psi d\mu \leq 0$. In other words,

$$\int \mathrm{e}^{\lambda g} d\mu \leq \mathrm{e}^{\lambda \int f d\mu + C\lambda^2/2}.$$

When F is Lipschitz with $\|F\|_{\mathrm{Lip}} \leq 1$, one may choose $F = g = f$ so that the latter amounts to (3.2). Since

$$\mathrm{Ent}_\mu(\phi) = \sup \int \phi\psi d\mu$$

where the supremum is running over all ψ's such that $\int \mathrm{e}^\psi d\mu \leq 1$, the preceding argument clearly indicates that (3.2) is actually equivalent to (3.1). The proof of Proposition 3.1 is complete. $\square$

The same may be proved on the alternate (more classical and easily equivalent) characterization of W_1 as

$$\mathrm{W}_1(\mu, \nu) = \sup \left[\int F d\mu - \int F d\nu \right]$$

where the supremum is running over all 1-Lipschitz functions F on (X, d). However, the preceding choice is motivated by the analagous argument for the quadratic cost (in $X = \mathbb{R}^n$). Namely, the Monge-Kantorovitch theorem states similarly (cf. [R-R], [Vi]) that

$$\mathrm{W}_2(\mu, \nu)^2 = \sup \left[\int g d\nu - \int f d\mu \right] \tag{3.3}$$

where the supremum is running over all bounded functions f and g such that

$$g(x) \leq f(y) + \frac{1}{2} |x - y|^2$$

for every $x, y \in \mathbb{R}^n$. In the infimum-convolution notation,

$$g(x) = \inf_{y \in \mathbb{R}^n} \left[f(y) + \frac{1}{2} |x - y|^2 \right]$$

achieves the optimal choice. Therefore, the proof of Proposition 3.1 may be exactly repeated so to yield the analogous statement for W_2.

Proposition 3.2 *Let μ be a Borel probability measure on $\mathbb{R}^n$. Then*

$$\mathrm{W}_2(\mu, \nu) \leq \sqrt{C\,\mathrm{H}(\nu \,|\, \mu)}$$

for some $C > 0$ and all ν if and only if for all bounded measurable functions f on $\mathbb{R}^n$,

$$\int \mathrm{e}^{Qf} d\mu \leq \mathrm{e}^{\int f d\mu}$$

where Qf is here the infimum-convolution

$$Qf(x) = \inf_{y \in \mathbb{R}^n} \left[f(y) + \frac{1}{2C} |x - y|^2 \right], \quad x \in \mathbb{R}^n.$$

The infimum-convolution inequality of Proposition 3.2 has to be compared to (2.7). For the matter of comparison with Proposition 3.1, observe also that whenever F is Lipschitz,

$$QF \geq F - \frac{C}{2} \|F\|_{\mathrm{Lip}}^2.$$

3.2 Hypercontractivity

To go deeper into the relationships, we actually have to connect the quadratic transportation cost inequality to logarithmic Sobolev inequalities. The key point in this project will the concept of hypercontractivity. From now on, we only consider measures on the Borel sets of $\mathbb{R}^n$ equipped with its standard Euclidean topology.

The fundamental work by L. Gross [Gros] put forward the equivalence between logarithmic Sobolev inequalities and hypercontractivity of the associated heat semigroup. Let us consider a probability measure μ on the Borel sets of $\mathbb{R}^n$ satisfying the logarithmic Sobolev inequality

$$\mathrm{Ent}_\mu(f^2) \leq 2C \int |\nabla f|^2 d\mu \tag{3.4}$$

for some $C > 0$ and all smooth enough functions f on $\mathbb{R}^n$.

For simplicity, assume furthermore that μ has a strictly positive smooth density which may be written e^{-V} for some smooth potential V on $\mathbb{R}^n$. For example, we may restrict below to the canonical Gaussian. Denote by L the second order diffusion operator $\mathrm{L} = \Delta - \nabla V \cdot \nabla$ with invariant measure μ. Integration by parts for L is described by

$$\int f(-\mathrm{L}g) d\mu = \int \nabla f \cdot \nabla g \, d\mu$$

for every smooth functions f, g. Under mild growth conditions on V (that will always be satisfied in applications throughout this work), one may consider the time reversible (with respect to μ) semigroup $(P_t)_{t\geq 0}$ with generator L (cf. [Ev], [F-O-T]). In the Gaussian case, we know that the associated Ornstein-Uhlenbeck $(P_t)_{t\geq 0}$ has the explicit integral representation (2.16). Given f (in the domain of L), $u = u(x,t) = P_t f(x)$ is the fundamental solution of the initial value problem (heat equation with respect to L)

$$\begin{cases} \dfrac{\partial u}{\partial t} - \mathrm{L}u = 0 & \text{in } \mathbb{R}^n \times (0,\infty), \\ \qquad\quad u = f & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases}$$

Logarithmic Sobolev inequalities are a powerful tool to study exponential decay to equilibrium in relative entropy. Namely, since differentiation of entropy along the heat semigroup yields Fisher information, (3.4) implies (is equivalent to the fact) that for every non-negative function f such that $\int f d\mu = 1$,

$$\mathrm{Ent}_\mu(P_t f) \leq \mathrm{e}^{-2t/C} \mathrm{Ent}_\mu(f)$$

for every $t \geq 0$. This estimate is the prototype of exponential entropy dissipation (cf. [Vi]).

One of the main results of the contribution [Gros] by L. Gross is that the logarithmic Sobolev inequality (3.4) for μ holds if and only if the associated heat semigroup $(P_t)_{t\geq 0}$

is hypercontractive in the sense that, for every (or some) $1 < p < q < \infty$, and every f (in L^p),

$$\|P_t f\|_q \leq \|f\|_p \tag{3.5}$$

for every $t > 0$ large enough so that

$$\mathrm{e}^{2t/C} \geq \frac{q-1}{p-1}\,. \tag{3.6}$$

In (3.5), L^p-norms are understood with respect to the measure μ. The key idea of the proof is to consider a function $q(t)$ of $t \geq 0$ such that $q(0) = p$ and to take the derivative in time of $F(t) = \|P_t f\|_{q(t)}$ (for a non-negative smooth function f on $\mathbb{R}^n$). Since the derivative of L^p-norms gives rise to entropy, due to the heat equation $\frac{\partial}{\partial t} P_t f = \mathrm{L} P_t f$ and integration by parts, one gets that

$$\begin{aligned} q(t)^2\, F(t)^{q(t)-1} F'(t) & \\ &= q'(t)\,\mathrm{Ent}_\mu\big((P_t f)^{q(t)}\big) + q(t)^2 \int (P_t f)^{q(t)-1} \mathrm{L} P_t f d\mu \\ &= q'(t)\,\mathrm{Ent}_\mu\big((P_t f)^{q(t)}\big) - 2\big(q(t)-1\big) \int \frac{q(t)^2}{2}\, |\nabla P_t f|^2 (P_t f)^{q(t)-2} d\mu. \end{aligned} \tag{3.7}$$

By the logarithmic Sobolev inequality applied to $(P_t f)^{q(t)/2}$, it follows that $F'(t) \leq 0$ as soon as $q'(t) = 2(q(t)-1)/C$, that is $q(t) = 1 + (p-1)\,\mathrm{e}^{2t/C}$, $t \geq 0$, which yields the claim. It is classical and easy to see that the same argument shows that the logarithmic Sobolev inequality (3.4) is also equivalent to

$$\|\,\mathrm{e}^{P_t f}\|_{\mathrm{e}^{2t/C}} \leq \|\,\mathrm{e}^f\|_1$$

for every $t \geq 0$ and f. For further comparison, observe that by linearity

$$\|\,\mathrm{e}^{P_t f}\|_{a\mathrm{e}^{2t/C}} \leq \quad (\text{resp. } \geq) \quad \|\,\mathrm{e}^f\|_a$$

according as $a \geq 0$ (resp. $a \leq 0$).

The main observation in our project is that a similar relationship for the solutions of Hamilton-Jacobi partial differential equations holds. Consider namely the basic Hamilton-Jacobi initial value problem

$$\begin{cases} \dfrac{\partial v}{\partial t} + \dfrac{1}{2}\,|\nabla v|^2 = 0 & \text{in } \mathbb{R}^n \times (0,\infty), \\ \qquad\qquad\quad v = f & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{3.8}$$

Solutions of (3.8) are described by the Hopf-Lax representation formula as infimum-convolutions. Namely, given a (Lipschitz continuous) function f on $\mathbb{R}^n$, define the one-parameter family of infimum-convolutions of f with the quadratic cost as

$$Q_t f(x) = \inf_{y \in \mathbb{R}^n} \Big[f(y) + \frac{1}{2t}\,|x-y|^2 \Big], \quad t > 0,\, x \in \mathbb{R}^n. \tag{3.9}$$

The family $(Q_t)_{t\geq 0}$ defines a semigroup with infinitesimal (non-linear) generator $-\frac{1}{2}\,|\nabla f|^2$. That is, $v = v(x,t) = Q_t f(x)$ is a solution of the Hamilton-Jacobi initial value problem (3.8) (at least almost everywhere). Actually, if in addition f is bounded, the Hopf-Lax formula $Q_t f$ is the pertinent mathematical solution of (3.8), that is its unique viscosity solution (cf. [Ev]).

Once this has been recognized, it is not difficult to try to follow Gross's idea for the Hamilton-Jacobi equation. Namely, letting now $F(t) = \|\,\mathrm{e}^{Q_t f}\|_{\lambda(t)}$, $t \geq 0$, for some function $\lambda(t)$ with $\lambda(0) = a$, $a \in \mathbb{R}$, the analogue of (3.7) reads as

$$\lambda(t)^2\, F(t)^{\lambda(t)-1} F'(t) = \lambda'(t)\,\mathrm{Ent}_\mu\big(\mathrm{e}^{\lambda(t)Q_t f}\big) - \int \frac{\lambda(t)^2}{2}\,\big|\nabla Q_t f\big|^2 \mathrm{e}^{\lambda(t)Q_t f} d\mu. \tag{3.10}$$

By the logarithmic Sobolev inequality (3.4) applied to $\mathrm{e}^{\lambda(t)Q_t f}$, $F'(t) \leq 0$ as soon as $\lambda'(t) = 1/C$, $t \geq 0$. As a result, the logarithmic Sobolev inequality (3.4) shows that, for every $t \geq 0$, every $a \in \mathbb{R}$ and every (say bounded) function f,

$$\|\,\mathrm{e}^{Q_t f}\|_{a+t/C} \leq \|\,\mathrm{e}^{f}\|_{a}. \tag{3.11}$$

Conversely, if (3.11) holds for every $t \geq 0$ and some $a \neq 0$, then the logarithmic Sobolev inequality (3.4) holds. We may thus state the following main result (cf. [B-G-L] for the detailed proof).

Theorem 3.3. *Assume that μ is absolutely continuous with respect to Lebesgue measure and that for some $C > 0$ and all smooth enough functions f on $\mathbb{R}^n$,*

$$\mathrm{Ent}_\mu(f^2) \leq 2C \int |\nabla f|^2 d\mu. \tag{3.12}$$

Then, for every *bounded measurable function f on $\mathbb{R}^n$, every $t \geq 0$ and every $a \in \mathbb{R}$,*

$$\|\,\mathrm{e}^{Q_t f}\|_{a+t/C} \leq \|\,\mathrm{e}^{f}\|_{a}\,. \tag{3.13}$$

Conversely, if (3.13) holds for some $a \neq 0$, all f's and all $t \geq 0$, then the logarithmic Sobolev inequality (3.12) holds.

An alternate proof of this result may be provided by the tool of vanishing viscosity (cf. [Ev]). We only briefly outline the principle that requires some further technical arguments. The idea is to add a small noise to the Hamilton-Jacobi equation to turn it after an exponential change of functions into the heat equation. Given a smooth function f, and $\varepsilon > 0$, denote namely by $v^\varepsilon = v^\varepsilon(x,t)$ the solution of the initial value partial differential equation

$$\begin{cases} \dfrac{\partial v^\varepsilon}{\partial t} + \dfrac{1}{2}\,|\nabla v^\varepsilon|^2 - \varepsilon \mathrm{L} v^\varepsilon = 0 & \text{in } \mathbb{R}^n \times (0,\infty), \\ v^\varepsilon = f & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases}$$

As $\varepsilon \to 0$, it is expected that v^ε approaches in a reasonable sense the solution v of (3.8) It is easy to check that $u^\varepsilon = \mathrm{e}^{-v^\varepsilon/2\varepsilon}$ is a solution of the heat equation $\frac{\partial u^\varepsilon}{\partial t} = \varepsilon \mathrm{L} u^\varepsilon$ (with initial value $\mathrm{e}^{-f/2\varepsilon}$). Therefore,

$$u^\varepsilon = P_{\varepsilon t}\big(\mathrm{e}^{-f/2\varepsilon}\big).$$

We may then apply, with some modifications (cf. [B-G-L]), classical hypercontractivity to u^ε to recover Theorem 3.3. It must be emphasized that the perturbation argument by a small noise has a clear picture in the probabilistic language of large deviations. Namely, the asymptotic of

$$v^\varepsilon = -2\varepsilon \log P_{\varepsilon t}\big(\mathrm{e}^{-f/2\varepsilon}\big)$$

as $\varepsilon \to 0$ is a Laplace-Varadhan asymptotic with rate described precisely by the infimum convolution of f with the quadratic large deviation rate function for the heat semigroup.In this limit, the second order Laplace operator is the leading term in the definition of $\mathrm{L} = \Delta - \nabla V \cdot \nabla$ so that the limiting solution u given by the infimum-convolution $Q_t f$ is independent of the potential U and thus of μ. (In particular, this asymptotic is explicit on the basic Ornstein-Uhlenbeck example.)

With respect to classical hypercontractivity, it is worthwhile noting that Q_t is defined independently of the underlying measure μ. Due to the homogeneity property $Q_t(sf) = sQ_{st}f$, $s, t > 0$, (3.13) may be rewritten equivalently as

$$\|\, \mathrm{e}^{Q_1 f} \|_{r+1/C} \leq \|\, \mathrm{e}^{f} \|_r \tag{3.14}$$

for $r \in \mathbb{R}$. If (3.14) holds for either every $r > 0$ (or only large enough) or every $r < 0$ (or only large enough), then the logarithmic Sobolev inequality (3.12) holds.

When $a = 0$ in (3.13), or $r = 0$ in (3.14), these two inequalities actually amount to the infimum-convolution inequality

$$\int \mathrm{e}^{Q_C f} d\mu \leq \mathrm{e}^{\int f d\mu} \tag{3.15}$$

holding for every bounded (or integrable) function f. Now, as we saw in Proposition 3.2, (3.15) is exactly the quadratic transportation cost inequality

$$\mathrm{W}_2(\mu, \nu) \leq \sqrt{C\, \mathrm{H}(\nu \,|\, \mu)} \tag{3.16}$$

holding for all probability measures ν absolutely continuous with respect to μ.

That the transportation cost inequality (3.16) follows from the logarithmic Sobolev inequality (3.12) was established by F. Otto and C. Villani [O-V1]. While the arguments developed in [O-V1] do involve PDE's methods, the approach presented here only relies on the basic Hamilton-Jacobi equation and presents a clear view of the connection between logarithmic Sobolev inequalities and transportation cost inequalities. One feature of the preceding approach is the systematic use of the Monge-Kantorovitch dual version of the transportation cost inequality involving infimum-convolution rather than Wasserstein distances. While optimal mass transportation is not used in the argument above, it is however implicit at various stages (see [O-V2], [Vi]). For example, optimality in (3.3)

$$\int Q_1 f d\nu - \int f d\mu = \mathrm{W}_2(\mu,\nu)^2 = \frac{1}{2}\int |x-T|^2 d\mu$$

is achieved at a function f such that

$$Q_1 f \circ T = f + \frac{1}{2}\,|x-T|^2$$

where $T = \nabla\phi$ is the Brenier-McCann optimal transport of μ to ν by the gradient of a convex function $\phi : \mathbb{R}^n \to \mathbb{R}$. Moreover, $\nabla f = T^{-1} - x$.

It is an open problem (although probably with negative answer) to know whether the critical case (3.15) (equivalently the quadratic transportation inequality (3.16)) is also equivalent to the logarithmic Sobolev inequality (3.12) (with a possibly different constant C). Partial converses are discussed in [O-V1] through a new family of inequalities, called HWI inequalities (relating entropy, Wasserstein distance and Fisher information). Extension to Riemannian manifolds are discussed in [O-V1], [B-G-L], [Wa].

3.3 Brunn-Minkowski revisited

The final step in our program is to come back to the geometric Brunn-Minskowski inequalities. Actually, it turns out that the Brunn-Minkowski theorem may be used to prove the hypercontractive inequalities of Theorem 3.3 for the class of measures with strictly logconcave densities. Assume, as in Theorem 2.2, 2.4 and 2.6, that $d\mu = \mathrm{e}^{-V}dx$ where $V : \mathbb{R}^n \to \mathbb{R}$ is smooth and such that for some $c > 0$, uniformly in $x \in \mathbb{R}^n$, $\mathrm{Hess}\,(V)(x) \geq c\,\mathrm{Id}$ in the sense of symmetric matrices. This condition is thus the condition under which both the quadratic transportation cost and logarithmic Sobolev inequalities for μ hold with $C = \frac{1}{c}$. What we observe here is that the classical Brunn-Minkowski inequality, in its functional form (Theorem 2.1), may be used to provide a simple proof of the hypercontractive estimates (3.13) (with $a = 1$), and thus of the logarithmic Sobolev inequality (and the quadratic transportation cost inequality (3.16)). Namely, given a (bounded) function f on $\mathbb{R}^n$, apply Theorem 2.1 to the functions

$$u(x) = \mathrm{e}^{\frac{1}{\theta}Q_{(1-\theta)/c\theta} f(x) - V(x)}, \quad v(y) = e^{-V(y)}, \quad w(z) = \mathrm{e}^{f(z)-V(z)}.$$

(Note the somewhat different choice in the functions u, v, w with respect to the one used in Section 2.1: the one here will lead to optimal constants.) The convexity assumption on V amounts to

$$\theta V(x) + (1-\theta)V(y) - V\big(\theta x + (1-\theta)y\big) \geq \frac{c\,\theta(1-\theta)}{2}\,|x-y|^2 \tag{3.17}$$

for every $0 < \theta < 1$ and $x, y \in \mathbb{R}^n$ so that condition (2.3) on u, v, w will be satisfied by the very definition of the infimum-convolution $Q_{(1-\theta)/c\theta} f$. Therefore,

$$\int \mathrm{e}^f d\mu \geq \left(\int \mathrm{e}^{\frac{1}{\theta} Q_{(1-\theta)/c\theta} f} d\mu \right)^\theta .$$

Setting $\frac{1}{\theta} = 1+ct$, $t \geq 0$, immediately yields (3.13) with $C = \frac{1}{c}$ and $a = 1$. In particular the logarithmic Sobolev inequality for μ holds with $C = \frac{1}{c}$. We thus recover Theorem 2.6 from the Hamilton-Jacobi approach.

Note that the same argument holds when considering an arbitrary norm in (3.17) to yield the logarithmic Sobolev inequality (3.12) with the dual norm of the gradient. This type of logarithmic Sobolev inequalities are out of reach of the heat semigroup tools.

3.4 Recent developments

Recent extensions of transportation cost and generalized entropic inequalities take a natural form in a PDE setting where they are used to control decays to equilibrium. The following comments, and references, are far from exhaustive, and we refer in particular to the recent monograph by C. Villani [Vi] and the references therein for a more complete account.

There, the reference measure is Lebesgue measure, and functions are replaced by probability densities. In particular, the heat equation $\frac{\partial}{\partial t} - \mathrm{L} = 0$ with respect to the operator $\mathrm{L} = \Delta - \nabla V \cdot \nabla$ is turned into the Fokker-Planck equation

$$\frac{\partial u}{\partial t} - \nabla \cdot \big[u \nabla (\log u + V) \big] = 0.$$

If $d\mu = \mathrm{e}^{-V} dx$ is the invariant self-adjoint probability measure for L, and if f is a non-negative function with $\int f d\mu = 1$, perform the change of functions $\rho = f\mathrm{e}^{-V}$ so that ρ is a probability density with respect to Lebesgue measure on $\mathbb{R}^n$. The logarithmic Sobolev inequality (2.14) takes the form

$$\int \rho \log \rho dx + \int V \rho dx \leq \frac{C}{2} \int \big| \nabla (\log \rho + V) \big|^2 \rho dx \tag{3.18}$$

for every smooth probability density ρ on $\mathbb{R}^n$. In this form, it may be generalized in various directions. For every probability density ρ, define the entropy of ρ as

$$\mathrm{H}(\rho) = \int U(\rho) dx + \int V \rho dx + \frac{1}{2} \int \int W(x-y) \rho(x) \rho(y) dx dy$$

where $U : \mathbb{R}^n \to \mathbb{R}_+$, $U(0) = 0$, is strictly convex and such that $\lambda \to \lambda^n U(\lambda^{-n})$ is convex and increasing, and W is convex and even. The typical examples of U are

$U(x) = x \log x$ (which corresponds to the classical case) and the power type function $U(x) = \frac{x^m}{m-1}$ corresponding to porous medium equations, while the convolution interaction W is associated with granular media. Denote by ρ_∞ the unique probability density that minimizes the functional H and set $\mathrm{H}(\rho \,|\, \rho_\infty) = \mathrm{H}(\rho) - \mathrm{H}(\rho_\infty)$.

The Fisher information on the right-hand side of (3.18) may also be extended to

$$I(\rho) = \int \Big[\nabla\big(U'(\rho) + V + W * \rho\big) \cdot \nabla c^*\big(\nabla\big(U'(\rho) + V + W * \rho\big)\big)\Big]\rho dx$$

where $c : \mathbb{R}^n \to \mathbb{R}_+$ is a cost function on $\mathbb{R}^n$ and c^* its Legendre transform. (See [Vi] for admissible costs extending the classical example of $c(x) = \frac{|x|^2}{2}$, $x \in \mathbb{R}^n$.)

In this framework, the generalized entropic inequality extending (3.18) takes the form

$$\mathrm{H}\big(\rho \,|\, \rho_\infty\big) \leq C\, I(\rho) \tag{3.19}$$

for some constant $C > 0$ and all probability densities ρ, while the transportation cost inequality generalizing (2.12) expresses that

$$\mathrm{W}_c(\rho, \rho_\infty) \leq C\, \mathrm{H}\big(\rho \,|\, \rho_\infty\big) \tag{3.20}$$

where

$$\mathrm{W}_c(\mu, \nu) = \inf \iint c(x - y) d\pi(x, y)$$

with the infimum taken over all probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals μ and ν.

These generalized entropic and transportation cost inequalities have been studied extensively in a vast recent PDE literature (some, but not all, being referenced here). In the geometric framework of dissipative evolution equations introduced by F. Otto [Ot], interpolation along mass transport and displacement convexity have been used to this task. Alternatively, along the lines of the mass transportation proofs of Section 2, both (3.19) and (3.20) have been established in [CE-G-H] under various convexity conditions on V and W (basically uniform convexity of V with respect to the cost c and convexity of W). The approaches cover in the same way various families of inequalities, including in particular HWI inequalities [O-V1], [C-MC-V], [CE-G-H], and the classical Sobolev and Gagliardo-Nirenberg inequalities with sharp constants for both Euclidean and arbitrary norms on $\mathbb{R}^n$ [CE-N-V]. Even sharper further developments have been obtained recently in [A-G-K], [Gh]... These estimates may then be used to produce exponential decays to equilibrium, in relative entropy and in the Wasserstein distance, of solutions of families of degenerate parabolic equations (Fokker-Plank, porous medium and fast diffusion, p-Laplacian), with explicit rates [Ot], [C-J-M-T-U], [A-M-T-U], [C-MC-V], [DP-D1], [DP-D2]... We refer to these works, and to the monograph [Vi], for a complete discussion, proofs and precise hypotheses in these recent results and developments.

In the dual formulation, the generalized entropic inequality (3.19) has been shown, by I. Gentil and F. Malrieu [G-M], to be equivalent, as in the classical case, to a family of hypercontractive bounds on the solutions

$$Q_t f(x) = \inf_{y \in \mathbb{R}^n} \Big[f(y) + tc\Big(\frac{x-y}{t}\Big)\Big], \quad x \in \mathbb{R}^n, \ t > 0,$$

of the Hamilton-Jacobi equation

$$\begin{cases} \dfrac{\partial v}{\partial t} + c^*(\nabla v) = 0 & \text{in } \mathbb{R}^n \times (0, \infty), \\ \hfill v = f & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases}$$

In particular, this family includes (3.20) extending to this more general framework the implication on [O-V1] from the generalized entropic inequality (3.19) to the transportation cost inequality (3.20).

The connection between classical Sobolev inequalities and ultracontractive bounds of solutions of Hamilton-Jacobi equations is developed in [Ge], in analogy with the connection with heat kernel bounds (cf. [Da], [Bak1]). See also [Bak2], [DP-D1], [DP-D2], [DP-D-G]... for various regularization properties of non-linear generalizations (porious medium, fast diffusion, p-Laplacian...) of the heat equation, with in particular in [Bak2] the extension of the semigroup proof of Theorem 2.5.

References

[A-G-K] M. Agueh, N. Ghoussoub, X. Kang. Geometric inequalities via a duality between certain quasilinear PDE's and Fokker-Plank equations (2002).

[An] C. Ané et al. Sur les inégalités de Sobolev logarithmiques. Panoramas et Synthèses, vol. 10. Soc. Math. de France (2000).

[A-M-T-U] A. Arnold, P. Markowich, G. Toscani, A. Unterreiter. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Plank type equations. Comm. Partial Differential Equations 26, 43–100 (2001).

[Bak1] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. Ecole d'Eté de Probabilités de St-Flour. Lecture Notes in Math. 1581, 1–114 (1994). Springer.

[Bak2] D. Bakry. Functional inequalities for Markov semigroups (2002).

[Bal] K. Ball. An elementary introductionto modern convex geometry. Flavors of Geometry. Math. Sci. Research Inst. Publ. 31. Cambridge Univ. Press (1997).

[Bar] F. Barthe. Autour de l'inégalité de Brunn-Minkowski (2003). Ann. Fac. Sci. Toulouse, to appear.

[Bl] G. Blower. The Gaussian isoperimetric inequality and transportation (1999). Positivity, to appear.

[B-G-L] S. Bobkov, I. Gentil, M. Ledoux. Hypercontractivity of Hamilton-Jacobi equations. J. Math. Pures Appl. 80, 669–696 (2001).

[B-G] S. Bobkov, F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. 163, 1–28 (1999).

[Br] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. Comm. Pure Appl. Math. 44, 375–417 (1991).

[Ca] L. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. Comm. Math. Phys. 214, 547–563 (2000).

[C-J-M-T-U] J. Carillo, A. Jüngel, P. Markowich, G. Toscani, A. Unterreiter. Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities. Monatsh. Math. 133, 1–82 (2001).

[C-MC-V] J. Carillo, R. McCann, C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates (2001). Rev. Math. Iberoamericana, to appear.

[CE] D. Cordero-Erausquin. Some applications of mass transport to Gaussian type inequalities. Arch. Rational Mech. Anal.161, 257–269 (2002).

[CE-G-H] D. Cordero-Erausquin, W. Gangbo, C. Houdré. Inequalities for generalized entropy and optimal transportation (2001). Procedings of the workshop: Mass transportation methods in kinetic theory and hydrodynamics, to appear.

[CE-MC-S] D. Cordero-Erausquin, R. McCann, M. Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. Invent. math. 146, 219–257 (2001).

[CE-N-V] D. Cordero-Erausquin, B. Nazaret, C. Villani. A mass-transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities (2002).

[DG] S. Das Gupta. Brunn-Minkowski inequality and its aftermath. J. Multivariate Anal. 10, 296–318 (1980).

[Da] E. B. Davies. Heat kernel and spectral theory. Cambridge Univ. Press (1989).

[DP-D1] M. Del Pino, J. Dolbeault. Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions. J. Math. Pures Appl. 81, 847–875 (2002).

[DP-D2] M. Del Pino, J. Dolbeault. Asymptotic behaviour of nonlinear diffusions (2002).

[DP-D-G] M. Del Pino, J. Dolbeault, I. Gentil. Nonlinear diffusions, hypercontractivity and the optimal L^p-Euclidean logarithmic Sobolev inequality (2002).

[Du] R. M. Dudley. Real analysis and probability. Chapman & Hall (1989).

[F-O-T] M. Fukushima, Y. Oshima, M. Takeda. Dirichlet forms and symmetric Markov processes. De Gruyter (1994).

[Ev] L. C. Evans. Partial differential equations. Amer. Math. Soc. (1997).

[Ge] I. Gentil. Ultracontractive bounds of Hamilton-Jacobi solutions. Bull. Sc. math. 126, 507–524 (2002).

[G-M] I. Gentil, F. Malrieu. Equations de Hamilton-Jacobi et inégalités entropiques généralisées. C. R. Acad. Sci. Paris 332, 437–440 (2002).

[Gh] N. Ghoussoub. An inequality for interacting gazes involving energy, transport cost, entropy and center of mass (2002).

[Grom] M. Gromov. Metric structures for Riemannian and non-Riemannian spaces. Birkhäuser (1998).

[Gros] L. Gross. Logarithmic Sobolev inequalities. Amer. J. Math. 97, 1061–1083 (1975).

[Le] M. Ledoux. The concentration of measure phenomenon. Mathematical Surveys and Monographs 89. AMS (2001).

[MC] R. J. McCann. Existence and uniqueness of monotone measure-preserving maps. Duke Math. J. 80, 309–323 (1995).

[MD] C. McDiarmid. Concentration. Probabilistic methods for algorithmic discrete mathematics, 195–248. Springer (1998).

[Mar1] K. Marton. A measure concentration inequality for contracting Markov chains. Geom. funct. anal. 6, 556-571 (1997).

[Mar2] K. Marton. Measure concentration for a class of random processes. Probab. Theory Relat. Fields 110, 427–439 (1998).

[Mau] B. Maurey. Some deviations inequalities. Geom. funct. anal. 1, 188–197 (1991).

[Mi1] V. D. Milman. New proof of the theorem of Dvoretzky on sections of convex bodies. Funct. Anal. Appl. 5, 28–37 (1971).

[Mi2] V. D. Milman. The heritage of P. Lévy in geometrical functional analysis. Colloque Paul Lévy sur les processus stochastiques. Astérisque 157-158, 273–302 (1988).

[Mi3] V. D. Milman. Dvoretzky's theorem - Thirty years later (Survey). Geom. funct. anal. 2, 455–479 (1992).

[Mi4] V. D. Milman. Topics in asymptotic geometric analysis. Proceedings of "Visions in Mathematics – Towards 2000". GAFA 2000, Special Volume, 792–815 (2000).

[M-S] V. D. Milman, G. Schechtman. Asymptotic theory of finite dimensional normed spaces. Lecture Notes in Math. 1200 (1986). Springer.

[Ot] F. Otto. The geometry of dissipative evolution equation: the porous medium equation. Comm. Part. Diff. Eq. 26, 101–174 (2001).

[O-V1] F. Otto, C. Villani. Generalization of an inequality by Talagrand, and links with the logarithmic Sobolev inequality. J. Funct. Anal. 173, 361–400 (2000).

[O-V2] F. Otto, C. Villani. Comment on: Hypercontractivity of Hamilton-Jacobi equations [J. Math. Pures Appl. 80, 669–696 (2001)] by S. Bobkov, I. Gentil, M. Ledoux. J. Math. Pures Appl. 80, 697–700 (2001).

[Pi] M. S. Pinsker. Information and information stability of random variables and processes. Holden-Day, San Franscico (1964).

[R-R] S. T. Rachev, L. Rüschendorff. Mass transportation problems, vol. 1 & 2. Springer (1998).

[Sc] G. Schechtman. Concentration, results and applications. The Handbook in the Geometry of Banach Spaces. Elsevier (2003).

[St] J. M. Steele. Probability theory and combinatorial optimization. CBMS-NSF Regional Conference Series in Applied Mathematics 69 (1996). SIAM.

[Ta1] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. Publications Mathématiques de l'I.H.E.S. 81, 73–205 (1995).

[Ta2] M. Talagrand. Transportation cost for Gaussian and other product measures. Geom. funct. anal. 6, 587–600 (1996).

[Vi] C. Villani. Topics in optimal transportation. Graduate Studies in Mathematics 58. AMS (2003).

[Wa] F.-Y. Wang. Probability distance inequalities on Riemannian manifolds and path spaces (2002).

Institut de Mathématiques, Université Paul-Sabatier, 31062 Toulouse, France
E-mail address: ledoux@math.ups-tlse.fr
http://www.lsp.ups-tlse.fr/Ledoux/