



HAL
open science

Concentration of measure principle and entropy-inequalities

Paul-Marie Samson

► **To cite this version:**

Paul-Marie Samson. Concentration of measure principle and entropy-inequalities . Eric Carlen, Mokshay Madiman, Elisabeth M. Werner. Convexity and Concentration, 161 (Chapter 3), Springer, New York, NY, 2017, The IMA Volumes in Mathematics and its Applications, 9781493970049 • 9781493970056 (online). 10.1007/978-1-4939-7005-6 . hal-01365314

HAL Id: hal-01365314

<https://hal.science/hal-01365314>

Submitted on 13 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONCENTRATION OF MEASURE PRINCIPLE AND ENTROPY-INEQUALITIES

PAUL-MARIE SAMSON

ABSTRACT. The concentration measure principle is presented in an abstract way to encompass and unify different concentration properties. We give a general overview of the links between concentration properties, transport-entropy inequalities, and logarithmic Sobolev inequalities for some specific transport costs. By giving few examples, we emphasize optimal weak transport costs as an efficient tool to establish new transport inequality and new concentration principles for discrete measures (the binomial law, the Poisson measure, the uniform law on the symmetric group).

1. INTRODUCTION

Of isoperimetric inspiration, the concentration of measure phenomenon has been pushed forward by V. Milman in the 70's in the study of the asymptotic geometry of Banach spaces and then in-depth studied by many authors including Gromov [GM83, Gro99], Talagrand [Tal95], Maurey [Mau91], Ledoux [Led97, BL97], Bobkov [Bob97, BL00]. This principle has applications in numerous fields of mathematics. The book by M. Ledoux [Led01] is devoted to this subject. It presents numerous examples and probabilistic, analytical and geometrical technics related to this notion. We also refer to the monographs [BLM13, Mas07] for more applications of this principle in statistics and probability theory. We also warmly recommend the surveys [GL10, Goz15] by Gozlan and Léonard about transport-entropy inequalities. The main purpose of this paper is to complement these surveys in view of the recent developments.

In the present paper, the concentration of measure principle is formalized in an abstract way to encompass and unify different concentration properties investigated in the literature. The definition of this principle with enlargements of sets takes its origin from the papers by M. Talagrand [Tal95, Tal96b, Tal96a]. We propose a functional formulation of the concentration principle, rigorously introduced in [GRST14b]. We emphasize three types of cost functions that provide most of the enlargements of sets considered in the literature, the usual cost functions, the barycentric cost functions and the universal cost functions.

Date: September 13, 2016.

1991 Mathematics Subject Classification. 60E15, 32F32, 39B62, 26D10.

Key words and phrases. Concentration of measure, Duality, Transport inequalities, logarithmic-Sobolev inequalities.

Supported by the grants ANR 2011 BS01 007 01, ANR 10 LABX-58.

The concentration properties associated to usual cost functions and its related functional inequalities have been widely studied these last years. Now, it is a challenge to develop new concentration inequalities that could capture precise dimensional concentration behavior for particular classes of functions, especially in discrete setting. In this document, we present some concentration results for discrete measures associated to the above weak transport costs. In the spirit of the early works by Talagrand [Tal95, Tal96b, Tal96a], we believe that entropy-functional inequalities associated to new weak transport costs could be adapted to understand some concentration challenging problems.

The third section of this paper put forward the transport-entropy inequality (also called transport inequality) associated to the above different costs functions, as a fundamental tool in the study of concentration properties in product spaces. This entropy-inequality is an alternative to the logarithmic Sobolev inequality and its variants, to establish concentration properties in product spaces. The main feature of these two inequalities is for each, a tensorization property, that provides concentration results in high dimension spaces. We will briefly recall the links between these two kinds of entropy-inequalities and the concentration of measure principle.

Section 4 is focussed on the concentration properties and transport inequalities related to the so-called barycentric costs. These transport costs, weaker than the usual one, are adapted to derive new transport inequalities for discrete measures (see Section 4.2). Indeed, let us recall that the Talagrand’s transport inequality \mathbf{T}_2 is never satisfied by discrete measures. Recently, in the context of the study of curvature notion in discrete spaces, other transport inequalities have been proposed, mainly in the works by Erbar-Maas [Maa11, EM12]. However, due to the very abstract definition of the optimal transport costs, the associated concentration of measure phenomenon remains difficult to interpret.

Barycentric optimal transport costs can be expressed using optimal transport costs by considering the notion of convex order on probability measures (see Proposition 4.1). Moreover, most of the results with usual transport costs can be adapted for barycentric transport costs. First, barycentric transport inequalities are equivalent to logarithmic Sobolev inequalities restricted to a class of convex or concave functions (see Section 4.1). In an other direction, on the real line, as for usual transport costs, for any optimal barycentric transport cost, there exists an optimal coupling which is independent of the convex function involved in the barycentric cost (see Section 4.3). This independence property allows to characterize the probability measures satisfying a barycentric transport inequality on the real line. As a byproduct of this characterization, the “convex” Poincaré inequality on the real line is equivalent to a barycentric transport inequality with some specific convex cost function (see Section 4.4).

Section 5 is devoted to examples of universal transport inequalities such as the so-called Csizár-Kullback-Pinsker inequality. The most emblematic universal transport inequality of this document is the Marton’s transport inequality with its weak cost $\tilde{\mathcal{T}}_2$ [Mar96b]. In the papers [Sam03, Sam07], the Marton’s cost has been improved to reach optimal Bernstein bounds for suprema of empirical bounded independent processes (see Section 5.1). This method is an alternative to the so-called

Herbst's method, first used by Ledoux to get deviation bounds for suprema of empirical processes [Led97]. In Section 5.2, we recall the results of [Sam00] and [Pau14] that extends Marton's inequality to any measure on a product space, with weak dependences of its marginals.

The last Section 5.3 concerns recent transport inequalities obtained for the uniform probability measure on the symmetric group. These inequalities are obtained from the Csizár-Kullback-Pinsker inequality or the Marton's inequality $\widetilde{\mathbf{T}}_2$, by using other tensorization arguments. The proofs are inspired by the work by Talagrand on the symmetric group [Tal95].

The works presented in this survey could be extended in different directions.

A first challenge is to define other costs functions that may capture new concentration's properties, as for the uniform measure on the symmetric group, for Gibbs measure or for other non-product measures under dependence properties. The new costs functions presented in this survey are of particular interest in discrete setting (discrete cube, binomial law, Poisson measure) and we may use it in other discrete framework such as Poisson processes.

Another direction is to develop the multimarginal transport inequalities in discrete and continuous setting. We wonder whether this multimarginal approach allows to reach superconcentration-properties. In this field, the works by Dembo [Dem97] and Talagrand [Tal96a] are also a guideline.

These last years, the concept of curvature in discrete setting has emerged [Oll09, OV12, EM12, EMT15, Hil14] by analogy of the concept of lower bounded curvature in continuous setting in metric spaces [Vil09, AGS14, AGS15]. These notions could be revisited by relating the notion of curvatures to different cost functions, the paper [GRST14a] is a first attempt in that direction.

Acknowledgements. The author would like to thank Nathael Gozlan and an anonymous referee for their careful reading of the manuscript.

CONTENTS

1. Introduction	1
2. A general concentration of measure principle	4
The usual cost functions	5
The universal cost functions	5
The barycentric cost functions	6
3. Transport-entropy inequalities	8
3.1. Transport inequalities and concentration properties	10
3.2. Functional formulation of transport inequality, the dual Kantorovich Theorem	13
3.3. Tensorization - Characterization by dimension-free concentration properties	15

3.4. Connections with logarithmic Sobolev inequalities	17
4. Some results around “barycentric” costs	20
4.1. Barycentric transport inequality and logarithmic Sobolev inequalities	22
4.2. Barycentric transport inequalities for the binomial law and the Poisson measure	23
4.3. Optimal transport coupling for barycentric costs on \mathbb{R}	25
4.4. Characterization of probability measures satisfying a barycentric transport inequality on \mathbb{R}	26
5. Universal transport inequalities	27
5.1. Bernstein’s type of deviation inequalities for suprema of independent empirical processes	29
5.2. Marton’s transport inequality for weakly dependent random variables	32
5.3. Transport inequalities for the uniform law on the symmetric group.	34
References	38

2. A GENERAL CONCENTRATION OF MEASURE PRINCIPLE

Let (\mathcal{X}, d) be a Polish metric space (separable, complete), with Borel σ -field $\mathcal{B}(\mathcal{X})$. We denote by $\mathcal{P}(A)$ the set of probability measures on a subset $A \in \mathcal{B}(\mathcal{X})$, and for $q \geq 1$, we denote by $\mathcal{P}_q(A)$ the set of probability measures p on A such that $\int d(x_0, y)^q dp(y) < +\infty$ for a point x_0 .

We assume that a notion of pseudo-distance from a point $x \in \mathcal{X}$ to a subset $A \subset \mathcal{X}$ is given, denoted by $c(x, A) \in [0, +\infty]$, and such that $c(x, A) \geq 0$ and $c(x, A) = 0$ if $x \in A$. The usual example is

$$c(x, A) = d(x, A) = \inf_{y \in A} d(x, y).$$

One could also choose $c(x, A) = \alpha(d(x, A))$, where $\alpha : \mathbb{R}^+ \rightarrow [0, +\infty]$ is such that $\alpha(0) = 0$.

For $r \geq 0$, the enlargement of A associated to this pseudo-distance is defined by

$$A_{r,c} = \{x \in \mathcal{X}, c(x, A) \leq r\}.$$

Definition 2.1. Let $\beta : \mathbb{R}^+ \rightarrow [0, +\infty]$ be an increasing function with $\beta(0) = 0$. A probability measure μ satisfies a concentration principle with profile β and cost c , if there exist $a_1, a_2 > 0$ such that for all subsets $A \in \mathcal{B}(\mathcal{X})$ and for all $r \geq 0$,

$$\mu(A)^{a_1} \mu(\mathcal{X} \setminus A_{r,c})^{a_2} \leq e^{-\beta(r)}.$$

If μ satisfies a concentration principle with profile β , then for all $A \in \mathcal{B}(\mathcal{X})$ with measure $\mu(A) \geq 1/2$, one has

$$\mu(\mathcal{X} \setminus A_{r,c}) \leq 2^{a_1/a_2} e^{-\beta(r)/a_2} = e^{-\tilde{\beta}(r)}, \quad \forall r \geq 0.$$

This last property is the classical formulation of a concentration of measure principle. Actually, these two formulations are equivalent, up to constants, as soon as

$$A \subset \mathcal{X} \setminus (\mathcal{X} \setminus A_{r,c})_r, \quad \forall r \geq 0, \forall A \in \mathcal{B}(\mathcal{X}),$$

(see Lemma 5.6 [GRST14b]). This inclusion depends of the kind of enlargement and is not satisfied for some enlargements.

In all this document, the pseudo-distance $c(x, A)$ is defined from a *cost function*.

Definition 2.2. *A cost function is a function*

$$c : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \rightarrow [0, +\infty],$$

such that the function $p \mapsto c(x, p)$ is convex, and $c(x, \delta_x) = 0$ for all $x \in \mathcal{X}$ (δ_x denotes the Dirac measure at point x). Then we define the pseudo-distance c by

$$c(x, A) = \inf_{p \in \mathcal{P}(A)} c(x, p), \quad x \in \mathcal{X}, \quad A \in \mathcal{B}(\mathcal{X}).$$

Here, for the sake of simplicity, the same notation c is used for the pseudo-distance and the cost function c .

In this document, we consider three kinds of cost functions with following definitions.

The usual cost functions. Usually, the transport cost is a function defined on $\mathcal{X} \times \mathcal{X}$ rather than $\mathcal{X} \times \mathcal{P}(\mathcal{X})$.

Definition 2.3. *A cost function $c : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \rightarrow [0, +\infty]$ is called usual if there exists a measurable function $\omega : \mathcal{X} \times \mathcal{X} \rightarrow [0, +\infty]$ such that for all $x \in \mathcal{X}$ and all $p \in \mathcal{P}(\mathcal{X})$, $\omega(x, x) = 0$ and*

$$c(x, p) = \int \omega(x, y) dp(y).$$

In that case, $p \mapsto c(x, p)$ is an affine function. The pseudo-distance $\inf_{p \in \mathcal{P}(A)} c(x, p)$ is reached at Dirac measures, the extremal points of the convex set $\mathcal{P}(A)$, therefore

$$c(x, A) = \inf_{y \in A} \omega(x, y).$$

By the way, the pseudo-distance is exactly the classical one associated to the cost ω . The most studied cost functions ω are $\omega(x, y) = d(x, y)^q$, $q > 0$ or $\omega(x, y) = \alpha(d(x, y))$, with $\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$.

In the case $\omega(x, y) = d(x, y)$ one has $c(x, A) = d(x, A)$ and we simply denote by A_r the enlargement $A_{r,c}$.

The universal cost functions. These cost functions have been introduced by Talagrand [Tal96b] and Marton [Mar96b] in order to solve different types of concentration's problems, for example the deviations of suprema of empirical processes, of the largest increasing subsequence, the bin-packing problem, etc ... These transport costs are the main tools of the so-called "convex hull" method of [Tal96b, Tal96a] and in the papers [Mar97, Sam00, Sam03, Sam07, Pau14].

We denote by $\mathbb{1}_{x \neq y}$ the Hamming distance between two points x and y in \mathcal{X} , defined by $\mathbb{1}_{x \neq y} = \gamma(d(x, y))$, where $\gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is the function defined by $\gamma(h) = 1$ if $h \neq 0$ and $\gamma(0) = 0$.

Definition 2.4. Let $\alpha : \mathbb{R}^+ \rightarrow [0, +\infty]$ be a lower semi-continuous convex function and let μ_0 be a probability measure on \mathcal{X} . We define two classes of universal cost functions.

- A cost function $c : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \rightarrow [0, +\infty]$ is called universal and associated to the function α , if

$$c(x, p) = \alpha \left(\int \mathbb{1}_{x \neq y} dp(y) \right),$$

for all $x \in X$ and for all $p \in \mathcal{P}(\mathcal{X})$.

- A cost function $c : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \rightarrow [0, +\infty]$ is called universal, associated to the function α and to the measure μ_0 , if

$$c(x, p) = \int \alpha \left(\mathbb{1}_{x \neq y} \frac{dp}{d\mu_0}(y) \right) d\mu_0(y),$$

for all $(x, p) \in \mathcal{X} \times \mathcal{P}(\mathcal{X})$ such that p is absolutely continuous with respect to μ_0 on the set $X \setminus \{x\}$, and $c(x, p) = +\infty$ otherwise.

These universal cost functions are independent of the distance d on \mathcal{X} , and therefore of the geometry of the space \mathcal{X} .

The barycentric cost functions. The so-called barycentric costs are defined on $\mathcal{X} = \mathbb{R}^n$ equipped with the Euclidean distance, $d(x, y) = |x - y|$, $x, y \in \mathbb{R}^n$. They have been introduced in the paper [GRST14b] to reach optimal concentration properties for discrete measures (see Section 4.2). As explained in [GRST14b], they are also related to the convex (τ)-property by Maurey [Mau91].

Definition 2.5. A cost function $c : \mathcal{X} \times \mathcal{P}_1(\mathcal{X}) \rightarrow [0, +\infty]$ is called barycentric if there exists a lower semi-continuous convex function $\theta : \mathbb{R}^n \rightarrow [0, +\infty]$ such that for all $x \in \mathbb{R}^n$ and all $p \in \mathcal{P}_1(\mathbb{R}^n)$,

$$c(x, p) = \theta \left(x - \int y dp(y) \right).$$

Let us observe that the concentration property associated to this cost function is weaker than the one associated to the usual cost function with $\omega(x, y) = \theta(x - y)$, $x, y \in \mathcal{X}$, since by Jensen's inequality

$$c(x, p) \leq \int \theta(x - y) dp(y).$$

A functional formulation of the concentration principle of Definition 2.1 is presented in [GRST14b]. This second definition is associated to the following type of infimum-convolution operator, introduced in [Sam07, GRST14b]: for any measurable function $\varphi : \mathcal{X} \rightarrow \mathbb{R} \cup \{\infty\}$ bounded from below

$$(1) \quad R_c \varphi(x) = \inf_{p \in \mathcal{P}(\mathcal{X})} \left\{ \int \varphi dp + c(x, p) \right\}, \quad x \in \mathcal{X}.$$

Since $c(x, \delta_x) = 0$, one has $R_c \varphi(x) \leq \varphi(x)$.

For a usual cost function, $c(x, p) = \int \omega(x, y) dp(y)$, since the function $p \mapsto \int \varphi dp + c(x, p)$ is affine, the operator $R_c \varphi$ is the classical infimum-convolution operator associated to the cost function ω ,

$$R_c \varphi(x) = \inf_{y \in \mathcal{X}} \{\varphi(y) + \omega(x, y)\} = Q_\omega \varphi(x).$$

The functional formulation of the concentration principle is given by the following result.

Proposition 2.1. *Let $a_1, a_2 > 0$ and $\beta : \mathbb{R}^+ \rightarrow [0, +\infty]$ be a function. The following properties are equivalent.*

(i) *For all $A \in \mathcal{B}(\mathcal{X})$, and all $r \geq 0$,*

$$\mu(A)^{a_1} \mu(\mathcal{X} \setminus A_{r,c})^{a_2} \leq e^{-\beta(r)}.$$

(ii) *For all measurable functions $\varphi : \mathcal{X} \rightarrow \mathbb{R} \cup \{\infty\}$ bounded from below,*

$$\mu(\varphi \leq m)^{a_1} \mu(R_c \varphi > m + r)^{a_2} \leq e^{-\beta(r)} \quad \forall m \in \mathbb{R}, \forall r \geq 0.$$

Proof. Given $A \in \mathcal{B}(\mathcal{X})$, let i_A be the zero function on A and equal to $+\infty$ on $\mathcal{X} \setminus A$. By applying (ii) with the function $\varphi = i_A$ and with $m = 0$, we get (i) since $\{\varphi \leq 0\} = A$ and $R_c i_A(x) = c(x, A)$, $x \in \mathcal{X}$.

Conversely, given a function φ , we apply (1) with $A = \{\varphi \leq m\}$. Then, (2) follows from the fact that $\{R_c \varphi > m + r\} \subset (\mathcal{X} \setminus A_{r,c})$. Indeed, if $x \in A_{r,c}$, then for all $\varepsilon > 0$, there exists $p_\varepsilon \in \mathcal{P}(A)$ such that $c(x, p_\varepsilon) \leq r + \varepsilon$. Since

$$R_c \varphi(x) \leq \int \varphi dp_\varepsilon + c(x, p_\varepsilon) \leq m + r + \varepsilon,$$

when ε goes to 0, we get $x \in \{R_c \varphi \leq m + r\}$. As a consequence $A_{r,c} \subset \{R_c \varphi \leq m + r\}$ and $\mu(R_c \varphi > m + r) \leq \mu(\mathcal{X} \setminus A_{r,c})$. \square

For a usual cost function of type $c(x, p) = \int \alpha(d(x, y)) dp(y)$, $x \in \mathcal{X}$, $p \in \mathcal{P}(\mathcal{X})$, where $\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is one-to-one, the common way to write the concentration principle is to use the classical enlargement $A_t = \{x \in \mathcal{X}, d(x, A) \leq t\}$. Since $A_{r,c} = A_t$ for $r = \alpha(t)$, (i) can be rewritten as follows: for all $A \in \mathcal{B}(\mathcal{X})$ and all $t \geq 0$,

$$\mu(A)^{a_1} \mu(\mathcal{X} \setminus A_t)^{a_2} \leq e^{-\beta(\alpha(t))}.$$

In that case, going back to early P. Lévy's ideas, we may formalize the concentration property by using the class of 1-Lipschitz functions f (cf. [Led01]). As in the previous proof, by choosing $A = \{f \leq m\}$, $m \in \mathbb{R}$, we show that $A_t \subset \{f \leq m + t\}$, $t \geq 0$. This provides the following equivalent functional formulation : for all 1-Lipschitz functions $f : \mathcal{X} \rightarrow \mathbb{R}$,

$$(2) \quad \mu(f \leq m)^{a_1} \mu(f > m + t)^{a_2} \leq e^{-\beta(\alpha(t))}, \quad \forall m \in \mathbb{R}, \forall t \geq 0.$$

Let us assume moreover that α is convex. The inequality (2) can be also derived from (ii) applied to the fonction $\varphi = \lambda f$, $\lambda > 0$, assuming first that f is bounded

from below. Since f is 1-Lipschitz, $R_c\varphi$ is close to φ and its closeness is controlled by λ . More precisely, one has for any $x \in \mathcal{X}$

$$\begin{aligned} R_c\varphi(x) &= \inf_{y \in \mathcal{X}} \{\varphi(y) + \alpha(d(x, y))\} \\ &\geq \varphi(x) - \sup_{y \in \mathcal{X}} \{\lambda d(x, y) - \alpha(d(x, y))\} \geq \varphi(x) - \alpha^*(\lambda), \end{aligned}$$

with $\alpha^*(\lambda) = \sup_{v \geq 0} \{\lambda v - \alpha(v)\}$. Therefore, by replacing m by λm , (ii) provides: for all 1-Lipschitz function f , bounded from below, and for all $\lambda \geq 0$,

$$\mu(f \leq m)^{a_1} \mu(\lambda f > \lambda m + r + \alpha^*(\lambda))^{a_2} \leq e^{-\beta(r)}, \quad \forall m \in \mathbb{R}, \forall r \geq 0.$$

Since $\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is one to one convex and $\alpha(0) = 0$, α is increasing and for all $t > 0$, $\partial\alpha(t) \subset (0, +\infty)$, where $\partial\alpha(t)$ denotes the subdifferential of α at point t . As a consequence, the last inequality implies property (2) for all 1-Lipschitz functions bounded from below by choosing by choosing $r = \alpha(t)$ and $\lambda \in \partial\alpha(t)$ such that $\lambda t = \alpha(t) + \alpha^*(\lambda)$. Then, by monotone convergence, the property (2) extends to all 1-Lipschitz functions.

Finally, the property (2) implies the classical concentration property for 1-Lipschitz functions f around their median m_f (see [Led01], Chapter 1). By applying (2) to f or to $-f$, and by choosing $m = m_f$ or $m = -m_f$, we get

$$\mu(|f - m_f| > t) \leq 2.2^{a_1/a_2} e^{-\beta(\alpha(t))/a_2}, \quad \forall t \geq 0.$$

3. TRANSPORT-ENTROPY INEQUALITIES

This section emphasizes the transport-entropy inequalities in the study of concentration of measure phenomenon on product spaces. As for the logarithmic Sobolev inequalities, the tensorization properties of the transport inequalities make it an effective tool to prove concentration properties in product spaces. The last part of this section briefly recalls the links between the transport-entropy inequalities and the logarithmic Sobolev inequalities.

Let us first recall the original links between isoperimetric properties and concentration of measure properties. Let μ be a measure on a metric space (\mathcal{X}, d) . For any Borel set A , the *surface measure* of A is defined by

$$\mu^+(\partial A) = \liminf_{t \rightarrow 0^+} \frac{\mu(A_t) - \mu(A)}{t}.$$

The isoperimetric problem is to determine the smaller surface measure $\mu^+(\partial A)$ among all Borel set A of fixed measure $\mu(A)$. Namely, we want to find the largest function, denoted by $I_\mu : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that for all $A \in \mathcal{B}(\mathcal{X})$,

$$(3) \quad \mu^+(\partial A) \geq I_\mu(\mu(A)).$$

The function I_μ is called *isoperimetric profile* of the measure μ .

If $I_\mu \geq v' \circ v^{-1}$, where $v : \mathbb{R} \rightarrow [0, \mu(X)]$ is an increasing smooth function, the *isoperimetric inequality* (3) provides a lower estimate of the measure of A_t (see Proposition 2.1 [Led01]): for all $t \geq 0$,

$$(4) \quad \mu(A_t) \geq v(v^{-1}(\mu(A)) + t).$$

Therefore, if μ is a probability measure, we get the following concentration property, for all $A \in \mathcal{B}(\mathcal{X})$ with $\mu(A) \geq 1/2$,

$$\mu(\mathcal{X} \setminus A_t) \leq 1 - v(v^{-1}(1/2) + t), \quad \forall t \geq 0.$$

When $(X, d) = (\mathbb{R}^n, |\cdot|)$ is the Euclidean space and μ is the Lebesgue measure, or when $X = \mathbb{S}^n$ is the unit Euclidean sphere of \mathbb{R}^{n+1} with its geodesic distance d and $\mu = \sigma^n$ is the uniform law on \mathbb{S}^n , the isoperimetric profile is given by $I_\mu = v' \circ v^{-1}$ where for all $r \geq 0$, $v(r)$ is the measure of a ball of radius r (see [Lév51]). As a consequence, if $\mu(A) = \mu(B) = v(r)$ where B is a ball of \mathcal{X} , then

$$v(v^{-1}(\mu(A)) + t) = v(r + t) = \mu(B_t).$$

Therefore the property (4) implies that the balls are extremal sets with following meaning: for all Borel sets A and for all balls B with measure $\mu(B) = \mu(A)$,

$$\mu(A_t) \geq \mu(B_t),$$

for all $t \geq 0$. Actually, this property is equivalent to the isoperimetric inequality since it implies,

$$\mu^+(\partial A) = \liminf_{t \rightarrow 0^+} \frac{\mu(A_t) - \mu(A)}{t} \geq \liminf_{t \rightarrow 0^+} \frac{\mu(B_t) - \mu(B)}{t} = I_\mu(\mu(A)),$$

by using $\mu(A) = \mu(B)$.

In this way, the concentration profile of σ^n is given by the estimate of the measures of the spherical balls: for all $A \in \mathcal{B}(\mathbb{S}^n)$ with $\sigma^n(A) \geq 1/2$,

$$(5) \quad \sigma^n(\mathcal{X} \setminus A_t) \leq 1 - \sigma^n(B_t) = v(v^{-1}(1/2) + t) \leq e^{-(n-1)t^2/2}, \quad \forall t \geq 0,$$

where B is a half-sphere, $\sigma^n(B) = 1/2$. The proof of the estimate given by the last inequality is given after Corollary 2.2 in [Led01].

By volume expansion, since the uniform law of the sphere of \mathbb{R}^{n+1} of radius \sqrt{n} goes to the canonical Gaussian measure on \mathbb{R}^n , we get the isoperimetric profile of the Gaussian measure (cf. [Led93, Led01]). Half hyperplanes are extremal sets for the standard Gaussian measure γ^n on \mathbb{R}^n and one has

$$I_{\gamma^n}(s) = \varphi \circ \Phi^{-1}(s) \underset{0^+}{\sim} s \sqrt{2 \log \frac{1}{s}},$$

with $\Phi(r) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^r e^{-u^2/2} du$, $r \in \mathbb{R}$, and $\varphi = \Phi'$, Φ is the cumulative distribution function of the standard Gaussian law on \mathbb{R} . It provides the following concentration property: for all $A \in \mathcal{B}(\mathbb{R}^n)$ such that $\gamma^n(A) \geq 1/2$,

$$(6) \quad \gamma^n(\mathbb{R}^n \setminus A_t) \leq 1 - \Phi(t) \leq e^{-t^2/2}, \quad t \geq 0.$$

On the discrete cube, $\mathcal{X} = \{0, 1\}^n$, equipped with the uniform probability measure μ^n and the Hamming distance defined by

$$d(x, y) = \sum_{i=1}^n \mathbb{1}_{x_i \neq y_i}, \quad x, y \in \{0, 1\}^n,$$

the extremal sets A minimizing $\mu^n(A_t)$ for $\mu^n(A) \geq 1/2$, have been identified (see [Har66, WW77]). This provides the following concentration property: for any

subset $A \subset \{0, 1\}^n$ such that $\mu^n(A) \geq 1/2$,

$$(7) \quad \mu^n(\mathcal{X} \setminus A_t) \leq e^{-2t^2/n}, \quad t \geq 0.$$

In the last three basic examples, the sphere (5), the Gaussian space (6) and the discrete cube (7), we observe that in high dimension, $n \gg 1$, dimension is a crucial parameter that quantifies the measure concentration phenomenon.

In high dimension, isoperimetric problems are often hard to establish and few of them are solved. Therefore, we need other methods to prove concentration properties. Moreover, it is well known that the concentration property do not tensorize properly. For that purpose, “entropic” methods are efficient alternative tools. They enable to enlarge considerably the class of examples of concentration properties on high dimensional spaces, thanks to the tensorization properties of the entropy.

For concentration, the main two useful entropic methods are the one associated to the logarithmic Sobolev inequality with the so-called “Herbst’s argument” (cf. chapter 5. [Led01]), and the one based on transport inequalities with the so-called “Marton’s argument” (cf. chapter 6. [Led01]). This paper mainly concerns this second one. For example, the concentration result (6) for the Gaussian measure on \mathbb{R}^n , or (7) for the uniform law on the discrete cube, are easy consequences of the tensorization property of the transport inequalities.

3.1. Transport inequalities and concentration properties. Let $\Pi(\mu, \nu)$ denote the set of probability measures on the product space $\mathcal{X} \times \mathcal{X}$, with first marginal μ and second marginal ν . The probability space $\mathcal{P}(\mathcal{X})$ is endowed with the σ -field generated by the applications

$$\begin{array}{ccc} \mathcal{P}(\mathcal{X}) & \rightarrow & ([0, 1], \mathcal{B}) \\ \nu & \mapsto & \nu(A), \end{array}$$

where A is any Borel set of \mathcal{X} and \mathcal{B} is the Borel σ -field on $[0, 1]$.

Since \mathcal{X} is a Polish space, any measure $\pi \in \Pi(\mu, \nu)$ can be decomposed as follows

$$d\pi(x, y) = d\mu(x)dp_x(y),$$

where $p : x \in \mathcal{X} \mapsto p_x \in \mathcal{P}(\mathcal{X})$ is a measurable map μ -almost-surely uniquely determined; p is a probability kernel satisfying

$$\mu p(A) = \int p_x(A)d\mu(x) = \nu(A), \quad \forall A \in \mathcal{B}(\mathcal{X}).$$

The cost function c defines the following optimal transport cost T_c , introduced in [GRST14b].

Definition 3.1. *The optimal transport cost between two probability measures μ and ν on \mathcal{X} , associated to $c : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \rightarrow [0, +\infty]$, is the quantity*

$$T_c(\nu|\mu) = \inf_{\pi \in \Pi(\mu, \nu)} \int c(x, p_x)d\mu(x),$$

where, given $\pi \in \Pi(\mu, \nu)$, the kernel $p = (p_x)_{x \in \mathcal{X}}$ is such that $d\pi(x, y) = d\mu(x)dp_x(y)$.

At first sight, $T_c(\nu|\mu)$ is not a symmetric quantity of μ and ν . As example, for a usual cost,

$$c(x, p_x) = \int \omega(x, y) dp_x(y), \quad x \in \mathcal{X},$$

the optimal cost $T_c(\nu|\mu)$ corresponds to the usual optimal transport cost linked to the cost function ω ,

$$\mathcal{T}_\omega(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \iint \omega(x, y) d\pi(x, y) = \inf_{\pi \in \Pi(\mu, \nu)} \int c(x, p_x) d\mu(x) = T_c(\nu|\mu).$$

If the function ω is symmetric, $\omega(x, y) = \omega(y, x)$ for all $x, y \in \mathcal{X}$, then \mathcal{T}_ω and therefore T_c is symmetric, $\mathcal{T}_\omega(\mu, \nu) = \mathcal{T}_\omega(\nu, \mu)$.

Let us present transport inequalities associated to the optimal transport costs $T_c(\nu|\mu)$. We emphasize a general version that exactly provides the concentration of measure property of Definition 2.1 by the Marton's argument.

Definition 3.2. *Let $a_1, a_2 > 0$ and let $\beta : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be a non-decreasing function. The probability measure $\mu \in \mathcal{P}(\mathcal{X})$ satisfies the transport inequality $\mathbf{T}_{c, \beta}(a_1, a_2)$ if*

$$\mathbf{T}_{c, \beta}(a_1, a_2) : \quad \beta(T_c(\nu_1|\nu_2)) \leq a_1 H(\nu_1|\mu) + a_2 H(\nu_2|\mu), \quad \forall \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}),$$

where $H(\nu_1|\mu)$ is the relative entropy of ν_1 with respect to μ defined by

$$H(\nu_1|\mu) = \int \log \frac{d\nu_1}{d\mu} d\nu_1,$$

if ν_1 is absolutely continuous with respect to μ ($\nu_1 \ll \mu$), and $H(\nu_1|\mu) = +\infty$ otherwise.

In most cases, the inequality $\mathbf{T}_{c, \beta}(a_1, a_2)$ is called weak transport inequality, and for some particular costs $c : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \rightarrow [0, +\infty]$, with inequality is called barycentric transport inequality (see Section 4), or universal transport inequality (see Section 5).

Generally, the transport inequalities $\mathbf{T}_{c, \beta}(0, a_2)$ or $\mathbf{T}_{c, \beta}(a_1, 0)$ do not make sense. Indeed, they should imply $\beta(T_c(\nu_1|\mu)) = 0, \forall \nu_1 \in \mathcal{P}(\mathcal{X})$, or even $\beta(T_c(\mu|\nu_2)) = 0, \forall \nu_2 \in \mathcal{P}(\mathcal{X})$, which are never satisfied, except in degenerated cases (for example $c = 0$ or $\beta = 0$). However, with the convention $0 \cdot \infty = 0$, the transport inequality $\mathbf{T}_{c, \beta}(b, \infty)$ corresponds to the common transport inequality

$$\mathbf{T}_{c, \beta}^+(b) : \quad \beta(T_c(\nu|\mu)) \leq bH(\nu|\mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}),$$

and $\mathbf{T}_{c, \beta}(\infty, b)$ corresponds to the common transport inequality

$$\mathbf{T}_{c, \beta}^-(b) : \quad \beta(T_c(\mu|\nu)) \leq bH(\nu|\mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}).$$

These two inequalities are identical for symmetric optimal transport costs. When β is the identity, we simply denote by $\mathbf{T}_c(a_1, a_2)$, $\mathbf{T}_c^+(b)$ and $\mathbf{T}_c^-(b)$ the last transport inequalities.

Let us recall the Marton's argument. Given $A \in \mathcal{B}(\mathcal{X})$, if ν_1 is the renormalized restriction of μ to A and ν_2 the renormalized restriction of μ to $B = \mathcal{X} \setminus A_{r, c}$, $r \geq 0$,

$$\frac{d\nu_1}{d\mu} = \frac{\mathbb{1}_A}{\mu(A)} \quad \text{and} \quad \frac{d\nu_2}{d\mu} = \frac{\mathbb{1}_B}{\mu(B)},$$

then $H(\nu_1|\mu) = -\log \mu(A)$, $H(\nu_2|\mu) = -\log \mu(B)$, and $T_c(\nu_1|\nu_2) \geq r$ (since for all $x \in \mathcal{X} \setminus A_{r,c}$ and all $p \in \mathcal{P}(A)$, $c(x,p) \geq r$). Consequently, since β is non-decreasing, the transport inequality $\mathbf{T}_{c,\beta}(a_1, a_2)$ provides

$$\beta(r) \leq \log(\mu(A)^{-a_1}) + \log(\mu(B)^{-a_2}), \quad \forall r \geq 0.$$

which is the concentration property of Definition 2.1.

For a better comprehension, let us illustrate Definition 3.2 by few examples of transport inequalities.

If c is the usual cost function $c(x,p) = \int d(x,y)^q p(dy)$, $q \geq 1$, then $T_c(\nu|\mu) = T_c(\mu|\nu)$ is associated to the Wasserstein distance W_q of order q ,

$$T_c(\nu|\mu) = W_q^q(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \iint d(x,y)^q \pi(dx, dy).$$

The transport inequality $\mathbf{T}_2(b)$, first considered by Talagrand [Tal96c], and satisfied by the standard Gaussian measure $\mu = \gamma^n$ on \mathbb{R}^n for $b = 2$, corresponds to the transport inequalities $\mathbf{T}_c^+(b)$ or $\mathbf{T}_c^-(b)$ with $q = 2$,

$$\mathbf{T}_2(b) : \quad W_2^2(\mu, \nu) \leq bH(\nu|\mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}).$$

A special feature of the inequality $\mathbf{T}_2(b)$ is its equivalence to the family of transport inequalities $\mathbf{T}_c(b/t, b/(1-t))$, for $t \in (0, 1)$. Indeed, if μ satisfies $\mathbf{T}_2(b)$, then by the triangular inequality for the Wasserstein metric,

$$W_2^2(\nu_1, \nu_2) \leq (W_2(\nu_1, \mu) + W_2(\mu, \nu_2))^2 \leq b \left(\sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)} \right)^2.$$

From the identity $(\sqrt{u} + \sqrt{v})^2 = \inf_{t \in (0,1)} \left\{ \frac{u}{t} + \frac{v}{1-t} \right\}$, we get that for all $t \in (0, 1)$, μ satisfies $T_c(b/t, b/(1-t))$,

$$W_2^2(\nu_1, \nu_2) \leq \frac{b}{t} H(\nu_1|\mu) + \frac{b}{1-t} H(\nu_2|\mu), \quad \forall \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}).$$

Conversely if μ verifies $T_c(b/t, b/(1-t))$ for all $t \in (0, 1)$, then by choosing $\nu_2 = \mu$ and then when t goes to 1, we recover the transport inequality $\mathbf{T}_2(b)$.

More generally, assume that c is a usual cost of type $c(x,p) = \int \alpha(d(x,y))p(dy)$, where $\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}$ is a convex function. In that case we note $\mathcal{T}_\alpha(\mu, \nu) = T_c(\nu|\mu) = T_c(\mu|\nu)$. If moreover α is increasing, $\alpha(0) = \alpha'(0) = 0$ and α satisfies the following Δ_2 -condition, [RR91]: there exists a positive constant C such that

$$(8) \quad \alpha(2h) \leq C\alpha(h), \quad \forall h \geq 0,$$

then, we may use the following change of metric Lemma given in [GRS13].

Lemma 3.1. *With the above conditions, setting $p_\alpha = \sup_{h>0} \frac{h\alpha'_+(h)}{\alpha(h)}$, the function $h \mapsto \alpha^{1/p_\alpha}(h)$ is sub-additive, namely*

$$\alpha^{1/p_\alpha}(h+k) \leq \alpha^{1/p_\alpha}(h) + \alpha^{1/p_\alpha}(k), \quad \forall h, k \in \mathbb{R}^+.$$

As a consequence, $d_\alpha(x, y) = \alpha^{1/p_\alpha}(d(x, y))$, $x, y \in \mathcal{X}$ is a distance on \mathcal{X} .

This Lemma together with the triangular inequality gives for all $\nu_1, \nu_2 \in \mathcal{P}(\mathcal{X})$,

$$\begin{aligned} \mathcal{T}_\alpha(\nu_1, \nu_2) &= W_{p_\alpha}^{p_\alpha}(\nu_1, \nu_2) \\ &\leq (W_{p_\alpha}(\nu_1, \mu) + W_{p_\alpha}(\mu, \nu_2))^{p_\alpha} = \left(\mathcal{T}_\alpha^{1/p_\alpha}(\nu_1, \mu) + \mathcal{T}_\alpha^{1/p_\alpha}(\mu, \nu_2) \right)^{p_\alpha}, \end{aligned}$$

where the Wasserstein distance W_{p_α} is understood with respect to the distance d_α of Lemma 3.1. Then, observing that $p_\alpha > 1$ when the function α is not linear, and using the identity

$$\left(u^{1/p} + v^{1/p}\right)^p = \inf_{t \in (0,1)} \left\{ \frac{u}{t^{p-1}} + \frac{v}{(1-t)^{p-1}} \right\}, \quad p > 1,$$

we get that μ satisfies the usual transport inequality

$$\mathcal{T}_\alpha(\mu, \nu) \leq H(\nu|\mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}),$$

if and only if μ satisfies the following transport inequalities: for all $t \in (0, 1)$,

$$(9) \quad \mathcal{T}_\alpha(\nu_1, \nu_2) \leq \frac{H(\nu_1|\mu)}{t^{p_\alpha-1}} + \frac{H(\nu_2|\mu)}{(1-t)^{p_\alpha-1}}, \quad \forall \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}).$$

Here is another example of transport inequality. When $c(x, p) = 2 \int \mathbb{1}_{x \neq y} dp(y)$, the universal optimal transport cost $T_c(\nu|\mu)$ is in fact the total variation distance between the measures μ and ν

$$T_c(\nu|\mu) = \|\mu - \nu\|_{TV} = 2 \sup_{A \subset \mathcal{X}} |\mu(A) - \nu(A)|.$$

The Csiszár-Kullback-Pinsker inequality [Pin64, Csi67, Kul67] that holds for any (reference) probability measure μ ,

$$\|\mu - \nu\|_{TV}^2 \leq 2 H(\nu|\mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}).$$

corresponds to the transport inequalities $\mathbf{T}_{c,\beta}^+(b)$ or $\mathbf{T}_{c,\beta}^-(b)$ with $\beta(r) = r^2/2$, $r \geq 0$. Here again, this inequality is equivalent to $\mathbf{T}_{c,\beta}(b/t, b/(1-t))$ for all $t \in (0, 1)$. This inequality and its improvements are known for their numerous applications in probability, in analysis and in information theory (cf. [Vil09], page 636).

As a last example, let us consider the universal cost function

$$c(x, p) = \left(\int \mathbb{1}_{x \neq y} dp(y) \right)^2, \quad x \in \mathcal{X}, \quad p \in \mathcal{P}(\mathcal{X}).$$

Then, the transport inequalities $\mathbf{T}_c^+(2)$ and $\mathbf{T}_c^-(2)$ correspond to the weak transport inequalities introduced by Marton [Mar96b]. As for the Csiszár-Kullback-Pinsker inequality, $\mathbf{T}_c^+(2)$ and $\mathbf{T}_c^-(2)$ hold for any (reference) probability measure μ . In that case, $\mathbf{T}_c^+(b)$ and $\mathbf{T}_c^-(b)$ are equivalent to the family of transport inequalities $\mathbf{T}_c(b/t, b/(1-t))$ for $t \in (0, 1)$, since the weak-transport cost T_c , also denoted by $\widetilde{\mathcal{T}}_2$, satisfies the following triangular inequality [Mar97],

$$(10) \quad \sqrt{\widetilde{\mathcal{T}}_2(\nu_1|\nu_2)} \leq \sqrt{\widetilde{\mathcal{T}}_2(\nu_1|\mu)} + \sqrt{\widetilde{\mathcal{T}}_2(\mu|\nu_2)}, \quad \forall \mu, \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}).$$

3.2. Functional formulation of transport inequality, the dual Kantorovich Theorem. The dual functional formulation of usual transport inequalities has been obtained by Bobkov and Götze [BG99] and then expanded in the paper [GL10].

This dual form is based on the duality between the relative entropy and the log-Laplace transform. Namely, for any continuous bounded function $g : \mathcal{X} \rightarrow \mathbb{R}$,

$$(11) \quad \log \int e^g d\mu = \sup_{\nu \in \mathcal{P}(\mathcal{X})} \left\{ \int g d\nu - H(\nu|\mu) \right\}.$$

A simple proof of this identity is given in [GL10] and one more general in [GRS11a].

The second argument is the dual Kantorovich Theorem. This theorem is well-known for usual lower semi-continuous cost functions $\omega : \mathcal{X} \times \mathcal{X} \rightarrow (-\infty, +\infty]$ (cf. [Vil09])

$$(12) \quad \mathcal{T}_\omega(\nu_1, \nu_2) = \sup_{\varphi \in \mathcal{C}_b(\mathcal{X})} \left\{ \int Q_\omega \varphi d\nu_2 - \int \varphi d\nu_1 \right\}, \quad \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}),$$

where $\mathcal{C}_b(\mathcal{X})$ is the set of continuous bounded functions on \mathcal{X} and

$$Q_\omega \varphi(y) = \inf_{x \in \mathcal{X}} \{ \varphi(x) + \omega(x, y) \}, \quad y \in \mathcal{X}.$$

In the paper [GRST14b], as the function $p \in \mathcal{P}(\mathcal{X}) \rightarrow c(x, p)$ is convex, this result is extended to weak transport costs T_c , under weak regularity additional assumptions on the cost $c : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \rightarrow [0, +\infty]$ (see Theorem 3.5, [GRST14b]). Overall, the result is the following,

$$(13) \quad T_c(\nu_1 | \nu_2) = \sup_{\varphi \in \mathcal{C}_b(\mathcal{X})} \left\{ \int R_c \varphi d\nu_2 - \int \varphi d\nu_1 \right\}, \quad \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}),$$

where $R_c \varphi$ is the infimum-convolution operator (1) previously defined,

$$R_c \varphi(x) = \inf_{p \in \mathcal{P}(\mathcal{X})} \left\{ \int \varphi dp + c(x, p) \right\}.$$

To be precise, we should slightly modify the sets $\mathcal{C}_b(\mathcal{X})$ and $\mathcal{P}(\mathcal{X})$, depending on the type of involved cost function c (see [GRST14b]).

The two duality identities (11) and (13) provide the following functional formulation of the transport-entropy inequality $\mathbf{T}_{c,\beta}(a_1, a_2)$.

Proposition 3.1. *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\beta : \mathbb{R}_+ \rightarrow [0, +\infty]$ be a lower semi-continuous convex function such that $\beta(0) = 0$. The following statements are equivalent.*

- (1) *The probability measure μ satisfies $\mathbf{T}_{c,\beta}(a_1, a_2)$.*
- (2) *For all functions $\varphi \in \mathcal{C}_b(\mathcal{X})$ and for all $\lambda \geq 0$,*

$$\left(\int e^{\frac{\lambda R_c \varphi}{a_2}} d\mu \right)^{a_2} \left(\int e^{-\frac{\lambda \varphi}{a_1}} d\mu \right)^{a_1} \leq e^{\beta^*(\lambda)},$$

$$\text{with } \beta^*(\lambda) = \sup_{t \geq 0} \{ \lambda t - \beta(t) \}.$$

Point (2) generalizes the infimum-convolution view point of transport inequalities introduced by Maurey [Mau91], the so-called (τ) -property.

Idea of the proof (1) \Rightarrow (2). One has $\beta(t) = \sup_{\lambda \geq 0} \{ \lambda t - \beta^*(\lambda) \}$, $\forall t \geq 0$. If μ satisfies $\mathbf{T}_{c,\beta}(a_1, a_2)$, then the general dual Kantorovich identity (13) implies that for all $\varphi \in \mathcal{C}_b(\mathcal{X})$ and all $\lambda \geq 0$,

$$\lambda \left(\int R_c \varphi d\nu_2 - \int \varphi d\nu_1 \right) - \beta^*(\lambda) \leq a_2 H(\nu_2 | \mu) + a_1 H(\nu_1 | \mu), \quad \forall \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}).$$

Point (2) follows by reordering the terms of this inequality, by optimizing over all probability measures ν_1 and ν_2 , and then by applying the dual formula (11), with the function $g = \lambda R_c \varphi / a_2$ and with the function $g = -\lambda \varphi / a_1$. \square

By density of the set of the bounded continuous functions in $L^1(\mu)$, and then by monotone convergence, (2) also holds for all measurable functions $\varphi : \mathcal{X} \rightarrow (-\infty, +\infty]$ bounded from below.

Given $A \in \mathcal{B}(\mathcal{X})$, let us consider the function i_A equal to 0 on A , and to $+\infty$ on its complement. Applying (2) to the function $\varphi = i_A$, since $R_c i_A(x) = c(x, A)$, $x \in \mathcal{X}$, the transport inequality $\mathbf{T}_{c,\beta}(a_1, a_2)$ provides the Talagrand's formulation of concentration properties (cf. [Tal95, Tal96a, Tal96b]):

$$\int e^{\frac{\lambda c(x,A)}{a_2}} d\mu \leq \frac{e^{\beta^*(\lambda)/a_2}}{\mu(A)^{a_1/a_2}}, \quad \forall \lambda \geq 0, \quad \forall A \in \mathcal{B}(\mathcal{X}).$$

When the function β is the identity, one has $\beta^* = i_{(-\infty,1]}$ and Proposition 3.1 is writing as follows.

Proposition 3.2. [GRST14b] *The following statements are equivalent.*

- (1) *The probability μ satisfies $\mathbf{T}_c(a_1, a_2)$.*
- (2) *For all function $\varphi \in \mathcal{C}_b(\mathcal{X})$,*

$$\left(\int e^{\frac{R_c \varphi}{a_2}} d\mu \right)^{a_2} \left(\int e^{-\frac{\varphi}{a_1}} d\mu \right)^{a_1} \leq 1.$$

3.3. Tensorization - Characterization by dimension-free concentration properties. The transport entropy inequalities tensorize, and this provides concentration results in high dimension.

Proposition 3.3. *Let \mathcal{X}_1 and \mathcal{X}_2 be Polish spaces. Let $\beta_1 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, $\beta_2 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be convex functions and let $\beta : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be defined by*

$$\beta(t) = \beta_1 \square \beta_2(t) = \inf\{\beta_1(t_1) + \beta_2(t_2), t = t_1 + t_2\}, \quad t \geq 0.$$

If $\mu_1 \in \mathcal{P}(\mathcal{X}_1)$ and $\mu_2 \in \mathcal{P}(\mathcal{X}_2)$ satisfy respectively the transport inequalities $\mathbf{T}_{c_1, \beta_1}(a_1, a_2)$ and $\mathbf{T}_{c_2, \beta_2}(a_1, a_2)$, then $\mu_1 \otimes \mu_2 \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ satisfies the transport inequality $\mathbf{T}_{c, \beta}(a_1, a_2)$ with for all $x = (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, and all $p \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ with marginals $p_1 \in \mathcal{P}(\mathcal{X}_1)$ and $p_2 \in \mathcal{P}(\mathcal{X}_2)$

$$c(x, p) = c_1 \oplus c_2(x, p) = c_1(x_1, p_1) + c_2(x_2, p_2).$$

The proof of this proposition exactly follows the one of Theorem 4.11 [GRST14b] for which $\beta_1(t) = \beta_2(t) = \beta(t) = t$, $t \geq 0$. We could also follow the tensorization proof given in [Sam07] on the dual functional form of such transport inequalities.

This tensorization property is a consequence of the tensorization properties of the relative entropy and of the optimal transport cost: for any measure $\nu \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ with decomposition $d\nu(x_1, x_2) = d\nu_1(x_1)d\nu_2^{x_1}(x_2)$, one has

$$H(\nu|\mu) = H(\nu_1|\mu_1) + \int H(\nu_2^{x_1}|\mu_2)d\nu_1(x_1),$$

and for any other measure $\nu' \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ with decomposition $d\nu'(x'_1, x'_2) = d\nu_1(x'_1)d\nu_2^{x'_1}(x'_2)$, for all $\varepsilon \geq 0$, there exists $\pi_1^\varepsilon \in \Pi(\nu_1, \nu'_1)$ such that

$$T_c(\nu|\nu') \leq T_{c_1}(\nu_1|\nu'_1) + \iint T_{c_2}(\nu_2^{x_1}|\nu_2^{x'_1}) d\pi_1^\varepsilon(x_1, x'_1) + \varepsilon,$$

where $c = c_1 \oplus c_2$. The error term ε can be chosen equal to 0 when \mathcal{X}_1 and \mathcal{X}_2 are compact spaces.

Therefore, if $\mu \in \mathcal{P}(\mathcal{X})$ satisfies the transport inequality $\mathbf{T}_{c,\beta}(a_1, a_2)$, then $\mu^n = \mu \otimes \cdots \otimes \mu \in \mathcal{P}(\mathcal{X}^n)$ satisfies $\mathbf{T}_{c^n, \beta \square^n}(a_1, a_2)$, with for all $p \in \mathcal{P}(\mathcal{X}^n)$, with marginals $p_i \in \mathcal{P}(\mathcal{X}), i \in \{1, \dots, n\}$,

$$c^n(x, p) = c^{\oplus n}(x, p) := c(x_1, p_1) + \cdots + c(x_n, p_n), \quad x = (x_1, \dots, x_n) \in \mathcal{X}^n,$$

and since β is convex,

$$\beta^{\square n}(t) := \beta \square \cdots \square \beta(t) = n\beta(t/n), \quad t \geq 0.$$

From the transport inequality $\mathbf{T}_{c^n, \beta \square^n}(a_1, a_2)$, we get that μ^n satisfies the following concentration property, for all $A \in \mathcal{B}(\mathcal{X}^n)$,

$$\mu^n(A)^{a_1} \mu^n(\mathcal{X} \setminus A_{r, c^n})^{a_2} \leq e^{-n\beta(r/n)}, \quad \forall r \geq 0.$$

The concentration profile in the right-hand side is independent of n if and only if β is linear. In that case, we say that μ satisfies a *dimension-free concentration property*.

Definition 3.3. *A measure $\mu \in \mathcal{P}(\mathcal{X})$ satisfies a dimension-free concentration property associated to a cost function $c : \mathcal{X} \times \mathcal{P}(\mathcal{X})$, if for all $n \geq 1$, and for all $A \in \mathcal{B}(\mathcal{X}^n)$,*

$$\mu^n(A)^{a_1} \mu^n(\mathcal{X} \setminus A_{r, c^n})^{a_2} \leq e^{-r}, \quad \forall r \geq 0.$$

Actually, Gozlan has proved that for usual enlargements associated to the cost function $\omega(x, y) = \alpha(d(x, y))$, this dimension-free concentration property is equivalent to a transport inequality (see [Goz09]). Its proof is based on large deviation technics. In the paper [GRST14b], a simpler approach, starting from the dual formulation of transport inequalities, allows to extend Gozlan's result to any transport inequality $\mathbf{T}_c(a_1, a_2)$.

Proposition 3.4. [GRST14b] *The following statements are equivalent.*

(i) μ satisfies $\mathbf{T}_c(a_1, a_2)$: for all functions $\psi \in \mathcal{C}_b(\mathcal{X})$,

$$\left(\int e^{\frac{R_c \psi}{a_2}} d\mu \right)^{a_2} \left(\int e^{-\frac{\psi}{a_1}} d\mu \right)^{a_1} \leq 1.$$

(ii) For all integers $n \geq 1$ and for all functions $\varphi \in \mathcal{C}_b(\mathcal{X}^n)$,

$$(14) \quad \mu^n(\varphi \leq m)^{a_1} \mu^n(R_{c^n} \varphi > m + r)^{a_2} \leq e^{-r}, \quad \forall m \in \mathbb{R}, \quad \forall r \geq 0.$$

Idea of the proof. As already explained, (i) \Rightarrow (ii) is a consequence of the tensorization properties of the inequality $\mathbf{T}_c(a_1, a_2)$.

In order to get (ii) \Rightarrow (i), we estimate the product of exponential moments of $R_{c^{\oplus n}} \varphi$ and $-\varphi$ using the tail distribution estimates given by (ii). More precisely, (ii) provides: for all $\varepsilon > 0$,

$$\left(\int e^{\frac{R_{c^{\oplus n}} \varphi}{(1+\varepsilon)a_2}} d\mu^n \right)^{a_2} \left(\int e^{-\frac{\varphi}{(1-\varepsilon)a_1}} d\mu^n \right)^{a_1} \leq K(\varepsilon, a_1, a_2),$$

where $K(\varepsilon, a_1, a_2)$ is a constant independent of n . We want to “tighten” this inequality by replacing this constant by 1. For that purpose, let us choose $\varphi(x) = \psi(x_1) + \cdots + \psi(x_n)$, $x = (x_1, \dots, x_n) \in \mathcal{X}^n$, for which

$$R_{c^n} \varphi(x) = R_c \psi(x_1) + \cdots + R_c \psi(x_n).$$

By independence, the last inequality can be rewritten as follows

$$\left(\int e^{\frac{R_c \psi}{(1+\varepsilon)^{a_2}}} d\mu \right)^{a_2} \left(\int e^{-\frac{\psi}{(1-\varepsilon)^{a_1}}} d\mu \right)^{a_1} \leq K(\varepsilon, a_1, a_2)^{1/n}.$$

The result follows from this inequality as n goes to $+\infty$ and then ε goes to 0. \square

3.4. Connections with logarithmic Sobolev inequalities. In this section, we assume that the closed balls of the Polish metric space (\mathcal{X}, d) are compact. In this part, the transport costs are associated to usual cost functions on a metric space (\mathcal{X}, d) :

$$c(x, p) = \int \alpha(d(x, y)) dp(y), \quad x \in \mathcal{X}, p \in \mathcal{P}(\mathcal{X}),$$

where $\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is a convex function such that $\alpha(0) = \alpha'(0) = 0$ satisfying the Δ_2 -condition (8). In this case, we note $\mathcal{T}_\alpha(\mu, \nu) = T_c(\mu|\nu)$ and $\mathbf{T}_\alpha(b)$ the transport inequality $\mathbf{T}_c^+(b)$ that coincides with $\mathbf{T}_c^-(b)$.

For any locally Lipschitz function $f : \mathcal{X} \rightarrow \mathbb{R}$, the gradient norms of f at a non-isolated point $x \in \mathcal{X}$ are defined by

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

and $|\nabla^+ f|(x) = |\nabla^- f|(x) = 0$ if x is an isolated point. If \mathcal{X} is a Riemannian manifold and f is smooth, $|\nabla^+ f|(x)$ and $|\nabla^- f|(x)$ are the norm of $\nabla f(x)$ in the tangent space $T_x \mathcal{X}$ at point x .

Definition 3.4. A measure $\mu \in \mathcal{P}(\mathcal{X})$ satisfies the modified logarithmic Sobolev inequality $\mathbf{LogSob}_\alpha^+(b)$, $b \geq 0$, associated to the cost α , if for any locally Lipschitz function $f : \mathcal{X} \rightarrow \mathbb{R}$, one has

$$\mathbf{LogSob}_\alpha^+(b) : \quad \text{Ent}_\mu(e^f) \leq b \int \alpha^*(|\nabla^+ f|) e^f d\mu,$$

where $\alpha^*(h) = \sup_{t \geq 0} \{ht - \alpha(t)\}$ and for any function $g : \mathcal{X} \rightarrow \mathbb{R}^+$,

$$\text{Ent}_\mu(g) = \int g \log g d\mu - \int g d\mu \log \int g d\mu.$$

In the same way, we define the logarithmic Sobolev inequality $\mathbf{LogSob}_\alpha^-(b)$ by replacing $|\nabla^+ f|$ by $|\nabla^- f|$. If \mathcal{X} is a Riemannian manifold, we simply note $\mathbf{LogSob}_\alpha(b)$. When α is quadratic, $\alpha(t) = t^2$, $t \geq 0$, $\alpha^*(h) = h^2/4$, $h \geq 0$, the logarithmic Sobolev inequalities are denoted by $\mathbf{LogSob}_2^+(b)$ or $\mathbf{LogSob}_2^-(b)$.

As a first result, the well-known Otto-Villani Theorem asserts that the Talagrand’s transport inequality is a consequence of the logarithmic Sobolev inequality.

Theorem 3.1. [OV00] Let \mathcal{X} be a Riemannian manifold. If $\mu \in \mathcal{P}_2(\mathcal{X})$ satisfies the logarithmic Sobolev inequality $\mathbf{LogSob}_2(b)$ then μ satisfies the Talagrand’s transport inequality $\mathbf{T}_2(b)$.

In the heuristic part of [OV00], Otto and Villani give the idea of their proof of this result by interpreting the Wasserstein space $(\mathcal{P}_2(X), W_2)$ as a Riemannian manifold and by considering the gradient flow of the relative entropy $\nu \mapsto H(\nu|\mu)$. Bobkov, Gentil and Ledoux [BGL01] give another proof based on the Hopf-Lax formula for the solutions of the Hamilton-Jacobi equation. More precisely, on a Riemannian manifold \mathcal{X} , the infimum-convolution operator

$$v(x, t) = Q_t f(x) = \inf_{y \in \mathcal{X}} \left\{ f(y) + \frac{1}{2t} d(x, y)^2 \right\}, \quad x \in \mathcal{X}, \quad t > 0,$$

is a semi-group, solution of the Hamilton-Jacobi equation

$$\frac{\partial v}{\partial t} = -\frac{1}{2} |\nabla v|^2, \quad \text{avec} \quad v(x, 0) = f(x), \quad \forall x \in \mathcal{X}.$$

A counter example, showing that Otto-Villani's Theorem can not be reversed in full generality, has been given by Cattiaux and Guillin [CG06] (see also [Goz07]).

Then Otto-Villani's result has been complemented by Gozlan and al. in a series of papers [GRS11b, GRS13, GRS14]. Following the Hamilton-Jacobi approach by Bobkov-Gentil-Ledoux, the modified logarithmic Sobolev inequality $\mathbf{LogSob}_\alpha^-(b)$ is characterized in terms of hypercontractivity property of the operator $Q_t f$ defined by

$$Q_t f(x) = \inf_{y \in \mathcal{X}} \left\{ f(y) + t\alpha \left(\frac{d(x, y)}{t} \right) \right\}, \quad x \in \mathcal{X},$$

for all bounded function $f : \mathcal{X} \rightarrow \mathbb{R}$.

Theorem 3.2. [GRS14] *Assume that α satisfies the Δ_2 -condition (8). Then the exponents $r_\alpha \leq p_\alpha$ defined by*

$$r_\alpha = \inf_{x>0} \frac{x\alpha'_-(x)}{\alpha(x)} \geq 1 \quad \text{and} \quad 1 < p_\alpha = \sup_{x>0} \frac{x\alpha'_+(x)}{\alpha(x)}$$

are both finite. Moreover, the measure μ satisfies $\mathbf{LogSob}_\alpha^-(b)$ if and only if for all $t > 0$, for all $t_o \leq b(p_\alpha - 1)$ and for all bounded continuous functions $f : X \rightarrow \mathbb{R}$,

$$\|e^{Q_t f}\|_{k(t)} \leq \|e^f\|_{k(0)},$$

with

$$k(t) = \begin{cases} \left(1 + \frac{b^{-1}(t-t_o)}{p_\alpha-1}\right)^{p_\alpha-1} \mathbf{1}_{t \leq t_o} + \left(1 + \frac{b^{-1}(t-t_o)}{r_\alpha-1}\right)^{r_\alpha-1} \mathbf{1}_{t > t_o} & \text{if } r_\alpha > 1 \\ \min\left(1; \left(1 + \frac{b^{-1}(t-t_o)}{p_\alpha-1}\right)^{p_\alpha-1}\right) & \text{if } r_\alpha = 1 \end{cases},$$

where $\|g\|_k = (\int |g|^k d\mu)^{1/k}$ for $k \neq 0$ and $\|g\|_0 = \exp(\int \log g d\mu)$.

By choosing $t_o = b(p_\alpha - 1)$ and after some easy computations this theorem implies the following Otto-Villani Theorem, extended to any metric space and for any cost function α satisfying the Δ_2 -condition.

Theorem 3.3. [GRS14] *Suppose that α verifies the Δ_2 -condition (8). If μ verifies $\mathbf{LogSob}_\alpha^-(b)$, then it verifies $\mathbf{T}_\alpha(B)$, with*

$$B = \max\left(\left((p_\alpha - 1)b\right)^{r_\alpha-1}; \left((p_\alpha - 1)b\right)^{p_\alpha-1}\right),$$

where the numbers r_α, p_α are defined in Theorem 3.2.

This result exactly recovers Otto-Villani Theorem 3.1 since $p_\alpha = r_\alpha = 2$ for $\alpha(h) = h^2$, $h \geq 0$.

Popular functions α appearing as cost functions in the literature are the functions $\alpha = \alpha_{p_1, p_2}$, with $p_1 \geq 2$ and $p_2 \geq 1$ defined by

$$\alpha_{p_1, p_2}(h) = \begin{cases} h^{p_1} & \text{if } 0 \leq h \leq 1, \\ \frac{p_1}{p_2} h^{p_2} + 1 - \frac{p_1}{p_2} & \text{if } h \geq 1. \end{cases}$$

Any such function satisfies the Δ_2 -condition with $r_\alpha = \min(p_1, p_2)$ and $p_\alpha = \max(p_1, p_2)$.

As examples, the best known measures on \mathbb{R}^n satisfying the logarithmic Sobolev inequality $\mathbf{LogSob}_{\alpha_2, p}(b)$ for some $b > 0$ are the standard Gaussian measure for $p = 2$ [Gro75], the exponential measure for $p = 1$ [BL97], and more generally the probability measures $d\mu_p = e^{-|t|^p} / Z_p dt$, for $p \geq 1$ (see [GGM07, BR08, Goz07]). For these measures, Theorem 3.2 provides the related transport inequalities obtained in different papers [Tal91, BK08, GGM05].

To end the comparisons between logarithmic Sobolev inequalities and transport inequalities, let us recall the reversed Otto-Villani's Theorem obtained in the papers [GRS11b, GRS13, GRS14]. It characterizes the transport inequalities in terms of modified logarithmic Sobolev inequalities restricted to a class of $K - \alpha$ -convex functions. By definition, a function $f : \mathcal{X} \rightarrow \mathbb{R}$ is $K - \alpha$ -convex if there exists a function $h : \mathcal{X} \rightarrow \mathbb{R}$ such that

$$f(x) = \sup_{y \in \mathcal{X}} \{h(y) - K\alpha(d(x, y))\} = P_\alpha^K h(x), \quad \forall x \in \mathcal{X}.$$

On the Euclidean space, if $\alpha(h) = h^2$, $h \geq 0$, then a smooth $K - \alpha$ -convex function is exactly a function with Hessian bounded from below.

Let us summarize the results of Theorem 1.12 [GRS11b], Theorem 5.1 [GRS13] when (\mathcal{X}, d) is a *geodesic space*: for any $x, y \in \mathcal{X}$, there exists a path $(\gamma_t)_{t \in [0, 1]}$ in \mathcal{X} , such that $\gamma_0 = x$, $\gamma_1 = y$ and $d(\gamma_s, \gamma_t) = |t - s|d(x, y)$, for all $s, t \in [0, 1]$.

Theorem 3.4. [GRS11b, GRS13] *Let (\mathcal{X}, d) be a geodesic space and $\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be a convex function satisfying the Δ_2 -condition (8) and such that $\alpha(0) = \alpha'(0) = 0$. The following properties are equivalent.*

- (1) *There exists $C_1 > 0$, such that μ satisfies the transport inequality $\mathbf{T}_\alpha(C_1)$.*
- (2) *There exist $C_2 > 0$ and $\lambda > 0$, such that μ satisfies the following (τ) -log-Sobolev inequality: for all locally Lipschitz functions $f : \mathcal{X} \rightarrow \mathbb{R}$,*

$$(\tau) - \mathbf{LogSob}_\alpha(C_2, \lambda) \quad \text{Ent}_\mu(e^f) \leq C_2 \int (f - Q_\alpha^\lambda f) e^f d\mu,$$

where $Q_\alpha^\lambda f(x) := \inf_{y \in \mathcal{X}} \{f(y) + \lambda\alpha(d(x, y))\}$, $x \in \mathcal{X}$.

- (3) *There exist $C_3 > 0$ and $\lambda > 0$, such that μ satisfies the following restricted modified logarithmic Sobolev inequality: for all $K - \alpha$ -convex functions $f : \mathcal{X} \rightarrow \mathbb{R}$, with $0 \leq K < \lambda$*

$$\mathbf{r} - \mathbf{LogSob}_\alpha(C_3, \lambda) \quad \text{Ent}_\mu(e^f) \leq C_3 \int \alpha^*(|\nabla^+ f|) e^f d\mu.$$

The logarithmic Sobolev inequality at point (2) is called $(\tau) - \mathbf{LogSob}_\alpha(C_2, \lambda)$, as a reference to the (τ) -property by Maurey [Mau91] for which the infimum-convolution operator also occurs.

A main application of this characterization is the following perturbation result.

Corollary 3.1. (Theorem 1.9 [GRS11b]) *Let (\mathcal{X}, d) be a geodesic space and $\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be a convex function satisfying the Δ_2 -condition (8) and such that $\alpha(0) = \alpha'(0) = 0$. Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\tilde{\mu} \in \mathcal{P}(\mathcal{X})$ with density e^ϕ with respect to μ , $\phi : \mathcal{X} \rightarrow \mathbb{R}$. If μ satisfies $\mathbf{T}_\alpha(C)$ then $\tilde{\mu}$ satisfies $\mathbf{T}_\alpha(8Ce^{\text{Osc } \phi})$ with $\text{Osc } \phi = \sup \phi - \inf \phi$.*

This type of perturbation's result has been established by Holley and Stroock [HS87] for usual logarithmic Sobolev inequalities. This corollary follows by applying their arguments to logarithmic Sobolev inequalities restricted to a class of functions.

4. SOME RESULTS AROUND "BARYCENTRIC" COSTS

In all this part, c is a barycentric cost function,

$$c(x, p) = \theta \left(x - \int y dp(y) \right), \quad x \in \mathbb{R}^n, p \in \mathcal{P}_1(\mathbb{R}^n),$$

where $\theta : \mathbb{R}^n \rightarrow \mathbb{R}^+$ is a convex function. Most of the results of this section extend to lower semi-continuous convex functions $\theta : \mathbb{R}^n \rightarrow [0, +\infty]$.

In that case, the optimal transport cost $T_c(\nu|\mu)$ between μ and ν in $\mathcal{P}(\mathbb{R}^n)$ is denoted by $\bar{T}_\theta(\nu|\mu)$. The following specific Kantorovich dual expression of $\bar{T}_\theta(\nu|\mu)$ has been obtained in [GRST14b] (see Theorem 2.11)

$$(15) \quad \bar{T}_\theta(\nu|\mu) = \sup \left\{ \int Q\varphi d\mu - \int \varphi d\nu; \varphi \text{ convex, Lipschitz, bounded from below} \right\}.$$

In the supremum, $Q\varphi$ is the usual infimum-convolution operator,

$$Q\varphi(x) = \inf_{y \in \mathbb{R}^n} \{ \varphi(y) + \theta(x - y) \}, \quad x \in \mathbb{R}^n.$$

Therefore, by a restriction to convex functions, the operator $Q\varphi$ replaces the operator $R_c\varphi$ in the dual formula (13).

Let γ and ν be two probability measures on \mathbb{R}^n . By definition, the measure γ is dominated by ν in the convex order, and we note $\gamma \preceq \nu$, if for all convex functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\int f d\gamma \leq \int f d\nu.$$

The following Strassen's Theorem provides an alternative definition of the convex order.

Theorem 4.1. [Str65] *Let γ and ν be two probability measures on \mathbb{R}^n , then $\gamma \preceq \nu$ if and only if there exists a martingale (X, Y) for which X has law γ and Y has law ν . Namely, if π is the law of the couple (X, Y) , with decomposition $d\pi(x, y) =$*

$d\gamma(x)dp_x(y)$, where p is a Markov Kernel such that $\gamma p = \nu$, then one has for γ -almost every x ,

$$\int y dp_x(y) = x.$$

Idea of the proof. A simple proof follows from the dual Kantorovich expression of the optimal barycentric cost:

$$\begin{aligned} \bar{\mathcal{T}}_1(\nu|\gamma) &= \inf_{\pi \in \Pi(\mu, \nu)} \int \left| x - \int y dp_x(y) \right| d\gamma(x) \\ &= \sup \left\{ \int f d\gamma - \int f d\nu; f \text{ convex, 1-lipschitz, lower bounded} \right\}, \end{aligned}$$

(see Proposition 3.2. [GRST14b]). Therefore, if $d\pi^*(x, y) = d\gamma(x)dp_x^*(y)$ is the law of (X, Y) , with marginals γ and ν , then

$$0 \leq \bar{\mathcal{T}}_1(\nu|\gamma) \leq \int \left| x - \int y dp_x^*(y) \right| d\gamma(x) = 0.$$

It follows that $\bar{\mathcal{T}}_1(\nu|\gamma) = 0$ and the dual expression of $\bar{\mathcal{T}}_1(\nu|\gamma)$ gives

$$\int f d\gamma \leq \int f d\nu,$$

for every convex, 1-Lipschitz, lower bounded function f . This inequality extends to any lower bounded convex function and then to any convex function by monotone convergence, which means that $\nu \preceq \gamma$. One way to prove this is to use the fact that if f is convex lower-bounded then the classical infimum convolution operator $Q_t f(x) := \inf_{y \in \mathbb{R}^n} \{f(y) + \frac{1}{t}|x - y|\}$ is convex $1/t$ -Lipschitz and $Q_t f(x)$ is increasing to $f(x)$ as t goes to 0 for all $x \in \mathbb{R}^n$. \square

Let $\mu \in \mathcal{P}(\mathbb{R}^n)$ be such that for any $\gamma \in \mathcal{P}(\mathbb{R}^n)$, there exists an optimal transport map $S^* : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $S^* \# \mu = \gamma$ and

$$\mathcal{T}_\theta(\gamma, \mu) = \inf_{\pi \in \Pi(\mu, \nu)} \int \theta(x - y) d\pi(x, y) = \int \theta(x - S^*(x)) d\mu(x).$$

This assumption is satisfied for example when μ is absolutely continuous with respect to the Lebesgue measure and θ is smooth and strictly convex (see e.g. [Vil09], Theorem 9.4, Theorem 10.28). Then the optimal barycentric transport cost $\bar{\mathcal{T}}_\theta$ can be expressed in terms of the usual transport cost \mathcal{T}_θ as follows.

Proposition 4.1. *Under the above conditions on the probability measure $\mu \in \mathcal{P}(\mathbb{R}^n)$, for any probability measure ν such that $\bar{\mathcal{T}}_\theta(\nu|\mu) < \infty$,*

$$\bar{\mathcal{T}}_\theta(\nu|\mu) = \inf_{\gamma \in \mathcal{P}(\mathbb{R}^n), \gamma \preceq \nu} \mathcal{T}_\theta(\gamma, \mu).$$

Proof. Let $\gamma \preceq \nu$. From the previous Strassen's Theorem, there exists a kernel p^* such that $\gamma p^* = \nu$ and $x = \int y dp_x^*(y)$ γ -almost surely. Furthermore, by hypotheses, there exists a transport map $S^* : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $S^* \# \mu = \gamma$ and

$$\mathcal{T}_\theta(\gamma, \mu) = \int \theta(x - S^*(x)) d\mu(x).$$

Let us consider the kernel defined by $p_x(dy) = p_{S^*(x)}^*(dy)$. We may simply check that $\mu p = \nu$ and for μ -almost every x , $\int y dp_x(y) = \int y dp_{S^*(x)}^*(y) = S^*(x)$. Therefore, $\mathcal{T}_\theta(\gamma, \mu) = \int \theta(x - \int y dp_x(y)) d\mu(x) \geq \overline{\mathcal{T}}_\theta(\nu|\mu)$, and by optimizing over all probability measures γ , with $\gamma \preceq \nu$, we get

$$\inf_{\gamma \in \mathcal{P}(\mathbb{R}^n), \gamma \preceq \nu} \mathcal{T}_\theta(\gamma, \mu) \geq \overline{\mathcal{T}}_\theta(\nu|\mu).$$

To prove the reverse inequality, let us consider a kernel p such that $\mu p = \nu$ and

$$\int \theta\left(x - \int y dp_x(y)\right) d\mu(x) < \infty.$$

Let $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the measurable map defined by $S(x) = \int y dp_x(y)$, for μ -almost every x . Let γ be the push forward measure of μ by the map S , $\gamma = S\#\mu$. Then, one has $\gamma \preceq \nu$, since by Jensen's inequality, for all convex functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\int f d\gamma = \int f\left(\int y dp_x(y)\right) d\mu(x) \leq \iint f(y) dp_x(y) d\mu(x) = \int f d\nu;$$

and moreover

$$\int \theta\left(x - \int y dp_x(y)\right) d\mu(x) = \int (x - S(x)) d\mu(x) \geq \mathcal{T}_\theta(\gamma, \mu) \geq \inf_{\gamma \in \mathcal{P}(\mathbb{R}^n), \gamma \preceq \nu} \mathcal{T}_\theta(\gamma, \mu).$$

We get the expected inequality by optimizing this inequality over all kernels p such that $\mu p = \nu$. \square

4.1. Barycentric transport inequality and logarithmic Sobolev inequalities. As for the usual transport costs, connections have been established between barycentric transport inequalities and logarithmic Sobolev inequalities restricted to a class of functions (see [GRST14b, AS15]). To simplify, in this section we only consider the case $\theta(h) = \|h\|^2$ where $\|\cdot\|$ is a fixed norm on \mathbb{R}^n . In that case we note $\overline{\mathcal{T}}_\theta = \overline{\mathcal{T}}_2$.

The next results have been obtained by Gozlan and al. [GRST14b] thanks to the Kantorovich dual expression (15) of $\overline{\mathcal{T}}_2$, and by applying the technics linked to the Hamilton-Jacobi equation satisfied by the semi-group $Q_t\varphi$,

$$Q_t\varphi(x) = \inf_{y \in \mathbb{R}^n} \left\{ \varphi(y) + \frac{1}{t} \|x - y\|^2 \right\}, x \in \mathbb{R}^n.$$

From the non-symmetry of the optimal transport cost $\overline{\mathcal{T}}_2(\nu|\mu)$, they establish two different results one corresponding to the transport inequality $\mathbf{T}_c^+(C)$ and the other associated to $\mathbf{T}_c^-(C)$.

Theorem 4.2. (see Theorem 8.15 [GRST14b]) *Let $\mu \in \mathcal{P}_1(\mathbb{R}^n)$. The following properties are equivalent.*

- (1) *There exists $C_1 > 0$ such that μ satisfies*

$$\overline{\mathcal{T}}_2(\nu|\mu) \leq C_1 H(\nu|\mu), \quad \forall \nu \in \mathcal{P}_1(\mathbb{R}^n).$$

- (2) *There exists $C_2 > 0$ such that for all convex Lipschitz functions $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ bounded from below,*

$$\int e^{-\varphi/C_2} d\mu \leq e^{-\int Q_1\varphi/C_2 d\mu}.$$

- (3) There exist $C_3 > 0$ and $\lambda > 0$, such that for all concave Lipschitz functions $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$, bounded from below and $\lambda \|\cdot\|^2$ -convex,

$$\text{Ent}_\mu(e^\psi) \leq C_3 \int \|\nabla \psi\|_*^2 e^\psi d\mu,$$

where $\|\cdot\|_*$ is the dual norm $\|\cdot\|$ on \mathbb{R}^n .

Recall that if $\|\cdot\| = |\cdot|$ is the Euclidean norm, a function ψ is $\lambda \|\cdot\|^2$ -convex if and only if its Hessian is bounded from below by $-2\lambda I$ (in the sense of quadratic forms).

Let us observe that in Theorem 4.2, point (1) is equivalent to point (2) with the same constant $C_1 = C_2$.

Theorem 4.3. (see Theorem 8.8 [GRST14b]) *Let $\mu \in \mathcal{P}_1(\mathbb{R}^n)$. The following properties are equivalent.*

- (1) There exists $C_1 > 0$ such that μ satisfies

$$\overline{T}_2(\mu|\nu) \leq C_1 H(\nu|\mu), \quad \forall \nu \in \mathcal{P}_1(\mathbb{R}^n).$$

- (2) There exists $C_2 > 0$ such that for all Lipschitz convex functions $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$, bounded from below,

$$\int e^{Q_1 \varphi / C_2} d\mu \leq e^{\int \varphi / C_2 d\mu}.$$

- (3) There exists $C_3 > 0$ such that for all Lipschitz convex functions $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$, bounded from below,

$$\text{Ent}_\mu(e^\varphi) \leq C_3 \int \|\nabla \varphi\|_*^2 e^\varphi d\mu,$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$ on \mathbb{R}^n .

- (4) There exists $C_4 > 0$ such that for all Lipschitz convex functions $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$, bounded from below,

$$\|e^{Q_t \varphi}\|_{a+t/(2C_4), (\mu)} \leq \|e^\varphi\|_{a, (\mu)}, \quad \forall t > 0,$$

where for all $h : \mathbb{R}^n \rightarrow \mathbb{R}$, $p \geq 0$, $\|h\|_{p, (\mu)} = (\int |h|^p d\mu)^{1/p}$.

Let us observe that in this theorem, point (1) is equivalent to point (2) with the same constant $C_1 = C_2$, and point (3) is equivalent to point (4) with the same constant $C_3 = C_4$. The other links between the constants in the two last theorems are given in [GRST14b].

The logarithmic Sobolev inequality restricted to the class of convex functions of point (3) has been investigated in [Ada05, AS15], where sufficient conditions on the probability measure μ are given for such an inequality to hold.

4.2. Barycentric transport inequalities for the binomial law and the Poisson measure. Discrete probability measures do not generally satisfy the Talagrand's transport inequality \mathbf{T}_2 . To be convinced, it suffices to consider μ_ρ , a convex combination of two Dirac measures at two distinct points a and b , $\mu_\rho =$

$\rho\delta_a + (1-\rho)\delta_b$, $\rho \in (0, 1)$. The measure μ_ρ satisfies $\mathbf{T}_2(C)$, $C > 0$ if and only if for all $q \in [0, 1]$

$$W_2^2(\mu_\rho, \mu_q) = \mathcal{T}_2(\mu_\rho, \mu_q) = |a-b|^2|q-\rho| \leq CH(\mu_q|\mu_\rho) = q \log \frac{q}{\rho} + (1-q) \log \frac{1-q}{1-\rho}.$$

We get a contradiction as q goes to ρ by observing that $H(\mu_q|\mu_\rho) = o(|q-\rho|)$.

However the measure μ_ρ satisfies the transport inequality

$$W_1^2(\mu_\rho, \nu) \leq \frac{d(a,b)^2}{2} H(\nu, \mu_\rho), \quad \forall \nu \ll \mu_\rho.$$

To get other transport inequalities, one strategy is to replace the usual transport cost by a barycentric cost. To simplify, let μ_ρ be the Bernoulli measure of parameter $\rho \in (0, 1)$ ($a=0$ and $b=1$). In [Sam03] and [GRST14b] (see Proposition 7.1), the barycentric transport inequality $\mathbf{T}_c(1/(1-t), 1/t)$, $t \in (0, 1)$ with cost

$$c(x, p) = \theta_{\rho,t} \left(x - \int y dp(y) \right), \quad x \in \mathbb{R}, \quad p \in \mathcal{P}(\mathbb{R}),$$

is established for the Bernoulli measure μ_ρ , with an optimal cost function $\theta_{\rho,t}$: one has,

$$\bar{\mathcal{T}}_{\theta_{\rho,t}}(\nu_1|\nu_2) \leq \frac{1}{1-t} H(\nu_1|\mu_\rho) + \frac{1}{t} H(\nu_2|\mu_\rho), \quad \forall \nu_1, \nu_2 \ll \mu_\rho.$$

By tensorization, it provides a barycentric transport inequality for the product measure $\mu_\rho^n = \mu_\rho \otimes \cdots \otimes \mu_\rho$ associated to the cost

$$c^n(x, p) = c(x_1, p_1) + \cdots + c(x_n, p_n), \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n, p \in \mathcal{P}_1(\mathbb{R}^n).$$

By projection, and observing that by convexity

$$c^n(x, p) \geq n c \left(\frac{\sum_i^n x_i}{n}, \frac{\sum_i^n p_i}{n} \right),$$

a barycentric transport inequality follows for the binomial law $\mu_{n,\rho}$ with parameters n and ρ (see [GRST14b], Corollary 7.7),

$$\bar{\mathcal{T}}_{\theta_{\rho,t,n}}(\nu_1|\nu_2) \leq \frac{1}{1-t} H(\nu_1|\mu_{n,\rho}) + \frac{1}{t} H(\nu_2|\mu_{n,\rho}), \quad \forall \nu_1, \nu_2 \ll \mu_{n,\rho},$$

where $\theta_{\rho,t,n}(h) = n\theta_{\rho,t}(h/n)$, $h \in \mathbb{R}$. Finally, by the weak convergence of the measure μ_{n,ρ_n} towards the Poisson measure p_λ with parameter $\lambda > 0$ when $\rho_n = \lambda/n$, an optimal barycentric transport inequality is obtained for the Poisson measure

$$\bar{\mathcal{T}}_{c_{\lambda,t}}(\nu_1|\nu_2) \leq \frac{1}{1-t} H(\nu_1|p_\lambda) + \frac{1}{t} H(\nu_2|p_\lambda), \quad \forall \nu_1, \nu_2 \ll p_\lambda,$$

with $c_{\lambda,t}(h) = \lim_{n \rightarrow \infty} n\theta_{\rho_n,t}(h/n)$, $h \in \mathbb{R}$. One specific feature of the cost function $c_{\lambda,t}$ is to be zero for $h \geq 0$. For more details, we refer to Proposition 7.11 [GRST14b].

One other famous strategy to establish transport inequalities in discrete setting is coming from the notion of curvature on discrete spaces introduced by Maas [Maa11]. Transport inequalities for invariant reversible measures of Markov chains are obtained from curvature type conditions (see also [EM12, EM14, EMT15]). The optimal transport cost is defined by an abstract Benamou-Brenier type formula which is not associated to a transport cost function. This optimal cost is not comparable to a barycentric cost.

In an other direction, transport inequalities for Poisson processes of different types are proposed by Ma and al. in [MSWW11].

4.3. Optimal transport coupling for barycentric costs on \mathbb{R} . This part concerns the construction of an optimal coupling π^* that optimizes the optimal barycentric cost on the real line (in dimension one).

The cost function $\theta : \mathbb{R} \rightarrow \mathbb{R}$ is assumed to be even (we believe that this condition can be removed). For any probability measure μ , we denote by F_μ its cumulative distribution function, $F_\mu(x) = \mu(-\infty, x]$, and by F_μ^{-1} its general inverse

$$F_\mu^{-1}(u) = \inf\{x \in \mathbb{R}, F_\mu(x) \geq u\}, \quad u \in [0, 1].$$

Let μ and γ be two probability measures on \mathbb{R} . Assume that μ has no atoms, then it is well-known that for all convex cost functions θ , the usual optimal transport cost $T_\theta(\mu, \gamma)$ is reached for the optimal deterministic coupling measure

$$d\pi^*(x, y) = d\mu(x)d\delta_{S^*(x)}(y),$$

where S^* is the monotone transport map defined by $S^*(x) = F_\gamma^{-1} \circ F_\mu(x)$, $x \in \mathbb{R}$. In other words, there exists a monotone transport map S^* , independent of θ and such that

$$\mathcal{T}_\theta(\mu, \gamma) = \int \theta(x - S^*(x)) d\mu(x).$$

We want the same kind of result of independence of the function θ for an optimal coupling of the barycentric cost $\bar{\mathcal{T}}_\theta(\nu|\mu)$. For that purpose, we will use the following preliminary result of [GRS⁺15] (Theorem 1.3).

Theorem 4.4. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. There exists $\hat{\gamma} \in \mathcal{P}_1(\mathbb{R})$ such that $\hat{\gamma} \preceq \nu$ and for any even convex function θ , one has*

$$\bar{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\hat{\gamma}, \mu).$$

This result is still available when μ has atoms and it seems that the even condition on θ can be removed.

Based on the facts set out above, if μ has no atoms,

$$\bar{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\hat{\gamma}, \mu) = \int \theta(x - S^*(x)) d\mu(x), \quad \text{with } S^* = F_{\hat{\gamma}}^{-1} \circ F_\mu,$$

and since $\hat{\gamma} \preceq \nu$, according to Strassen's Theorem 4.1, there exists a kernel p^* such that $\hat{\gamma}p^* = \nu$ and $\hat{\gamma}$ -almost surely $\int y dp_x^*(y) = x$. Since $\hat{\gamma}$ is independent of θ , le kernel p^* is also independent of θ . Moreover, since $S^* \# \mu = \hat{\gamma}$, we get for μ -almost every x ,

$$\int y dp_{S^*(x)}^*(y) = S^*(x),$$

and it finally gives

$$\bar{\mathcal{T}}_\theta(\nu|\mu) = \int \theta\left(x - \int y dp_{S^*(x)}^*(y)\right) d\mu(x).$$

This shows that if μ has no atoms, then the optimal barycentric cost $\bar{\mathcal{T}}_\theta(\nu|\mu)$ is reached for the optimal coupling

$$\pi^*(dx, dy) = \mu(dx)p_{S^*(x)}^*(dy),$$

which is independent of the even convex function θ .

4.4. Characterization of probability measures satisfying a barycentric transport inequality on \mathbb{R} . We know how to characterize the probability measures on \mathbb{R} satisfying different functional inequalities as the Poincaré inequality [Muc72], the logarithmic Sobolev inequality [BG99, BR03] (see also chapter 6, [ABC⁺00]), the usual transport inequalities [Goz12]. In each of these cases, the characterization can be given by criteria of Hardy type, on the tails of distribution and on the densities of the involved measures. This section concerns the characterization of the barycentric transport inequalities. The approach is the one introduced by Gozlan [Goz12].

Let τ be the exponential law on \mathbb{R} , with density $e^{-|x|}/2$. For any $\mu \in \mathcal{P}(\mathbb{R})$, let us note U_μ the unique left-continuous monotone transport map from the measure τ to the measure μ , $U_\mu = F_\mu^{-1} \circ F_\tau$, namely,

$$U_\mu(x) = \begin{cases} F_\mu^{-1} \left(1 - \frac{1}{2} e^{-|x|} \right) & \text{if } x \geq 0, \\ F_\mu^{-1} \left(\frac{1}{2} e^{-|x|} \right) & \text{if } x \leq 0. \end{cases}$$

Here is one of the main result of [Goz12]: a probability measure μ satisfies the transport inequality $\mathbf{T}_2(C)$ with $C > 0$ if and only if it satisfies the Poincaré inequality and the following condition: there exists $b \geq 0$ such that

$$\sup_{x \in \mathbb{R}} (U_\mu(x+u) - U_\mu(x)) \leq b \sqrt{1+u}, \quad \forall u \geq 0,$$

that enforces a particular behavior of the tails of distribution of the measure μ .

In [GRS⁺15], an analogous result is obtained for the barycentric transport inequality with costs \bar{T}_θ . Let us note $\theta^a(t) = \theta(at)$, $t \in \mathbb{R}$, for any $a > 0$.

Theorem 4.5. (Theorem 1.2, [GRS⁺15]) *Let θ be an even convex function such that $\theta(t) = t^2$ for all $|t| \leq t_0$, $t_0 > 0$. A probability measure μ satisfies the barycentric transport cost inequalities*

$$\bar{\mathbf{T}}_{\theta^a}^- : \quad \bar{T}_{\theta^a}(\mu|\nu) \leq H(\nu|\mu), \quad \forall \nu \in \mathcal{P}_1(\mathbb{R}),$$

and

$$\bar{\mathbf{T}}_{\theta^a}^+ : \quad \bar{T}_{\theta^a}(\nu|\mu) \leq H(\nu|\mu), \quad \forall \nu \in \mathcal{P}_1(\mathbb{R}),$$

for some positive constant a if and only if there exists $b \geq 0$ such that

$$(16) \quad \sup_x (U_\mu(x+u) - U_\mu(x)) \leq b \theta^{-1}(u + t_0^2) \quad \forall u \geq 0.$$

Measures satisfying a barycentric transport inequality do not necessarily verify a Poincaré inequality since their support is not necessarily connected (for example the Bernoulli and the binomial laws as explained in Section 4.2). However, the condition (16) of the above theorem implies for $u = 1$: there exists $h > 0$ such that

$$(17) \quad \sup_{x \in \mathbb{R}} (U_\mu(x+1) - U_\mu(x)) \leq h.$$

Bobkov and Götze [BG99] have proved that this condition is equivalent to the existence of a Poincaré inequality restricted to convex functions satisfied by the measure μ . Therefore, probability measures satisfying (16) necessarily satisfy a so-called “convex” Poincaré inequality. More precisely, the following result has been

established by Feldheim and al. [FMNW15], and also independently by Gozlan and al. [GRS⁺15], as an intermediate key result of their proof of Theorem 4.5.

Theorem 4.6. *Let μ be a probability measure on \mathbb{R} . The condition (17) is equivalent to each of the following properties.*

(a) *There exists $C > 0$ such that for all convex functions f on \mathbb{R} ,*

$$\mathrm{Var}_\mu(f) \leq C \int_{\mathbb{R}} f'^2 d\mu.$$

(b) *There exist $a, t_0 > 0$ such that*

$$\overline{\mathbf{T}}_{\theta_1^-} : \quad \overline{\mathcal{T}}_{\theta_1^-}(\mu|\nu) \leq H(\nu|\mu), \quad \forall \nu \in \mathcal{P}_1(\mathbb{R}),$$

and

$$\overline{\mathbf{T}}_{\theta_1^+} : \quad \overline{\mathcal{T}}_{\theta_1^+}(\nu|\mu) \leq H(\nu|\mu), \quad \forall \nu \in \mathcal{P}_1(\mathbb{R}),$$

where the function θ_1 is defined by $\theta_1(t) = \begin{cases} t^2 & \text{if } |t| \leq t_0, \\ 2|t|t_0 - t_0^2 & \text{if } |t| > t_0. \end{cases}$

5. UNIVERSAL TRANSPORT INEQUALITIES

We call universal any transport inequality that holds for any (reference) probability measure μ on \mathcal{X} .

The most popular and commonly used universal transport inequality, mentioned in Section 3.1, is the Csiszár-Kullback-Pinsker inequality [Csi67, Kul67, Pin64],

$$(18) \quad \frac{1}{2} \|\mu - \nu\|_{TV}^2 \leq H(\nu|\mu), \quad \forall \mu, \nu \in \mathcal{P}(\mathcal{X}),$$

where $\|\mu - \nu\|_{TV}$ is the total variation distance between μ and ν ,

$$\|\mu - \nu\|_{TV} = 2 \inf_{\pi \in \Pi(\mu, \nu)} \iint \mathbb{1}_{x \neq y} d\pi(x, y).$$

The functional dual formulation of the Csiszár-Kullback-Pinsker inequality is the following exponential inequality, for any function $f : \mathcal{X} \rightarrow \mathbb{R}$ such that $\sup_{x, y \in \mathcal{X}} |f(x) - f(y)| \leq c$,

$$\int e^{tf} d\mu \leq e^{t \int f d\mu + t^2 c^2 / 8}, \quad t \geq 0.$$

This inequality, commonly used, gives the Hoeffding inequality by applying Markov's inequality,

$$\mu \left(f \geq \int f d\mu + t \right) \leq e^{-2t^2/c^2} \quad t \geq 0.$$

The Csiszár-Kullback-Pinsker inequality (18) can be improved in different ways. We may change the function of the total variation distance on the left-hand side (see [FHT03, Gil10]), or we may replace the total variation distance by a comparable optimal weak transport cost of Definition 2.4. More precisely, given a convex function $\alpha : \mathbb{R}^+ \rightarrow [0, +\infty]$, and $\mu, \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X})$, we note

$$\widetilde{\mathcal{T}}_\alpha(\nu_1|\nu_2) = \inf_{\pi \in \Pi(\nu_2, \nu_1)} \left\{ \int c(x, p_x) d\nu_2(x), d\pi(x, y) = d\nu_2(x) dp_x(y) \right\},$$

when $c(x, p) = \alpha \left(\int \mathbb{1}_{x \neq y} dp(y) \right)$, $x \in \mathcal{X}$, $p \in \mathcal{P}(\mathcal{X})$, and

$$\widehat{\mathcal{T}}_\alpha(\nu_1|\nu_2) = \inf_{\pi \in \Pi(\nu_2, \nu_1)} \left\{ \int c(x, p_x) d\nu_2(x), d\pi(x, y) = d\nu_2(x) dp_x(y) \right\},$$

when

$$c(x, p) = \int \alpha \left(\mathbb{1}_{x \neq y} \frac{dp}{d\mu}(y) \right) d\mu(y),$$

if (x, p) is such that p is absolutely continuous with respect to μ on $\mathcal{X} \setminus \{x\}$, and $c(x, p) = +\infty$ otherwise. In [Sam07], Theorem 1.1 and 1.2 give the following variants of the Csizár-Kullback-Pinsker inequality.

Theorem 5.1. [Sam07, GRST14b] *Let \mathcal{X} be a compact metric space, $\mu \in \mathcal{P}(\mathcal{X})$ and $t \in (0, 1)$.*

(a) *For any probability measures ν_1 and ν_2 on \mathcal{X} , one has*

$$\widetilde{\mathcal{T}}_{\alpha_t}(\nu_1|\nu_2) \leq \frac{1}{1-t} H(\nu_1|\mu) + \frac{1}{t} H(\nu_2|\mu),$$

where the convex cost function $\alpha_t : \mathbb{R}^+ \rightarrow [0, +\infty]$ is defined by

$$\alpha_t(u) = \frac{t(1-u) \log(1-u) - (1-tu) \log(1-tu)}{t(1-t)}, \quad 0 \leq u \leq 1,$$

and $\alpha_t(u) = +\infty$ if $u > 1$.

As t goes to 0, it implies

$$\widetilde{\mathcal{T}}_{\alpha_0}(\nu_1|\mu) \leq H(\nu_1|\mu),$$

with $\alpha_0(u) = (1-u) \log(1-u) + u$ if $0 \leq u \leq 1$, and $\alpha_0(u) = +\infty$ if $u > 1$.

As t goes to 1, it implies

$$\widetilde{\mathcal{T}}_{\alpha_1}(\mu|\nu_2) \leq H(\nu_2|\mu),$$

with $\alpha_1(u) = -\log(1-u) - u$, if $0 \leq u < 1$ and $\alpha_1(u) = +\infty$ if $u \geq 1$.

(b) *For any probability measures ν_1 and ν_2 on \mathcal{X} , one has*

$$\widehat{\mathcal{T}}_{\beta_t}(\nu_1|\nu_2) \leq \frac{1}{1-t} H(\nu_1|\mu) + \frac{1}{t} H(\nu_2|\mu),$$

where the convex cost function $\beta_t : \mathbb{R}^+ \rightarrow [0, +\infty]$ is defined by

$$\beta_t(u) := \sup_{s \in \mathbb{R}^+} \{su - \beta_t^*(s)\}, \quad u \in \mathbb{R}^+,$$

with

$$\beta_t^*(s) = \frac{te^{(1-t)s} + (1-t)e^{-ts} - 1}{t(1-t)}, \quad s \in \mathbb{R}^+.$$

When t goes to 0 this implies

$$(19) \quad \widehat{\mathcal{T}}_{\beta_0}(\nu_1|\mu) \leq H(\nu_1|\mu),$$

with $\beta_0(u) = (1+u) \log(1+u) - u$, $u \geq 0$, and when t goes to 1, it implies

$$\widehat{\mathcal{T}}_{\beta_1}(\mu|\nu_2) \leq H(\nu_2|\mu),$$

with $\beta_1(u) = (1-u) \log(1-u) + u$, if $u \leq 1$ and $\beta_1(u) = +\infty$ if $u > 1$.

By using the estimate $\alpha_t(u) \geq u^2/2 = \alpha(u)$, for all $t \in [0, 1]$, $u \geq 0$, the transport inequalities of point (a) provide the Marton's transport inequalities [Mar96b] associated to the quadratic cost function α :

$$(20) \quad \widetilde{\mathcal{T}}_2(\nu_1|\mu) \leq 2H(\nu_1|\mu), \quad \widetilde{\mathcal{T}}_2(\mu|\nu_2) \leq 2H(\nu_2|\mu),$$

or even for every $t \in (0, 1)$,

$$\frac{1}{2}\widetilde{\mathcal{T}}_2(\nu_1|\nu_2) \leq \frac{1}{1-t}H(\nu_1|\mu) + \frac{1}{t}H(\nu_2|\mu).$$

By optimizing in t , this inequality can be rewritten

$$\frac{1}{2}\widetilde{\mathcal{T}}_2(\nu_1|\nu_2) \leq \left(\sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)} \right)^2.$$

As explained in Section 3.3, these transport inequalities can be tensorised on product spaces, and provide concentration results for product measures, or even weakly dependent measures. The concentration principle related to this kind of n -dimensional costs have been introduced by Talagrand [Tal96a, Tal96c], especially to prove deviations inequalities for suprema of empirical processes of Bernstein type.

The optimal deviation results for suprema of empirical processes that follow from Theorem 5.1 are briefly recalled in Section 5.1 below. Then in Section 5.2, we summarise results obtained in a weak dependence framework concerning the Marton's transport inequalities (20) in [Sam00] and [Pau14]. Finally, in Section 5.3, we suggest a different way to tensorize the Csiszár-Kullback-Pinsker inequality or the Marton's inequality. It provides new weak transport inequalities for the uniform law on the symmetric group (see [Sam15]). These results are guided by the concentration results by Talagrand [Tal95].

5.1. Bernstein's type of deviation inequalities for suprema of independent empirical processes. The first Bernstein's type of deviation inequalities for suprema of independent empirical processes have been obtained by Talagrand [Tal96b, Tal96a] with the so-called "convex hull" method. These inequalities are of particular interest in statistics [Mas00b, Mas07].

Ledoux [Led97] has proposed an "entropic" method that allows to simply recover the results by Talagrand. This approach is based on the tensorization property of the entropy and the so called Herbst's argument. Then, it has been widely developed, mainly to reach optimal deviation bounds for the suprema of independent empirical processes [Mas00a, BLM00, Rio01, BLM03, Rio02, Bou03, KR05, Rio12, Rio13, BLM13].

In the continuation of the works by Talagrand, Marton [Mar96a, Mar96b], Dembo [Dem97] and Maurey [Mau91], the transport-entropy method has been developed in [Sam07] as an alternative approach to achieve the best constants in the deviation inequalities of suprema of empirical processes.

Another approach has been proposed by Panchenko, based on symmetrization technics [Pan01, Pan02, Pan03]. Finally, to complete the picture, Stein's methods have been pushed forward by Chatterjee to reach similar concentration properties to the one by Talagrand [Cha05, Cha07, CD10, Pau14]. The main interest of this last method is that it extends to dependence cases, under Dobrushin type of conditions.

Let us present some concentration results for suprema of empirical processes that follow from the transport inequalities of Theorem 5.1 after tensorization. Let \mathcal{F} be a countable set and let $(X_{1,t})_{t \in \mathcal{F}}, \dots, (X_{n,t})_{t \in \mathcal{F}}$ be n independent processes. We are interested by the deviations of the random variable

$$Z = \sup_{t \in \mathcal{F}} \sum_{i=1}^n X_{i,t}.$$

Let us note

$$V = \sup_{t \in \mathcal{F}} \sum_{i=1}^n \mathbb{E} \left[[X_{i,t} - X'_{i,t}]^2 \mid X_{i,t} \right],$$

where $(X'_{i,t})_{t \in \mathcal{F}}$ is an independent copy of $X_i = (X_{i,t})_{t \in \mathcal{F}}$ and $\mathbb{E}[\cdot \mid X_{i,t}]$ denotes the conditional expectation, given $X_{i,t}$. In the following theorem, for all $t \in \mathcal{F}$ and all $i \in \{1, \dots, n\}$, $M_{i,t}$ and $m_{i,t}$ are numerical constants limiting the random variables $X_{i,t}$.

Theorem 5.2. (a) Assume that $X_{i,t} \leq M_{i,t}$, and $\mathbb{E}[(M_{i,t} - X_{i,t})^2] \leq 1$, for all i and all t , then for all $u \geq 0$,

$$\begin{aligned} \mathbb{P}(Z \geq \mathbb{E}[Z] + u) &\leq \exp \left[-\frac{u}{2 \left(1 + \varepsilon \left(\frac{u}{\mathbb{E}[V]}\right)\right)} \log \left(1 + \frac{u}{\mathbb{E}[V]}\right) \right] \\ &\leq \exp \left[-\frac{u^2}{2\mathbb{E}[V] + 2u} \right], \end{aligned}$$

with $\varepsilon(u) = \frac{\beta_0(u)}{(1+u) \log(1+u)}$ and $\beta_0(u) := (1+u) \log(1+u) - u$.

(b) Assume that $m_{i,t} \leq X_{i,t} \leq M_{i,t}$, with $M_{i,t} - m_{i,t} = 1$ for all i and all t , then for all $u \geq 0$,

$$\mathbb{P}(Z \leq \mathbb{E}[Z] - u) \leq \exp \left[-\mathbb{E}[V] \beta_0 \left(\frac{u}{\mathbb{E}[V]} \right) \right] \leq \exp \left[-\frac{u^2}{2\mathbb{E}[V] + \frac{2}{3}u} \right],$$

with $\beta_0(u) = (1+u) \log(1+u) - u$.

The optimality of these results is discussed in [Sam07]. Recall that by usual symmetrization's technics ([LT91], Lemma 6.3 and Theorem 4.12), the variance term $\mathbb{E}[V]$ can be estimated as follows

$$\mathbb{E}[V] \leq \sup_{t \in \mathcal{F}} \sum_{i=1}^n \text{Var}(X_{i,t}) + 16 \mathbb{E} \left[\sup_{t \in \mathcal{F}} \left| \sum_{i=1}^n (X_{i,t} - \mathbb{E}[X_{i,t}]) \right| \right].$$

In [Ada08], by using Hoffman-Jørgensen's inequality and some other results by Talagrand, Adamczak extends the concentration properties to suprema of unbounded random variables, by truncation arguments of the random variables.

Idea of the proof. We only present one elementary proof of (b) to show the links between the transport inequality with cost $\widehat{\mathcal{T}}_{\beta_t}$ and the deviations of a function around its mean.

Let μ_i denote the law of the process $X_i = (X_{i,t})_{t \in \mathcal{F}}$. Point (b) simply follows from the dual form of the tensorized transport inequality (19): $\mathcal{X} = \mathbb{R}^{\mathcal{F}}$, for any

function $g : \mathcal{X}^n \rightarrow \mathbb{R}$,

$$(21) \quad \int e^{-g} d\mu \leq \exp\left(-\int \widehat{Q}g d\mu\right),$$

where $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ and for all $x = (x_1, \dots, x_n) \in \mathcal{X}^n$

$$\widehat{Q}g(x) = \inf_{p \in \mathcal{P}(\mathcal{X}^n)} \left\{ \int g(y) dp(y) + \sum_{i=1}^n \int \beta_0 \left(\mathbb{1}_{x_i \neq y_i} \frac{dp_i}{d\mu_i}(y_i) \right) d\mu_i(y_i) \right\},$$

the probability measures p_i are the marginals of p . Let us choose $g = \lambda f$ with $\lambda \geq 0$ and

$$f(x) = \sup_{t \in \mathcal{F}} \sum_{i=1}^n x_{i,t}, \quad x = (x_1, \dots, x_n) \in \mathcal{X}^n.$$

To simplify, we assume that for all $x \in \mathcal{X}^n$, the supremum is reached at a single point $\tau(x) \in \mathcal{F}$:

$$\sup_{t \in \mathcal{F}} \sum_{i=1}^n x_{i,t} = \sum_{i=1}^n x_{i,\tau(x)}.$$

Then for all $x, y \in \mathcal{X}^n$

$$f(y) \geq f(x) + \sum_{i=1}^n (y_{i,\tau(x)} - x_{i,\tau(x)}) = f(x) + \sum_{i=1}^n (y_{i,\tau(x)} - x_{i,\tau(x)}) \mathbb{1}_{x_i \neq y_i}.$$

as a consequence, for all x ,

$$\begin{aligned} & \widehat{Q}g(x) \\ & \geq \lambda f(x) - \sup_p \left\{ \int \sum_{i=1}^n \lambda (x_{i,\tau(x)} - y_{i,\tau(x)}) \mathbb{1}_{x_i \neq y_i} dp(y) - \sum_{i=1}^n \int \beta_0 \left(\mathbb{1}_{x_i \neq y_i} \frac{dp_i}{d\mu_i}(y_i) \right) d\mu_i(y_i) \right\} \\ & = \lambda f(x) - \sum_{i=1}^n \sup_{p_i \in \mathcal{P}(\mathcal{X})} \left\{ \int \lambda (x_{i,\tau(x)} - y_{i,\tau(x)}) \mathbb{1}_{x_i \neq y_i} \frac{dp_i}{d\mu_i}(y_i) d\mu_i(y_i) - \int \beta_0 \left(\mathbb{1}_{x_i \neq y_i} \frac{dp_i}{d\mu_i}(y_i) \right) d\mu_i(y_i) \right\} \\ & \geq \lambda f(x) - \sum_{i=1}^n \int \sup_{h \geq 0} \{ \lambda (x_{i,\tau(x)} - y_{i,\tau(x)}) h - \beta_0(h) \} d\mu_i(y_i) \\ & = \lambda f(x) - \sum_{i=1}^n \int \beta_0^* (\lambda [x_{i,\tau(x)} - y_{i,\tau(x)}]_+) d\mu_i(y_i) \\ & \geq \lambda f(x) - \beta_0^*(\lambda) \sum_{i=1}^n \int [x_{i,\tau(x)} - y_{i,\tau(x)}]_+^2 d\mu_i(y_i) \\ & \geq \lambda f(x) - \beta_0^*(\lambda) \sup_{t \in \mathcal{F}} \sum_{i=1}^n \int [x_{i,t} - y_{i,t}]_+^2 d\mu_i(y_i), \end{aligned}$$

where $\beta_0^*(s) = e^s - s - 1$, $s \geq 0$. The second last inequality is a consequence of the fact that $[x_{i,\tau(x)} - y_{i,\tau(x)}]_+ \leq M_{i,t} - m_{i,t} \leq 1$ and $\beta_0^*(\lambda u) \leq u^2 \beta_0^*(\lambda)$ for $0 \leq u \leq 1$. By inserting the previous estimate of $\widehat{Q}g(x)$ in the transport inequality (21), we get for all $\lambda \geq 0$,

$$\mathbb{E} [e^{-\lambda Z}] \leq e^{-\lambda \mathbb{E}[Z] + \mathbb{E}[V] \beta_0^*(\lambda)}.$$

The deviation inequality of (b) directly follows by the Markov inequality, optimizing over all $\lambda \geq 0$. \square

5.2. Marton's transport inequality for weakly dependent random variables. The paper [Sam00] presents a tensorization scheme of the Marton's inequality (20) when μ is a probability measure on the product space \mathcal{X}^n , whose marginals are weakly dependent of each other, more precisely if μ is the law the n first random variables X_1, \dots, X_n of a Φ -mixing process. This tensorization scheme is based on the construction of couplings similar to the one of [Mar03].

To simplify, one may assume that \mathcal{X} is a finite set. Given a sequence of random variables X_1, \dots, X_n with values in \mathcal{X} , for $1 \leq i < j \leq n$, let us note

$$\mathcal{L}(X_j | X_1^{i-1} = x_1^{i-1}, X_i = x_i)$$

the law of (X_j, \dots, X_n) knowing that $X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i = x_i$, and let $\Gamma = (\gamma_{i,j})_{1 \leq i, j \leq n}$ be the upper triangular matrix defined by

$$\gamma_{ij}^2 = \sup_{x_1^{i-1}, x_i, y_i} \|\mathcal{L}(X_j | X_1^{i-1} = x_1^{i-1}, X_i = x_i) - \mathcal{L}(X_j | X_1^{i-1} = x_1^{i-1}, X_i = y_i)\|_{TV},$$

for $i < j$ and $\gamma_{ii} = 1$.

Theorem 5.3. [Sam00] *According to the previous notations, for all probability measures μ and ν on \mathcal{X}^n , one has*

$$\widetilde{\mathcal{T}}_2(\nu|\mu) \leq 2\|\Gamma\|^2 H(\nu|\mu) \quad \text{and} \quad \widetilde{\mathcal{T}}_2(\mu|\nu) \leq 2\|\Gamma\|^2 H(\nu|\mu),$$

where $\|\Gamma\|$ denotes the operator norm of the matrix Γ from $(\mathbb{R}^n, |\cdot|)$ to $(\mathbb{R}^n, |\cdot|)$.

Note that since the Marton's cost $\widetilde{\mathcal{T}}_2$ in dimension n also satisfies the triangular inequality (10), the two weak transport inequalities of this theorem are equivalent to the following family of transport inequalities, for all $t \in (0, 1)$, for all $\mu, \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}^n)$,

$$\frac{1}{2\|\Gamma\|^2} \widetilde{\mathcal{T}}_2(\nu_1|\nu_2) \leq \frac{1}{1-t} H(\nu_1|\mu) + \frac{1}{t} H(\nu_2|\mu),$$

or equivalently, applying Theorem 3.1 for all functions $g : \mathcal{X}^n \rightarrow \mathbb{R}$ bounded from below,

$$\left(\int e^{t\widetilde{Q}g} d\mu \right)^{1/t} \left(\int e^{-(1-t)g} d\mu \right)^{1/(1-t)} \leq 1,$$

where

$$\widetilde{Q}g(x) = \inf_{p \in \mathcal{P}(\mathcal{X}^n)} \left\{ \int g dp + \frac{1}{2\|\Gamma\|^2} \sum_{i=1}^n \left(\int \mathbb{1}_{x_i \neq y_i} dp(y) \right)^2 \right\}.$$

In particular, applying this inequality to the function $g = i_A$, with $A \subset \mathcal{X}^n$, we get the Talagrand's concentration result extended to any measure $\mu \in \mathcal{P}(\mathcal{X}^n)$: for all subsets $A \subset \mathcal{X}^n$, for all $t \in (0, 1)$,

$$(22) \quad \int e^{\frac{t}{2\|\Gamma\|^2} D_T^2(x, A)} d\mu(x) \leq \frac{1}{\mu(A)^{t/(1-t)}},$$

where $D_T(x, A)$ is the Talagrand's convex distance defined by

$$\begin{aligned} D_T^2(x, A) &= \left(\sup_{\alpha \in \mathbb{R}^n, |\alpha| \leq 1} \inf_{y \in A} \sum_{i=1}^n \alpha_i \mathbb{1}_{x_i \neq y_i} \right)^2 \\ &= \left(\sup_{\alpha \in \mathbb{R}^n, |\alpha| \leq 1} \inf_{p \in \mathcal{P}(A)} \sum_{i=1}^n \alpha_i \int \mathbb{1}_{x_i \neq y_i} dp(y) \right)^2 \\ &= \left(\inf_{p \in \mathcal{P}(A)} \sup_{\alpha \in \mathbb{R}^n, |\alpha| \leq 1} \sum_{i=1}^n \alpha_i \int \mathbb{1}_{x_i \neq y_i} dp(y) \right)^2 \\ &= \inf_{p \in \mathcal{P}(A)} \sum_{i=1}^n \left(\int \mathbb{1}_{x_i \neq y_i} dp(y) \right)^2 = \tilde{Q}i_A(x). \end{aligned}$$

The second equality follows from the linearity of the expression in p and from the fact that Dirac measures are the extremal points of the convex set $\mathcal{P}(A)$. The third equality is a consequence of Sion's minimax Theorem [Sio58, Kom88] since the expression is linear in p and α , and therefore convex in p and concave in α .

When μ is the law of some independent random variables, the property (22) exactly recovers Talagrand's concentration results [Tal95] since $\|\Gamma\| = \|\text{Id}\| = 1$.

Theorem 5.3 complements the results by Marton for contracting Markov chains [Mar96b, Mar97]. More generally, assume that the sequence $(X_k)_{k \geq 1}$ is a Doeblin recurrent Markov chain with kernel K ; in other words, there exists a probability measure m , an integer r and a real $\rho \in (0, 1)$ such that for all $x \in \mathcal{X}$ and all subsets $A \subset \mathcal{X}$

$$K^r(x, A) \geq \rho m(A).$$

Then the coefficient $\|\Gamma\|$ is bounded independently of n ,

$$\|\Gamma\| \leq \frac{\sqrt{2}}{1 - \rho^{1/2r}}.$$

In any case, if $(\Phi_k)_{k \geq 1}$ represents the sequence of Φ -mixing coefficients of the sequence of random variables X_1, \dots, X_n (see [Sam00], [Dou94]), then one has

$$\|\Gamma\| \leq \sum_{k=1}^n \sqrt{\Phi_k}.$$

After this result, many authors have obtained transport inequalities under weak different dependence assumptions, for example under Dobrushin type conditions [DGW04, Mar04, Mar10, Wu06, Kon12, Pau14, WW14, Wan14, Pau14].

Among these results, we want to emphasize a result by Paulin [Pau14] that exactly concerns the Talagrand's concentration property (22) obtained by using Stein's methods, following Chatterjee's approach [Cha07].

For $1 \leq i \leq n$, let us note X_{-i} the random vector defined by

$$X_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n).$$

The Dobrushin's interdependence matrix $D = (d_{ij})$ is a matrix of non-negative entries such that for every $i \in \{1, \dots, n\}$, for every $x, y \in \mathcal{X}^n$,

$$\|\mathcal{L}(X_i|X_{-i} = x_{-i}) - \mathcal{L}(X_i|X_{-i} = y_{-i})\|_{TV} \leq \sum_{j,j \neq i} d_{ij} \mathbb{1}_{x_j \neq y_j}.$$

Theorem 5.4. (Theorem 3.3, [Pau14]) Let $\|D\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n d_{ij}$ and $\|D\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n d_{ij}$. If $\|D\|_1 < 1$ and $\|D\|_\infty \leq 1$, then for all subsets $A \subset \mathcal{X}^n$, one has

$$\int e^{\frac{1-\|D\|_1}{26.1} D_T^2(x,A)} d\mu(x) \leq \frac{1}{\mu(A)}.$$

Examples of applications of this concentration result are presented in [Pau14] (the stochastic travelling salesman problem, Steiner trees).

5.3. Transport inequalities for the uniform law on the symmetric group.

In this section we present transport inequalities for the uniform law on the symmetric group S_n , that provide concentration results obtained by Maurey [Mau79] and Talagrand [Tal95].

Let μ be the uniform law on S_n , $\mu(\sigma) = \frac{1}{n!}$, $\sigma \in S_n$.

Theorem 5.5. [Mau79] Let d be the Hamming distance on the symmetric group, for all $\sigma, \tau \in S_n$,

$$d(\sigma, \tau) = \sum_{i=1}^n \mathbb{1}_{\sigma(i) \neq \tau(i)}.$$

Then for any subset $A \subset S_n$ such that $\mu(A) \geq 1/2$, and for all $t \geq 0$, one has

$$\mu(A_t) \geq 1 - 2e^{-\frac{t^2}{64n}},$$

where $A_t = \{y \in S_n, d(x, A) \leq t\}$.

This result has been generalized by Milman and Schechtman, to some groups whose distance is invariant by translation [MS86]. Talagrand has obtained another concentration property, stronger in terms of the dependence in n , obtained by the so-called ‘‘convex-hull’’ method. Here is the Talagrand's property with slightly modified notations. This property implies the one of the previous theorem up to constant.

Theorem 5.6. [Tal95] For any subset $A \subset S_n$,

$$\int_{S_n} e^{f(A,\sigma)/16} d\mu(\sigma) \leq \frac{1}{\mu(A)},$$

where the quantity $f(A, \sigma)$ measures the distance from σ to A as follows

$$f(A, \sigma) = \inf_{p \in \mathcal{P}(A)} \sum_{i=1}^n \left(\int \mathbb{1}_{\sigma(i) \neq \tau(i)} dp(\tau) \right)^2.$$

This result has been first generalized to product of symmetric groups by McDiarmid [McD02], and then further by Luczak and McDiarmid, to cover more general permutation groups which act suitably ‘‘locally’’ [LM03].

Theorems 5.5 and 5.6 are consequences of the following transport inequalities.

Theorem 5.7. [Sam15] *Let μ be the uniform law on the symmetric group.*

(a) *For all probability measures ν_1 and ν_2 on S_n ,*

$$\frac{1}{2(n-1)} W_1^2(\nu_2, \nu_1) \leq \left(\sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)} \right)^2,$$

$$\text{where } W_1(\nu_1, \nu_2) = \inf_{\pi \in \Pi(\nu_2, \nu_1)} \iint d(\sigma, \tau) d\pi(\sigma, \tau).$$

(b) *For all probability measures ν_1 and ν_2 on S_n ,*

$$\frac{1}{8} \widehat{\mathcal{T}}_2(\nu_1|\nu_2) \leq \left(\sqrt{H(\nu_1|\mu)} + \sqrt{H(\nu_2|\mu)} \right)^2,$$

$$\text{where } \widehat{\mathcal{T}}_2(\nu_1|\nu_2) = \inf_{\pi \in \Pi(\nu_2, \nu_1)} \int \sum_{i=1}^n \left(\int \mathbb{1}_{\sigma(i) \neq \tau(i)} dp_\sigma(\tau) \right)^2 d\nu_2(\sigma),$$

$$\text{with } \pi(\sigma, \tau) = \nu_2(\sigma) p_\sigma(\tau).$$

The proof of (b), inspired from the Talagrand's results, is given in the preprint [Sam15]. We present a simpler proof of (a) of the same nature at the end of this section. In fact, the dual formulation of the transport inequality of (a) is more popular: for all 1-Lipschitz functions $f : S_n \rightarrow \mathbb{R}$ (with respect to the Hamming distance d),

$$\int e^f d\mu \leq e^{\int f d\mu + (n-1)t^2/2}, \quad \forall t \geq 0.$$

This exponential inequality is a consequence of Hoeffding inequalities for bounded martingales. It is widely commented and a proof is given in the paper [BHT06].

Point (b) implies the following useful concentration property.

Theorem 5.8. [Sam15] *Let μ be the uniform law on the symmetric group S_n . Let $g : S_n \rightarrow \mathbb{R}$ and $\alpha_k : S_n \rightarrow \mathbb{R}^+$, $k \in \{1, \dots, n\}$ be functions such that for all $\tau, \sigma \in S_n$,*

$$g(\tau) - g(\sigma) \leq \sum_{k=1}^n \alpha_k(\tau) \mathbb{1}_{\tau(k) \neq \sigma(k)}.$$

Then, for all $t \geq 0$, one has

$$\mu \left(g \leq \int g d\mu - t \right) \leq \exp \left(- \frac{t^2}{8 \int |\alpha|^2 d\mu} \right),$$

and

$$\mu \left(g \geq \int g d\mu + t \right) \leq \exp \left(- \frac{t^2}{8 \sup_{\sigma \in S_n} |\alpha(\sigma)|^2} \right),$$

$$\text{with } |\alpha(\sigma)|^2 = \sum_{k=1}^n \alpha_k^2(\sigma), \quad \sigma \in S_n.$$

By applying this result to the particular function $g(\sigma) = \varphi(x_\sigma)$ where $\varphi : [0, 1]^n \rightarrow \mathbb{R}$ is a Lipschitz convex function and given $(x_1, \dots, x_n) \in [0, 1]^n$, $x_\sigma = (x_{\sigma(1)}, \dots, x_{\sigma(n)})$, we recover the deviation inequality by Adamczak, Chafaï and Wolff [ACW14] (Theorem 3.1) obtained from Theorem 5.6 by Talagrand. This concentration property plays a key role in their approach, to study the convergence of the empirical spectral measure of random matrices with exchangeable entries, when the size of these matrices is increasing.

Proof of point (a) in Theorem 5.7. Since the distance W_1 satisfies a triangular inequality, it suffices to prove that for all probability measures ν_1 on S_n ,

$$\frac{1}{2(n-1)}W_1^2(\nu_1, \mu) \leq H(\nu_1|\mu).$$

According to Proposition 3.1, the dual formulation of this inequality is the following, for all function φ on S_n and all $\lambda \geq 0$,

$$(23) \quad \int e^{\lambda Q\varphi} d\mu \leq e^{\int \lambda\varphi d\mu + (n-1)\lambda^2/2},$$

with

$$Q\varphi(\sigma) = \inf_{p \in \mathcal{P}(S_n)} \left\{ \int \varphi dp + \int d(\sigma, \tau) dp(\tau) \right\} = \inf_{p \in \mathcal{P}(S_n)} \left\{ \int \varphi dp + \sum_{k=1}^n \int \mathbb{1}_{\sigma(k) \neq \tau(k)} dp(\tau) \right\}.$$

We will prove the inequality (23) by induction on n .

When $n = 2$, S_n is the two point space

$$Q\varphi(\sigma) = \inf_{p \in \mathcal{P}(S_n)} \left\{ \int \varphi dp + 2 \int \mathbb{1}_{\sigma \neq \tau} dp(\tau) \right\}.$$

The inequality (23) corresponds exactly to the dual form of the Csizár-Kullback-Pinsker inequality given by Proposition 3.1: for any probability measure ν on a separable metric space \mathcal{X} , for any measurable function $f : \mathcal{X} \rightarrow \mathbb{R}$,

$$(24) \quad \int e^{\lambda Rf} d\nu \leq e^{\lambda \int f d\nu + \lambda^2/2}, \quad \forall \lambda \geq 0,$$

with $Rf(x) = \inf_{p \in \mathcal{P}(\mathcal{X})} \left\{ \int f dp + 2 \int \mathbb{1}_{x \neq y} dp(y) \right\}$, $x \in \mathcal{X}$.

The induction step is also a consequence of the dual form (24) of the Csizár-Kullback-Pinsker inequality. Let $(H_i)_{1 \leq i \leq n}$ be the partition of S_n defined by,

$$H_i = \{\sigma \in S_n, \sigma(i) = n\}$$

If p is a probability measure on S_n , it admits a unique decomposition defined by

$$p = \sum_{i=1}^n \hat{p}(i)p_i, \quad \text{with } p_i \in \mathcal{P}(H_i) \quad \text{and} \quad \hat{p}(i) = p(H_i).$$

Thus, we define a probability measure \hat{p} on $\{1, \dots, n\}$. In particular, for the uniform law μ on S_n , one has

$$\mu = \frac{1}{n} \sum_{i=1}^n \mu_i,$$

where μ_i is the uniform law on H_i , $\mu_i(\sigma) = \frac{1}{(n-1)!}$, for any $\sigma \in H_i$. Therefore, one has

$$\int e^{\lambda Q\varphi} d\mu = \frac{1}{n} \sum_{i=1}^n \int e^{\lambda Q\varphi(\sigma)} d\mu_i(\sigma).$$

For any function $f : H_i \rightarrow \mathbb{R}$, let us note

$$Q_{H_i}f(\sigma) = \inf_{p \in \mathcal{P}(H_i)} \left\{ \int f dp + \sum_{k \neq i} \int \mathbb{1}_{\sigma(k) \neq \tau(k)} dp(\tau) \right\}.$$

We denote by τ_{ij} the transposition that exchanges the indices i and j . The application from H_i to H_n defined by $\tau \mapsto \tau\tau_{in}$ is one to one, and therefore by a change of index in the sum, we get

$$\begin{aligned} Q_{H_i} f(\sigma) &= \inf_{p \in \mathcal{P}(H_i)} \left\{ \int f(\tau) dp(\tau) + \sum_{k \neq n} \int \mathbb{1}_{\sigma\tau_{in}(k) \neq \tau\tau_{in}(k)} dp(\tau) \right\} \\ &= \inf_{q \in \mathcal{P}(H_n)} \left\{ \int f(\tau\tau_{in}) dq(\tau) + \sum_{k \neq n} \int \mathbb{1}_{\sigma\tau_{in}(k) \neq \tau(k)} dq(\tau) \right\} = Q_{H_n} f^{\tau_{in}}(\sigma\tau_{in}). \end{aligned}$$

where $f^{\tau_{in}}(\tau) = f(\tau\tau_{in})$ for all $\tau \in H_n$. Consequently, by induction, one has for all function $f : H_i \rightarrow \mathbb{R}$, for all $\lambda \geq 0$,

$$\begin{aligned} \int e^{\lambda Q_{H_i} f} d\mu_i &= \int e^{\lambda Q_{H_n} f^{\tau_{in}}(\sigma\tau_{in})} d\mu_i(\sigma) = \int e^{\lambda Q_{H_n} f^{\tau_{in}}} d\mu_n \\ &\leq \exp \left[\lambda \int f^{\tau_{in}} d\mu_n + (n-2) \frac{\lambda^2}{2} \right] = \exp \left[\lambda \int f d\mu_i + (n-2) \frac{\lambda^2}{2} \right]. \end{aligned}$$

Then the proof relies on the following Lemma.

Lemma 5.1. *For any function $\varphi : H_i \rightarrow \mathbb{R}$ and any $\sigma \in H_i$, one has*

$$Q\varphi(\sigma) \leq \inf_{\hat{p} \in \mathcal{P}(\{1, \dots, n\})} \left\{ \sum_{l=1}^n Q_{H_i} \varphi^{\tau_{il}} \hat{p}(l) + 2 \sum_{l=1}^n \mathbb{1}_{l \neq i} \hat{p}(l) \right\}.$$

The proof of this lemma is by decomposition of the probability measures p on the H_j 's, we get that if $\sigma \in H_i$ then

$$\begin{aligned} Q\varphi(\sigma) &= \inf_{\hat{p} \in \mathcal{P}(\{1, \dots, n\})} \inf_{p_1 \in \mathcal{P}(H_1), \dots, p_n \in \mathcal{P}(H_n)} \\ &\quad \left\{ \sum_{l=1}^n \left[\int \varphi dp_l + \sum_{k \notin \{l, i\}} \int \mathbb{1}_{\sigma(k) \neq \tau\tau_{il}(k)} dp_l(\tau) \right] \hat{p}(l) + 2 \sum_{l=1}^n \mathbb{1}_{l \neq i} \hat{p}(l) \right\}. \end{aligned}$$

The proof of (a) continues by applying consecutively this lemma, the Hölder inequality, and the induction hypotheses, this gives

$$\begin{aligned} \int e^{\lambda Q\varphi(\sigma)} d\mu_i(\sigma) &\leq \inf_{\hat{p} \in \mathcal{P}(\{1, \dots, n\})} \left\{ \prod_{l=1}^n \left(\int e^{\lambda Q_{H_i} \varphi^{\tau_{il}}} d\mu_i \right)^{\hat{p}(l)} e^{2\lambda \sum_{l=1}^n \mathbb{1}_{l \neq i} \hat{p}(l)} \right\} \\ &\leq \exp \left[\inf_{\hat{p} \in \mathcal{P}(\{1, \dots, n\})} \left\{ \lambda \sum_{l=1}^n \left(\int \varphi^{\tau_{il}} d\mu_i \right) \hat{p}(l) + (n-2) \frac{\lambda^2}{2} + 2\lambda \sum_{l=1}^n \mathbb{1}_{l \neq i} \hat{p}(l) \right\} \right] \\ &= \exp \left[\lambda \inf_{\hat{p} \in \mathcal{P}(\{1, \dots, n\})} \left\{ \sum_{l=1}^n \hat{\varphi}(l) \hat{p}(l) + 2 \sum_{l=1}^n \mathbb{1}_{l \neq i} \hat{p}(l) \right\} + (n-2) \frac{\lambda^2}{2} \right], \end{aligned}$$

where $\hat{\varphi}(l) = \int \varphi^{\tau_{il}} d\mu_i = \int \varphi d\mu_l$. Let us consider again the above infimum-convolution $R\hat{\varphi}$ defined on the space $\mathcal{X} = \{1, \dots, n\}$, one has

$$R\hat{\varphi}(i) = \inf_{\hat{p} \in \mathcal{P}(\{1, \dots, n\})} \left\{ \sum_{l=1}^n \hat{\varphi}(l) \hat{p}(l) + 2 \sum_{l=1}^n \mathbb{1}_{l \neq i} \hat{p}(l) \right\}.$$

As a consequence, by applying (24) with the uniform law ν on $\{1, \dots, n\}$, the previous inequality gives

$$\begin{aligned} \int e^{\lambda Q\varphi} d\mu &= \frac{1}{n} \sum_{i=1}^n \int e^{\lambda Q\varphi(\sigma)} d\mu_i(\sigma) \leq \left(\frac{1}{n} \sum_{i=1}^n e^{\lambda R\hat{\varphi}(i)} \right) e^{(n-2)\lambda^2/2} \\ &\leq \exp \left[\frac{\lambda}{n} \sum_{i=1}^n \hat{\varphi}(i) + \frac{\lambda^2}{2} + (n-2)\frac{\lambda^2}{2} \right] = \exp \left[\lambda \int \varphi d\mu + (n-1)\frac{\lambda^2}{2} \right]. \end{aligned}$$

□

REFERENCES

- [ABC⁺00] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux.
- [ACW14] R. Adamczak, D. Chafaï, and P. Wolff. Circular law for random matrices with exchangeable entries. *ArXiv e-prints*, February 2014.
- [Ada05] R. Adamczak. Logarithmic Sobolev inequalities and concentration of measure for convex functions and polynomial chaoses. *Bull. Pol. Acad. Sci. Math.*, 53(2):221–238, 2005.
- [Ada08] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13:no. 34, 1000–1034, 2008.
- [AGS14] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Math. J.*, 163(7):1405–1490, 2014.
- [AGS15] L. Ambrosio, N. Gigli, and G. Savaré. Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds. *Ann. Probab.*, 43(1):339–404, 2015.
- [AS15] R. Adamczak and M. Strzelecki. Modified log-Sobolev inequalities for convex functions on the real line. Sufficient conditions. *ArXiv e-prints*, May 2015.
- [BG99] S.G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999.
- [BGL01] S.G. Bobkov, I. Gentil, and M. Ledoux. Hypercontractivity of Hamilton-Jacobi equations. *J. Math. Pures Appl. (9)*, 80(7):669–696, 2001.
- [BHT06] S. G. Bobkov, C. Houdré, and P. Tetali. The subgaussian constant and concentration inequalities. *Israel J. Math.*, 156:255–283, 2006.
- [BK08] F. Barthe and A. V. Kolesnikov. Mass transport and variants of the logarithmic Sobolev inequality. *J. Geom. Anal.*, 18(4):921–979, 2008.
- [BL97] S. G. Bobkov and M. Ledoux. Poincaré’s inequalities and Talagrand’s concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields*, 107(3):383–400, 1997.
- [BL00] S. G. Bobkov and M. Ledoux. From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.*, 10(5):1028–1052, 2000.
- [BLM00] S. Boucheron, G. Lugosi, and P. Massart. A sharp concentration inequality with applications. *Random Structures Algorithms*, 16(3):277–292, 2000.
- [BLM03] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, 31(3):1583–1614, 2003.
- [BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.
- [Bob97] S. G. Bobkov. An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space. *Ann. Probab.*, 25(1):206–214, 1997.
- [Bou03] O. Bousquet. Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 213–247. Birkhäuser, Basel, 2003.

- [BR03] F. Barthe and C. Roberto. Sobolev inequalities for probability measures on the real line. *Studia Math.*, 159(3):481–497, 2003.
- [BR08] F. Barthe and C. Roberto. Modified logarithmic Sobolev inequalities on \mathbb{R} . *Potential Anal.*, 29(2):167–193, 2008.
- [CD10] S. Chatterjee and P.S. Dey. Applications of Stein’s method for concentration inequalities. *Ann. Probab.*, 38(6):2443–2485, 2010.
- [CG06] P. Cattiaux and A. Guillin. On quadratic transportation cost inequalities. *J. Math. Pures Appl. (9)*, 86(4):341–361, 2006.
- [Cha05] S. Chatterjee. Concentration inequalities with exchangeable pairs (Ph.D. thesis). *ArXiv Mathematics e-prints*, July 2005.
- [Cha07] S. Chatterjee. Stein’s method for concentration inequalities. *Probab. Theory Related Fields*, 138(1-2):305–321, 2007.
- [Csi67] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.
- [Dem97] A. Dembo. Information inequalities and concentration of measure. *Ann. Probab.*, 25(2):927–939, 1997.
- [DGW04] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.*, 32(3B):2702–2732, 2004.
- [Dou94] P. Doukhan. *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994. Properties and examples.
- [EM12] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.*, 206(3):997–1038, 2012.
- [EM14] M. Erbar and J. Maas. Gradient flow structures for discrete porous medium equations. *Discrete Contin. Dyn. Syst.*, 34(4):1355–1374, 2014.
- [EMT15] M. Erbar, J. Maas, and P. Tetali. Discrete Ricci curvature bounds for Bernoulli-Laplace and random transposition models. *Ann. Fac. Sci. Toulouse Math. (6)*, 24(4):781–800, 2015.
- [FHT03] A. A. Fedotov, P. Harremoës, and F. Topsøe. Refinements of Pinsker’s inequality. *IEEE Trans. Inform. Theory*, 49(6):1491–1498, 2003.
- [FMNW15] N. Feldheim, A. Marsiglietti, P. Nayar, and J. Wang. A note on the convex infimum convolution inequality. *ArXiv e-prints*, May 2015.
- [GGM05] I. Gentil, A. Guillin, and L. Miclo. Modified logarithmic Sobolev inequalities and transportation inequalities. *Probab. Theory Related Fields*, 133(3):409–436, 2005.
- [GGM07] I. Gentil, A. Guillin, and L. Miclo. Modified logarithmic Sobolev inequalities in null curvature. *Rev. Mat. Iberoam.*, 23(1):235–258, 2007.
- [Gil10] G. L. Gilardoni. On Pinsker’s and Vajda’s type inequalities for Csiszár’s f -divergences. *IEEE Trans. Inform. Theory*, 56(11):5377–5386, 2010.
- [GL10] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Process. Related Fields*, 16(4):635–736, 2010.
- [GM83] M. Gromov and V. D. Milman. A topological application of the isoperimetric inequality. *Amer. J. Math.*, 105(4):843–854, 1983.
- [Goz07] N. Gozlan. Characterization of Talagrand’s like transportation-cost inequalities on the real line. *J. Funct. Anal.*, 250(2):400–425, 2007.
- [Goz09] N. Gozlan. A characterization of dimension free concentration in terms of transportation inequalities. *Ann. Probab.*, 37(6):2480–2498, 2009.
- [Goz12] N. Gozlan. Transport-entropy inequalities on the line. *Electron. J. Probab.*, 17:no. 49, 18, 2012.
- [Goz15] Gozlan, N. Transport inequalities and concentration of measure*. *ESAIM: Proc.*, 51:1–23, 2015.
- [Gro75] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.
- [Gro99] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 of *Progress in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1999.
- [GRS11a] N. Gozlan, C. Roberto, and P.-M. Samson. From concentration to logarithmic Sobolev and Poincaré inequalities. *J. Funct. Anal.*, 260(5):1491–1522, 2011.
- [GRS11b] N. Gozlan, C. Roberto, and P.-M. Samson. A new characterization of Talagrand’s transport-entropy inequalities and applications. *Ann. Probab.*, 39(3):857–880, 2011.

- [GRS13] N. Gozlan, C. Roberto, and P.-M. Samson. Characterization of Talagrand’s transport-entropy inequalities in metric spaces. *Ann. Probab.*, 41(5):3112–3139, 2013.
- [GRS14] N. Gozlan, C. Roberto, and P.-M. Samson. Hamilton Jacobi equations on metric spaces and transport entropy inequalities. *Rev. Mat. Iberoam.*, 30(1):133–163, 2014.
- [GRS⁺15] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali. Characterization of a class of weak transport-entropy inequalities on the line. *ArXiv e-prints*, September 2015.
- [GRST14a] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Displacement convexity of entropy and related inequalities on graphs. *Probab. Theory Related Fields*, 160(1-2):47–94, 2014.
- [GRST14b] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *ArXiv e-prints*, December 2014.
- [Har66] L. H. Harper. Optimal numberings and isoperimetric problems on graphs. *J. Combinatorial Theory*, 1:385–393, 1966.
- [Hil14] E. Hillion. Contraction of measures on graphs. *Potential Anal.*, 41(3):679–698, 2014.
- [HS87] R. Holley and D. Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *J. Statist. Phys.*, 46(5-6):1159–1194, 1987.
- [Kom88] H. Komiya. Elementary proof for Sion’s minimax theorem. *Kodai Math. J.*, 11(1):5–7, 1988.
- [Kon12] A. Kontorovich. Obtaining measure concentration from Markov contraction. *Markov Process. Related Fields*, 18(4):613–638, 2012.
- [KR05] T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077, 2005.
- [Kul67] S. Kullback. Lower bound for discrimination information in terms of variation. *IEEE Trans. Information Theory*, 4:126–127, 1967.
- [Led93] M. Ledoux. Inégalités isopérimétriques en analyse et probabilités. *Astérisque*, (216):Exp. No. 773, 5, 343–375, 1993. Séminaire Bourbaki, Vol. 1992/93.
- [Led97] M. Ledoux. On Talagrand’s deviation inequalities for product measures. *ESAIM Probab. Statist.*, 1:63–87 (electronic), 1997.
- [Led01] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
- [Lév51] P. Lévy. *Problèmes concrets d’analyse fonctionnelle. Avec un complément sur les fonctionnelles analytiques par F. Pellegrino*. Gauthier-Villars, Paris, 1951. 2d ed.
- [LM03] M. J. Luczak and C. McDiarmid. Concentration for locally acting permutations. *Discrete Math.*, 265(1-3):159–171, 2003.
- [LT91] M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. Isoperimetry and processes.
- [Maa11] J. Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.
- [Mar96a] K. Marton. Bounding \bar{d} -distance by informational divergence: a method to prove measure concentration. *Ann. Probab.*, 24(2):857–866, 1996.
- [Mar96b] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6(3):556–571, 1996.
- [Mar97] K. Marton. Erratum to: “A measure concentration inequality for contracting Markov chains” [Geom. Funct. Anal. **6** (1996), no. 3, 556–571; MR1392329 (97g:60082)]. *Geom. Funct. Anal.*, 7(3):609–613, 1997.
- [Mar03] K. Marton. Measure concentration and strong mixing. *Studia Sci. Math. Hungar.*, 40(1-2):95–113, 2003.
- [Mar04] K. Marton. Measure concentration for Euclidean distance in the case of dependent random variables. *Ann. Probab.*, 32(3B):2526–2544, 2004.
- [Mar10] K. Marton. Correction: Measure concentration for Euclidean distance in the case of dependent random variables [mr2078549]. *Ann. Probab.*, 38(1):439–442, 2010.
- [Mas00a] P. Massart. About the constants in Talagrand’s concentration inequalities for empirical processes. *Ann. Probab.*, 28(2):863–884, 2000.
- [Mas00b] P. Massart. Some applications of concentration inequalities to statistics. *Ann. Fac. Sci. Toulouse Math. (6)*, 9(2):245–303, 2000. Probability theory.

- [Mas07] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.
- [Mau79] B. Maurey. Construction de suites symétriques. *C. R. Acad. Sci. Paris Sér. A-B*, 288(14):A679–A681, 1979.
- [Mau91] B. Maurey. Some deviation inequalities. *Geom. Funct. Anal.*, 1(2):188–197, 1991.
- [McD02] Colin McDiarmid. Concentration for independent permutations. *Combin. Probab. Comput.*, 11(2):163–178, 2002.
- [MS86] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.
- [MSWW11] Y. Ma, S. Shen, X. Wang, and L. Wu. Transportation inequalities: from Poisson to Gibbs measures. *Bernoulli*, 17(1):155–169, 2011.
- [Muc72] B. Muckenhoupt. Hardy’s inequality with weights. *Studia Math.*, 44:31–38, 1972. Collection of articles honoring the completion by Antoni Zygmund of 50 years of scientific activity, I.
- [Oll09] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.
- [OV00] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.
- [OV12] Y. Ollivier and C. Villani. A curved Brunn-Minkowski inequality on the discrete hypercube, or: what is the Ricci curvature of the discrete hypercube? *SIAM J. Discrete Math.*, 26(3):983–996, 2012.
- [Pan01] D. Panchenko. A note on Talagrand’s concentration inequality. *Electron. Comm. Probab.*, 6:55–65 (electronic), 2001.
- [Pan02] D. Panchenko. Some extensions of an inequality of Vapnik and Chervonenkis. *Electron. Comm. Probab.*, 7:55–65 (electronic), 2002.
- [Pan03] D. Panchenko. Symmetrization approach to concentration inequalities for empirical processes. *Ann. Probab.*, 31(4):2068–2081, 2003.
- [Pau12] D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *ArXiv e-prints*, December 2012.
- [Pau14] D. Paulin. The convex distance inequality for dependent random variables, with applications to the stochastic travelling salesman and other problems. *Electron. J. Probab.*, 19:no. 68, 34, 2014.
- [Pin64] M. S. Pinsker. *Information and information stability of random variables and processes*. Translated and edited by Amiel Feinstein. Holden-Day Inc., San Francisco, Calif., 1964.
- [Rio01] E. Rio. Inégalités de concentration pour les processus empiriques de classes de parties. *Probab. Theory Related Fields*, 119(2):163–175, 2001.
- [Rio02] E. Rio. Une inégalité de Bennett pour les maxima de processus empiriques. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(6):1053–1057, 2002. En l’honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.
- [Rio12] E. Rio. Sur la fonction de taux dans les inégalités de Talagrand pour les processus empiriques. *C. R. Math. Acad. Sci. Paris*, 350(5-6):303–305, 2012.
- [Rio13] E. Rio. On McDiarmid’s concentration inequality. *Electron. Commun. Probab.*, 18:no. 44, 11, 2013.
- [RR91] M. M. Rao and Z. D. Ren. *Theory of Orlicz spaces*. Marcel Dekker Inc., 1991.
- [Sam00] P.-M. Samson. Concentration of measure inequalities for Markov chains and Φ -mixing processes. *Ann. Probab.*, 28(1):416–461, 2000.
- [Sam03] P.-M. Samson. Concentration inequalities for convex functions on product spaces. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 33–52. Birkhäuser, Basel, 2003.
- [Sam07] P.-M. Samson. Infimum-convolution description of concentration properties of product probability measures, with applications. *Ann. Inst. H. Poincaré Probab. Statist.*, 43(3):321–338, 2007.
- [Sam15] P.-M. Samson. Transport inequalities on the symmetric group. Preprint, 2015.
- [Sio58] M. Sion. On general minimax theorems. *Pacific J. Math.*, 8:171–176, 1958.

- [Str65] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.
- [Tal91] M. Talagrand. A new isoperimetric inequality and the concentration of measure phenomenon. In *Geometric aspects of functional analysis (1989–90)*, volume 1469 of *Lecture Notes in Math.*, pages 94–124. Springer, Berlin, 1991.
- [Tal95] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, (81):73–205, 1995.
- [Tal96a] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.
- [Tal96b] M. Talagrand. A new look at independence. *Ann. Probab.*, 24(1):1–34, 1996.
- [Tal96c] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3):587–600, 1996.
- [Vil09] C. Villani. *Optimal transport: old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.
- [Wan14] N.-Y. Wang. Concentration inequalities for Gibbs sampling under d_{l_2} -metric. *Electron. Commun. Probab.*, 19:no. 63, 11, 2014.
- [Wu06] L. Wu. Poincaré and transportation inequalities for Gibbs measures under the Dobrushin uniqueness condition. *Ann. Probab.*, 34(5):1960–1989, 2006.
- [WW77] D. L. Wang and P. Wang. Extremal configurations on a discrete torus and a generalization of the generalized Macaulay theorem. *SIAM. J. Appl. Math.*, 33(1):55–59, 1977.
- [WW14] N.-Y. Wang and L. Wu. Convergence rate and concentration inequalities for Gibbs sampling in high dimension. *Bernoulli*, 20(4):1698–1716, 2014.

SAMSON P.-M., UNIVERSITÉ PARIS-EST, LABORATOIRE D'ANALYSE ET DE MATHÉMATIQUES APPLIQUÉES (UMR 8050), UPEM, UPEC, CNRS, F-77454, MARNE-LA-VALLÉE, FRANCE

E-mail address: paul-marie.samson@u-pem.fr