Canad. J. Math. Vol. 69 (3), 2017 pp. 481–501
http://dx.doi.org/10.4153/CJM-2016-046-3

Transport Inequalities for Log-concave Measures, Quantitative Forms, and Applications

Dario Cordero-Erausquin

Abstract. We review some simple techniques based on monotone mass transport that allow us to obtain transport-type inequalities for any log-concave probability measures, and for more general measures as well. We discuss quantitative forms of these inequalities, with application to the Brascamp–Lieb variance inequality.

1 Introduction

Throughout the paper we work, when needed, with some fixed scalar product $\cdot$ and Euclidean norm $|\cdot|$ on $\mathbb{R}^n$. Although our main motivation is to analyse log-concave densities, meaning densities of the form e^{-V} with V convex, our results apply to more general situations, regardless of the convexity of the potential V. We can often work with a locally Lipschitz function $V:\mathbb{R}^n \to \mathbb{R}$ with the mild assumption that

$$\int (1+|x|^2+|\nabla V(x)|^2)e^{-V(x)}\,dx < +\infty. \tag{1.1}$$

Actually, when V is convex, we don't need these assumptions, but not much is lost by imposing it. Given such V, we introduce the probability measure μ_V defined by

$$d\mu_V(x) := \frac{e^{-V(x)}}{\int e^{-V}}\,dx.$$

Note that the density is by assumption everywhere strictly positive.

Following Kantorovich's idea, given a function $\mathbf{c}:\mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}$ (one interprets $\mathbf{c}(x,y)$ as the cost of moving a unit mass from x to y or of bringing back a unit mass from y to x), we can define a transportation cost $\mathcal{W}_{\mathbf{c}}$ between two Borel probability measures μ and ν on $\mathbb{R}^n$ by

$$\mathcal{W}_{\mathbf{c}}(\mu,\nu) := \mathcal{W}_{\mathbf{c}(x,y)}(\mu,\nu) := \inf_{\pi} \iint_{\mathbb{R}^n\times\mathbb{R}^n} \mathbf{c}(x,y)\,d\pi(x,y)$$

where the infimum is taken over all probability measures π on $\mathbb{R}^n\times\mathbb{R}^n$ projecting on μ and ν, respectively. From the definition of $\mathcal{W}_{\mathbf{c}}(\mu,\nu)$ and Fubini's theorem, we see that it suffices to show that the cost is well defined on $(\mathbb{R}^n\smallsetminus X)\times\mathbb{R}^n$ only, where $\mu(X)=0$. Under very mild hypotheses on $\mathbf{c}$, one can prove that there exists a coupling π that is

Received by the editors May 27, 2015; revised July 15, 2016.
Published electronically March 27, 2017.
AMS subject classification: **52A40**, 60E15, 49Q20.
Keywords: log-concave measures, transport inequality, Brascamp–Lieb inequality, quantitative inequalities.

optimal, that is, which achieves the infimum above (see [32, Chapters 4 and 5]). The cost $\mathbf{c}(x,y) = |y-x|^p$, $p \in [1,+\infty)$, is used for the definition of the L^p-Kantorovich–Rubinstein (or Wassertein) distance

$$W_p(\mu,\nu) := \big(\mathcal{W}_{|x-y|^p}(\mu,\nu)\big)^{1/p}.$$

Recall that given two probability measures μ and ν on $\mathbb{R}^n$, the relative entropy of ν with respect to μ is defined by

$$H(\nu\|\mu) := \begin{cases} \int f\log(f)\,d\mu & \text{if } d\nu(x) = f(x)d\mu(x) \text{ with } f\log_+(f) \in L^1(\mu), \\ +\infty & \text{otherwise.} \end{cases}$$

Accordingly, we should only consider probability measures that have a density, in short "absolutely continuous" probability measures. Recall also that the variance of a function $g \in L^2(\mu)$ is defined by

$$\mathrm{Var}_\mu(g) := \int \Big(g - \int g\,d\mu\Big)^2\,d\mu.$$

The inequality in the next proposition appeared in [6] where it was derived in the dual form (1.5) as a consequence of the Prékopa–Leindler inequality. By now it is folklore in optimal mass transportation theory and known to most specialists. The investigation of equality cases seems to be new.

Proposition 1.1 *Let $V : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function satisfying* (1.1). *Define for every y and almost every x in $\mathbb{R}^n$ the (asymmetric) cost*

$$\mathbf{c}_V(x,y) := V(y) - V(x) - \nabla V(x)\cdot(y-x). \tag{1.2}$$

Then for every (absolutely continuous) probability measure ν on $\mathbb{R}^n$, we have

$$\mathcal{W}_{\mathbf{c}_V}(\mu_V,\nu) \le H(\nu\|\mu_V). \tag{1.3}$$

Moreover, when V is convex, equality holds if and only if ν is a translate of μ_V.

Remark 1.2 Regarding the treatment of the equality case, we can prove a sharper result. Namely, we will establish the following statement. Let $V : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz function satisfying (1.1), and assume that μ_V has a positive Cheeger constant $h(\mu_V) > 0$ (see definition below; this is the case when V is convex). Then there is equality in the transport inequality (1.3) of Proposition 1.1 if and only if V is convex and ν is a translate of μ_V.

The fact that the convexity of V is necessary for equality cases is reminiscent of the equality cases in the Brunn–Minkowski inequality.

When V is convex, it is possible to define the cost for every x (and not just for almost every x) by using the subgradient $\partial V(x)$ at x of the convex function V (see [29] for background on subgradients):

$$\partial V(x) := \{w \in \mathbb{R}^n;\ V(x+h) \ge V(x) + w\cdot h,\ \forall h \in \mathbb{R}^n\}.$$

Indeed, the Proposition can then be stated with the following cost $\mathbf{c}_V$ in place of (1.2):

$$\forall (x,y) \in \mathbb{R}^n\times\mathbb{R}^n, \qquad \mathbf{c}_V(x,y) := \sup_{w\in\partial V(x)} \Big\{V(y) - V(x) - w\cdot(y-x)\Big\}. \tag{1.4}$$

Recall that V is locally Lipschitz and so differentiable μ-almost-everywhere.

Note that when V is convex, we have $\mathbf{c}_V(x,y) \geq 0$ with $\mathbf{c}_V(x,x) = 0$ in (1.2) and (1.4); when V is strictly convex, $\mathbf{c}_V(x,y) > 0$ if $x \neq y$.

Let us mention that by a simple and standard dualization procedure for transportation inequalities (see [21]), the statement of Proposition 1.1 is equivalent to the following infimal convolution inequality: for every (bounded) function $g: \mathbb{R}^n \to \mathbb{R}$,

$$\int e^{Q_{\mathbf{c}_V}(g)}\, d\mu_V \leq e^{\int g\, d\mu_V}, \tag{1.5}$$

where

$$Q_{\mathbf{c}_V}(g)(y) := \inf_x \left\{ g(x) + \mathbf{c}_V(x,y) \right\}. \tag{1.6}$$

We should also mention that transportation cost inequalities of the form stated above imply concentration of measure inequalities (for $\mathbf{c}_V$-neighborhoods); we refer the reader to [21] for details.

The interest of the statement in Proposition 1.1 resides in the fact that no uniform convexity of V is needed. This is reminiscent of the Brascamp–Lieb variance inequality [8] (anticipated in different context by Hörmander), which states that for a C^2 smooth convex function $V: \mathbb{R}^n \to \mathbb{R}$ with $\int e^{-V} < +\infty$, we have, for every locally Lipschitz function $g \in L^2(\mu_V)$,

$$\operatorname{Var}_{\mu_V}(g) \leq \int (D^2 V(x))^{-1} \nabla g(x) \cdot \nabla g(x)\, d\mu_V(x). \tag{1.7}$$

Since the cost $\mathbf{c}_V(x,y)$ in Proposition 1.1 behaves, when x and y are close to each other, like $\frac{1}{2} D^2 V(x)(y-x)\cdot(y-x)$, it follows by a standard linearization argument that Proposition 1.1 implies the Brascamp–Lieb inequality (1.7). We will recall the argument later.

Another interesting feature of Proposition 1.1 is that it is an *affinely invariant* statement, in the sense that it does not depend on the Euclidean structure we put on $\mathbb{R}^n$. More precisely, we do not need a scalar product in the statement: the gradient $w = \nabla f(s)$ (or a subgradient) comes from a linear form $\ell = df(x) \in (\mathbb{R}^n)^*$, and we can use $\ell(y-x)$ in place of $w \cdot (y-x)$. This also reflects in the fact that the Brascamp–Lieb inequality (1.7) shares the same affine invariance: if $\varphi: \mathbb{R}^n \to \mathbb{R}^n$ is an (invertible) affine map, then the functions $V_\varphi = V \circ \varphi^{-1}$ and $g_\varphi = g \circ \varphi^{-1}$ satisfy $\operatorname{Var}_{\mu_{V_\varphi}}(g_\varphi) = \operatorname{Var}_{\mu_V}(g)$ and

$$\int (D^2 V_\varphi(x))^{-1} \nabla g_\varphi(x) \cdot \nabla g_\varphi(x)\, d\mu_{V_\varphi}(x) = \int (D^2 V(x))^{-1} \nabla g(x) \cdot \nabla g(x)\, d\mu_V(x).$$

Other consequences of Proposition 1.1 are Talagrand's transportation inequalities for Gaussian-like measures. Observe that for the standard Gaussian measure γ, when $V(x) = |x|^2/2$, we have

$$\mathbf{c}_V(x,y) = |y-x|^2/2$$

and the inequality becomes exactly Talagrand's inequality [30]: for every probability density ν on $\mathbb{R}^n$,

$$\frac{1}{2} W_2^2(\gamma, \nu) \leq H(\nu \| \gamma),$$

with equality if and only if ν is a translate of γ. More generally, if V is C^2 with $D^2V \geq \lambda\,\mathrm{Id}$ on $\mathbb{R}^n$ for some $\lambda > 0$, then by second-order Taylor expansion we see that the cost satisfies

$$\mathbf{c}_V(x,y) \geq \lambda|y-x|^2/2,$$

and therefore we deduce that in this case, for every probability measure ν on $\mathbb{R}^n$, we have

$$\frac{\lambda}{2}\,W_2^2(\mu_V,\nu) \leq H(\nu\|\mu_V). \tag{1.8}$$

This inequality appeared in [3, 6, 28]. We refer the reader to [18, 21] for background and references on transportation inequalities.

The proof of Proposition 1.1 is very short; it is a minor adaptation of the transportation proof of Talagrand's inequality (1.8) given in [10]. With a little more effort one can actually prove a quantitative form of the inequality involving a remainder term. To state the result, we need some notation. Given a probability measure μ on $\mathbb{R}^n$, we denote by $h(\mu)$ the best (*i.e.*, largest) nonnegative constant for which the inequality

$$h(\mu)\int\left|g(x)-\int g\,d\mu\right|d\mu(x) \leq \int|\nabla g|\,d\mu \tag{1.9}$$

holds for every sufficiently smooth $g \in L^1(\mu)$. This constant, up to a factor 2, is also known as the Cheeger isoperimetric constant.

When μ is log-concave, then it is known that $h(\mu) > 0$, and $h(\mu)^2$ is actually equivalent, up to an universal constant, to the spectral gap of the Laplacian associated with μ (or equivalently the inverse of the Poincaré constant). More explicitly, if we denote by $\lambda(\mu)$ the best nonnegative constant for which the inequality

$$\lambda(\mu)\int\left|g(x)-\int g\,d\mu\right|^2 d\mu(x) \leq \int|\nabla g|^2\,d\mu$$

holds for every smooth enough $g \in L^2(\mu)$, then when μ is a log-concave measure on $\mathbb{R}^n$, we have

$$c\,h(\mu)^2 \leq \lambda(\mu) \leq C\,h(\mu)^2 \tag{1.10}$$

for some universal (numerical) constants $c, C > 0$, independent of μ and n; see [22, 25].

In the rest of the paper, we will adopt the lazy but convenient tradition from asymptotic functional analysis to call "a numerical constant c" any positive constant larger than 2 or smaller than 1/2 (c may even vary from line to line). So a numerical constant refers to a universal constant (in particular, it does not depend on n, V, μ, ν, etc.) whose exact value is irrelevant but who could *a priori* be computed explicitly.

There is another natural cost function associated with any measure having a positive Cheeger constant, namely the cost $\min(h(\mu)^2|y-x|^2, h(\mu)|y-x|) = \mathcal{N}(h(\mu)|y-x|)$, where

$$\forall t \geq 0, \qquad \mathcal{N}(t) := \min(t^2,\ t).$$

This cost (in this form or in some equivalent form) has been studied by several authors (see [18, 21] for details).

Since equality holds in Proposition 1.1 when ν is a translate of μ, it is natural, if we want a remainder term, to minimize over translations, or equivalently, to impose some centering.

The main result of this note is the following theorem.

Theorem 1.3 (General transport inequality with a remainder term) *Let $V: \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function satisfying (1.1) and let $\mathbf{c}_V$ be the cost defined by (1.2). Introduce the cost*

$$\widetilde{\mathbf{c}}_V(x,y) = \mathbf{c}_V(x,y) + c\,\mathcal{N}\big(h(\mu_V)|y-x|\big),$$

where $c > 0$ is an appropriate numerical constant.

Then for every probability measure ν on $\mathbb{R}^n$ such that $\int x\,d\nu = \int x\,d\mu_V$ we have

$$\mathcal{W}_{\widetilde{\mathbf{c}}_V}(\mu_V,\nu) \le H(\nu\|\mu_V). \tag{1.11}$$

As a consequence, we have the quantitative version of Proposition 1.1 when $\int x\,d\nu = \int x\,d\mu_V$:

$$H(\nu\|\mu_V) \ge \mathcal{W}_{\mathbf{c}_V}(\mu_V,\nu) + c\,\mathcal{W}_{\mathcal{N}(h(\mu_V)|y-x|)}(\mu_V,\nu), \tag{1.12}$$

and, in particular, we also have

$$H(\nu\|\mu_V) \ge \mathcal{W}_{\mathbf{c}_V}(\mu_V,\nu) + c\min\Big\{h(\mu_V)^2 W_1^2(\mu_V,\nu), h(\mu_V)W_1(\mu_V,\nu)\Big\}. \tag{1.13}$$

Note that unlike the quantities H and $\mathcal{W}_{\mathbf{c}_V}$, the cost $\mathcal{N}_{\mathcal{N}(h(\mu_V)|y-x|)}$ is very much dependent on the scalar product, which should therefore be chosen with care.

Let us explain how the consequences of (1.11) stated in the theorem are obtained. The first one (1.12) follows from a general and straightforward principle: given two costs c_1, c_2, we always have $\mathcal{W}_{c_1+c_2}(\cdot,\cdot) \ge \mathcal{W}_{c_1}(\cdot,\cdot) + \mathcal{W}_{c_2}(\cdot,\cdot)$. The "in particular", may seem more dubious. The reason that (1.13) follows indeed from (1.12) is that, up to numerical constants (see below) we can replace the function $\mathcal{N}(s) = \min(s^2, s)$ by a convex increasing function $\mathcal{F}(s)$, and then we can invoque Jensen's inequality to ensure that $\mathcal{W}_{\mathcal{F}(|y-x|)}(\nu,\mu) \ge \mathcal{F}(\mathcal{W}_{|y-x|}(\mu,\nu))$.

Note, however, that the form (1.13) is strictly weaker than the forms (1.11) and (1.12). In particular, we should note that the cost $\mathcal{N}(h(\mu_V)|y-x|)$ behaves like $h(\mu_V)^2|y-x|^2$ when x and y are close to each other, and this behavior is well adapted to linearization procedures.

Let us describe some consequences of Theorem 1.3 in the case where V is convex. Applied to Gaussian type measures, when $\mathbf{c}_V(x,y) \ge \lambda|x-y|^2/2$, it amounts to a quantitative version of the transport inequality (1.8).

Proposition 1.4 (Gaussian type transport with a remainder) *Let $V: \mathbb{R}^n \to \mathbb{R}$ be a C^2 convex function with $D^2V \ge \lambda\,\mathrm{Id}$ on $\mathbb{R}^n$ for some $\lambda > 0$ (we have mainly in mind the Gaussian measure, for which $\lambda = 1$). Then for every probability measure ν on $\mathbb{R}^n$ such that $\int x\,d\nu = \int x\,d\mu_V$, we have*

$$\begin{aligned} H(\nu\|\mu_V) - \frac{\lambda}{2}W_2^2(\mu_V,\nu) &\ge c\,\mathcal{W}_{\mathcal{N}(h(\mu_V)|y-x|)}(\mu_V,\nu) \\ &\ge \widetilde{c}\min\Big\{h(\mu_V)^2 W_1^2(\mu_V,\nu), h(\mu_V)W_1(\mu_V,\nu)\Big\} \end{aligned} \tag{1.14}$$

https://doi.org/10.4153/CJM-2016-046-3 Published online by Cambridge University Press

for some numerical constants $c, \widetilde{c} > 0$*. One can also replace* $h(\mu_V)^2$ *by* λ*, since* $h(\mu_V)^2 \geq c' \lambda$ *for some numerical constant* $c' > 0$*.*

Next, linearization of the inequality (1.11) in Theorem 1.3 leads to a reinforced Brascamp–Lieb inequality in the case of centered functions.

Proposition 1.5 *Let* $V: \mathbb{R}^n \to \mathbb{R}$ *be a* C^2 *convex function with* $\int e^{-V} < +\infty$*. For every locally Lipschitz function* $g \in L^2(\mu_V)$ *with*

$$\int x\Big(g(x) - \int g\, d\mu_V\Big)\, d\mu_V(x) = 0,$$

we have

$$\mathrm{Var}_{\mu_V}(g) \leq \int \big[D^2 V + c\, h(\mu_V)^2\, \mathrm{Id}\big]^{-1} \nabla g \cdot \nabla g\, d\mu_V,$$

for some numerical constant $c > 0$*. We can replace* $h(\mu_V)^2$ *by* $\lambda(\mu_V)$*, in view of* (1.10).

One can derive a similar result using Hörmander's L^2-method (see *e.g.*, [1]).

We should add, as is apparent from the proof, that the convexity of V is not really needed in Propostion 1.5. The correct assumption is that $D^2 V + c\, h(\mu_V)^2\, \mathrm{Id}$ is non-negative. In particular, the result applies to perturbed log-concave measures, provided $h(\mu_V) > 0$.

Equality cases in the Brascamp–Lieb inequality (1.7) are given, exactly, by the functions g of the form

$$g(x) = \nabla V(x) \cdot u_0 + c_0 \tag{1.15}$$

with $u_0 \in \mathbb{R}^n$ and $c_0 \in \mathbb{R}$. In order to have a nice quantitative version, one would like to get rid of the centering assumption and to measure, in some form, a "distance"

$$\inf_{u_0, c_0} d(g, \nabla V \cdot u_0 + c_0)$$

to the set of extremizers (1.15). Here is an attempt.

Proposition 1.6 (Brascamp–Lieb inequality with a remainder term) *Let* $V: \mathbb{R}^n \to \mathbb{R}$ *be a* C^2 *convex function with* $\int e^{-V} < +\infty$*. Then for every locally Lipschitz function* $g \in L^2(\mu_V)$*, if we denote*

$$g_0(x) := g(x) - \nabla V(x) \cdot u_0 - c_0, \quad c_0 := \int g\, d\mu_V \text{ and } u_0 := \int y\,(g(y) - c_0)\, d\mu_V(y),$$

we have

$$\int (D^2 V(x))^{-1} \nabla g(x) \cdot \nabla g(x)\, d\mu_V(x) - \mathrm{Var}_{\mu_V}(g) \geq \\ c\lambda(\mu_V) \int (D^2 V)^{-1} \big(D^2 V + c\lambda(\mu_V)\, \mathrm{Id}\big)^{-1} \nabla g_0 \cdot \nabla g_0\, d\mu_V,$$

https://doi.org/10.4153/CJM-2016-046-3 Published online by Cambridge University Press

where $c > 0$ is a numerical constant. As a consequence, if we denote by $\lambda_{\max}(x)$ the largest eigenvalue of the nonnegative operator $D^2V(x)$, we have

$$\int (D^2V(x))^{-1}\nabla g(x)\cdot\nabla g(x)\,d\mu_V(x) - \mathrm{Var}_{\mu_V}(g) \geq \frac{c\lambda(\mu_V)}{\sup_x \lambda_{\max}(x) + c\lambda(\mu_V)}\int |g_0|^2\,d\mu_V$$

and

$$\int (D^2V(x))^{-1}\nabla g(x)\cdot\nabla g(x)\,d\mu_V(x) - \mathrm{Var}_{\mu_V}(g) \geq \frac{\widetilde{c}\lambda(\mu_V)^2}{\int \lambda_{\max}(\lambda_{\max} + c\lambda(\mu_V))\,d\mu_V}\Big(\int |g_0|\,d\mu_V\Big)^2,$$

where $\widetilde{c} > 0$ is a numerical constant.

Let us recall that $\lambda(\mu_V)$ can be estimated by

$$\lambda(\mu_V) \geq \frac{c}{\int \lambda_{\min}^{-1}\,d\mu_V},$$

where $\lambda_{\min}(x)$ denotes the lowest eigenvalue of the nonnegative operator $D^2V(x)$ and $c > 0$ is a numerical constant (see [26,31]). Therefore, the constant in the previous proposition (which is an increasing function of $\lambda(\mu_V)$) can be lower bounded by some integrals of $\lambda_{\min}$ and $\lambda_{\max}$ with respect to μ_V. For instance, using the previous bound and the fact that $(\int \lambda_{\min}^{-1}\,d\mu_V)^{-1} \leq \int \lambda_{\max}\,d\mu_V$, we find

$$\frac{\widetilde{c}\lambda(\mu_V)^2}{\int \lambda_{\max}(\lambda_{\max} + c\lambda(\mu_V))\,d\mu_V} \geq \frac{c}{\big(\int \lambda_{\min}^{-1}\,d\mu_V\big)^2\int \lambda_{\max}^2\,d\mu_V}$$

for some numerical constant $c > 0$. This might provide a computable constant beyond the easy case where $\lambda \leq D^2V \leq R$ on $\mathbb{R}^n$.

We conclude this introduction with some bibliographical comments. Part of this note is rather elementary, and many arguments are known to specialists in mass transport, some having appeared implicitly or explicitly in recent or older works. For instance, we already said that Proposition 1.1 was folklore in the theory, and while writing these notes we heard about the work of Bolley, Gentil, and Guillin [7], which contains an analogue, in a less straightforward form, of the inequality of Proposition 1.1 together with its connection to the Brascamp–Lieb inequality. If we go back in time, the idea of using the remainder term in the transportation proof of [10] appears, in the case of dimension one, in the paper by Barthe and Kolesnikov [2]. Similar arguments in higher dimensions for unconditional measures were recently used in [20] and in a form very close to the one used here in [11]. Mass transport arguments combined with Poincaré inequalities (of a different nature than the one we use) were put forward to exhibit remainder terms in isoperimetric type inequalities in the far-reaching work of Figalli, Maggi, and Pratelli, in particular in [15] for the case of log-concave measures (or rather convex sets). Our treatment is in part very close to the recent work of Fathi, Indrei, and Ledoux [14] where the mass transport remainder term is combined with a Poincaré inequality in order to get a bound on the deficit for Talagrand's inequality (1.8) in the case of the Gaussian measure (they have also similar, but deeper, arguments for the log-Sobolev inequality, a case that was also considered in [4]).

The quantitative transport inequality obtained by Fathi, Indrei, and Ledoux [14] for the standard Gaussian measure $\mu = \gamma$ on $\mathbb{R}^n$ (a case where $\lambda(\mu) = 1$) is as follows: for any probability measure ν with $\int x\, d\nu(x) = \int x\, d\gamma(x) = 0$,

$$H(\nu\|\gamma) - \frac{1}{2}W_2^2(\gamma,\nu) \geq c\min\Big(\frac{W_{1,1}^2(\gamma,\nu)}{n}, \frac{W_{1,1}(\gamma,\nu)}{\sqrt{n}}\Big),$$

where $W_{1,1} := \mathcal{W}_{\|x-y\|_1}$ with $\|x-y\|_1 := \sum_{i=1}^n |x_i - y_i|$. If we compare with inequality (1.14) in Proposition 1.4 above applied to the Gaussian measure $\mu = \gamma$, we see that our result is formally stronger, since

$$W_1(\mu,\nu) \geq \frac{W_{1,1}(\mu,\nu)}{\sqrt{n}}.$$

Actually, our bound is significantly better in many cases, but both bounds are "equally bad" when ν is a product of centered measures being at a "large" distance from the one-dimensional Gaussian, since in this case one expects a remainder of order n and both results give something of order $\sqrt{n}$ (on the other hand, it is not clear to us that this situation is the most relevant one).

As to quantative versions of the variance Brascamp–Lieb inequality, Hargé [19] (by an L^2-method) and recently Bolley, Gentil, and Guillin [7] (by linearization of a transport inequality) obtained a remainder term which is, up to constants depending on μ_V, of the form

$$\Big(\int g\, V\, d\mu_V - \int g\, d\mu_V \int V\, d\mu_V\Big)^2 =: R_V(g).$$

Note that, unlike the remainder term in Proposition 1.6, this term $R_V(g)$ does not vanish only for extremizers. For instance, $R_V(g)$ is zero if V is even and g odd. Actually, the space where R_V vanishes is of co-dimension one in $L^2(\mu_V)$, whereas extremizers (1.15) form a $(n+1)$ dimensional subspace. Of course, it could be that such type of remainder is nonetheless sometimes better and more useful than the one we obtained. Bolley, Gentil, and Guillin also derive, in the same work [7] but by a different method (namely by linearization of a functional Brunn-Minkowski inequality), a second quantative form of the variance Brascamp–Lieb inequality with a remainder term that vanishes exactly for the extremizers (1.15), as expected. This remainder, however, is not an L^1 or L^2 distance to the space of extremizers, and so the comparison with the result of our Proposition 1.6 is not clear to us.

The plan of the paper is as follows. In the next section we prove Proposition 1.1 and Theorem 1.3. For this we recall some tools from the Brenier-McCann monotone mass transport theory, and prove a general lower bound for the remainder term (Lemma 2.2) that might be of independent interest. Then we prove Propositions 1.5 and 1.6.

2 Mass Transport, Minoration of the Remainder and Proofs of Proposition 1.1 and Theorem 1.3

The proofs of Theorems 1.1 and 1.3 use monotone transportation of measure in the spirit of [10].

Given two probability measures μ and ν on $\mathbb{R}^n$ with densities F and G, respectively, we know from Brenier [9] and McCann [23] that there exists a convex function ψ such that the map $\nabla\psi$ pushes forward μ onto ν. By the simple but useful weak-regularity theory of McCann [24] we have, for μ-almost any x,

$$F(x) = G(\nabla(\psi(x))\det D^2\psi(x). \tag{2.1}$$

Here, $D^2\psi(x)$ stands for the Hessian of the convex function ψ in the sense of Aleksandrov, that exists almost everywhere. There are several ways to use this equation to prove our inequalities. One can use the McCann weak theory of change of variables [24], as in [10]. The advantage is that it relies on simple arguments in convexity and Lebesgue measure theory. Alternatively, one can use results on the regularity of Monge-Ampère equation, in the spirit of those obtained by Caffarelli. This relies on more difficult and deeper arguments. However, partial regularity results for solutions of Monge-Ampère have been simplified and extended recently, and we shall favor this point of view.

Let us assume that μ and ν are supported on the whole $\mathbb{R}^n$, and that the densities are continuous and strictly positive (so locally bounded above and away from zero). Since the support of the target measure is convex (here $\mathbb{R}^n$), one can prove that the convex function ψ solves the Monge-Ampère equation (2.1) also in the sense of Aleksandrov (see the argument given in the proof of [16, Theorem 3.3], and by the assumption above on the densities, the local regularity of [27], say, applies). In particular, ψ is $W^{2,1}_{\mathrm{loc}}(\mathbb{R}^n)$. To prove the transport inequality of Proposition 1.1 for $d\mu_V = e^{-V(x)}/\int e^{-V}\,dx$, we assume that $d\nu = f(x)\,d\mu_V(x)$. It is sufficient to prove the inequalities in Proposition 1.1 and Theorem 1.3 in the case where f is continuous and strictly positive on $\mathbb{R}^n$, so that the previous assumptions are satisfied. We can also assume that ν has second moment.

We introduce the Brenier map $T = \nabla\psi$ between μ_V and ν. We have that $\psi \in W^{2,1}_{\mathrm{loc}}(\mathbb{R}^n)$ and that almost everywhere

$$e^{-V(x)} = f(T(x))e^{-V(T(x))}\det D^2\psi(x).$$

It is convenient to introduce the displacement $\nabla\theta(x) = T(x) - x = \nabla\psi(x) - x$ (*i.e.*, $\theta(x) := \psi(x) - |x|^2/2$). If we take the log in the previous equation and introduce $\mathbf{c}_V(x, T(x)) = V(T(x)) - V(x) - \nabla V(x)\cdot\nabla\theta(x)$, we find

$$\begin{aligned}\log(f(T(x))) - \mathbf{c}_V(x, T(x)) &= \nabla V(x)\cdot\nabla\theta(x) - \log\det D^2\psi(x)\\ &= \nabla V\cdot\nabla\theta - \Delta\theta + \Delta\theta - \log\det(\mathrm{Id} + D^2\theta).\end{aligned}$$

We integrate with respect to μ_V. Noticing that $\int \log(f \circ T) d\mu_V = \int f \log(f) d\mu_V$, we have

$$H(\nu\|\mu_V) - \int \mathbf{c}_V(x, T(x))\, d\mu_V = \int \big[\nabla V \cdot \nabla\theta - \Delta\theta\big]\, d\mu_V + \int \big[\Delta\theta - \log\det(\mathrm{Id} + D^2\theta)\big]\, d\mu_V.$$

The first term in the right-hand side vanishes after integration by parts, thanks to the integrability assumptions we have made. Let us justify this.

Fact 2.1 $\int \Delta\theta\, e^{-V} = \int \nabla\theta \cdot \nabla V\, e^{-V}$.

Proof Since $\int |\nabla\psi|^\alpha e^{-V} = \int |y|^\alpha\, d\nu(y)$ and ν as second moment, we have in view of our assumptions that $\int |\nabla\theta|^2 e^{-V} < +\infty$ and $\int |\nabla\theta| e^{-V} < +\infty$. Let h be a C^1 function on $\mathbb{R}^n$, with values on $[0,1]$, that is compactly supported and is identically one in a neighborhood of $0 \in \mathbb{R}^n$. Introduce the sequence $h_k(x) = h(x/k)$. We have $0 \le h_k \le 1$, $h_k(x) \uparrow 1$ for every $x \in \mathbb{R}^n$ and $\|\nabla h_k\|_\infty \to 0$ as $k \to +\infty$. We have

$$\int h_k \Delta\theta\, e^{-V} = -\int \nabla h_k \cdot \nabla\theta\, e^{-V} + \int h_k \nabla\theta \cdot \nabla V\, e^{-V}.$$

For the left-hand side, we wan write

$$\int h_k \Delta\theta\, e^{-V} = \int h_k \Delta\psi\, e^{-V} - n \int h_k e^{-V}$$

and each term converges using the monotone convergence theorem (since $\Delta\psi \ge 0$), giving $\int \Delta\psi\, e^{-V} - n \int e^{-V} = \int (\Delta\psi - n)\, e^{-V} = \int \Delta\theta\, e^{-V}$. The first term in the right-hand side tends to zero, since it is bounded by $\|\nabla h_k\|_\infty \int |\nabla\theta|\, e^{-V}$. For the last term, we conclude by using the dominated convergence theorem, since

$$2 \int |\nabla\theta \cdot \nabla V|\, e^{-V} \le \int |\nabla\theta|^2 e^{-V} + \int |\nabla V|^2 e^{-V} < +\infty. \qquad \blacksquare$$

So we have arrived at the elementary formula

$$\begin{aligned} H(\nu\|\mu_V) &= \int \mathbf{c}_V(x, T(x))\, d\mu_V(x) + \int \big[\Delta\theta - \log\det(\mathrm{Id} + D^2\theta)\big]\, d\mu_V \\ &= \int \mathbf{c}_V(x, T(x))\, d\mu_V(x) + \int \big[\operatorname{tr} D^2\theta - \operatorname{tr}(\log(\mathrm{Id} + D^2\theta))\big]\, d\mu_V \\ &= \int \mathbf{c}_V(x, T(x))\, d\mu_V(x) + \int \operatorname{tr}(\mathcal{F}(D^2\theta))\, d\mu_V, \end{aligned} \tag{2.2}$$

where $\mathcal{F}\colon [-1, +\infty[\to [0, +\infty[$ stands for the convex (increasing on $\mathbb{R}^+$) function defined by

$$\mathcal{F}(t) := t - \log(1+t), \qquad t \in \mathbb{R}^+.$$

Since by definition $\int \mathbf{c}_V(x, T(x))\, d\mu_V(x) \ge \mathcal{W}_{\mathbf{c}_V}(\mu_V, \nu)$ and $\mathcal{F} \ge 0$, we have proved, in particular, the inequality in Proposition 1.1.

The treatment of the cases of equality in Proposition 1.1 requires a bit of extra work (in particular since T was not *a priori* the $\mathbf{c}_V$-optimal map); we postpone it to the end of this section and go on with the proof of Theorem 1.3.

In order to prove Theorem 1.3, we have to play a bit with the second term in the right-hand side of (2.2), as was done in the works we mentioned in the introduction.

Indeed, a "remainder" term of this form appears in several mass transport proofs (for instance [2, 11, 14, 15, 20]), sometimes in equivalent forms such as

$$\sum\Big(|s_i| + \frac{1}{1+|s_i|} - 1\Big) \quad \text{or} \quad \sum \frac{s_i^2}{1+|s_i|}$$

(here s_i refer to the eigenvalues of $D^2\theta$). Anyway, the crucial property of the these functions and of the convex function $t - \log(1+t)$ is that it behaves like t^2 for t close to zero, and like t for t large. More precisely, we have, for every $t \in]-1, +\infty[$, that $\mathcal{F}(t) \geq \mathcal{F}(|t|)$ and that for every $s \geq 0$,

$$\frac{1}{4}\min(s^2, s) \leq \mathcal{F}(s) \leq \min(s^2, s). \tag{2.3}$$

But we find it more convenient to work with the convex function $\mathcal{F}(|t|)$ rather than with $\mathcal{N}(|t|) = \min(t^2, |t|)$.

The treatment of the remainder term is stated in the next, central, lemma, which is of independent interest.

Lemma 2.2 *Let μ be a probability mesure on $\mathbb{R}^n$ absolutely continuous with respect to the Lebesgue measure and $\theta \in W^{2,1}_{\mathrm{loc}}(\mathbb{R}^n)$ with $D^2\theta \geq -\mathrm{Id}$ almost everywhere. We assume that $|\nabla\theta| \in L^1(\mu)$ with $\int \nabla\theta \, d\mu = 0$. Then*

$$\int \mathrm{tr}(\mathcal{F}(D^2\theta))\, d\mu \geq c \int \mathcal{F}\big(h(\mu)|\nabla\theta|\big)\, d\mu \tag{2.4}$$

for some numerical constant $c > 0$.

Note that our assumption $\int x\, d\mu_V = \int x\, d\nu$ can be rewritten as $\int \nabla\theta\, d\mu_V = 0$, so if we use (2.2) and the previous lemma with $\mu = \mu_V$ and θ our displacement function, we find

$$H(\nu\|\mu_V) \geq \int \widetilde{\mathbf{c}}_V(x, T(x))\, d\mu_V \geq W_{\widetilde{\mathbf{c}}_V}(\mu_V, \nu),$$

as claimed in Theorem 1.3.

So it only remains to prove Lemma 2.2.

Denote by σ the uniform probability measure on S^{n-1}. Recall that for every vector $X \in \mathbb{R}^n$, we have

$$n \int_{S^{n-1}} (X\cdot u)^2\, d\sigma(u) = |X|^2$$

and that

$$c|X| \leq \sqrt{n} \int_{S^{n-1}} |X\cdot u|\, d\sigma(u) \leq |X| \tag{2.5}$$

for some numerical constant $c > 0$.

We will use the following fact, the proof of which is postponed below.

Fact 2.3 *Let A be a symmetric matrix with eigenvalues > -1. Then*

$$\mathrm{tr}(\mathcal{F}(A)) \geq \frac{1}{8}\int_{S^{n-1}} \mathcal{F}\big(\sqrt{n}|Au|\big)\, d\sigma(u).$$

We will combine this with the following isoperimetric type inequality. It is due to Bobkov and Houdré [5], where it is stated with the median. We will include a proof below for completeness.

Fact 2.4 (Bobkov–Houdré) *Let μ be a probability measure on $\mathbb{R}^n$. For every regular enough $f:\mathbb{R}^n \to \mathbb{R}$ we have*

$$\int \mathcal{F}\big(\big|f - \textstyle\int f\,d\mu\big|\big)\,d\mu \le c\int \mathcal{F}\Big(\frac{1}{h(\mu)}|\nabla f|\Big)\,d\mu \tag{2.6}$$

for some numerical constant $c > 0$ (for instance, $c = 3\times 4^3$ works).

With these two facts in hand, we can now finish the proof of (2.4). We have, by Fact 2.3 and Fubini's theorem, that

$$\int \operatorname{tr}(\mathcal{F}(D^2\theta))\,d\mu \ge \frac{1}{8}\int_{S^{n-1}}\int \mathcal{F}\big(\sqrt{n}|D^2\theta u|\big)\,d\mu\,d\sigma(u). \tag{2.7}$$

For any fixed vector $u \in S^{n-1}$, we have that the function $g(x) = h(\mu)\sqrt{n}\,\nabla\theta(x)\cdot u$ is $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n)$ with derivative $\nabla g(x) = h(\mu)\sqrt{n}\,(D^2\theta(x))u$, and $\int g\,d\mu_V = 0$. Applying Fact 2.4, we have that

$$\int \mathcal{F}\big(\sqrt{n}|(D^2\theta)u|\big)\,d\mu \ge \frac{1}{3\times 4^3}\int \mathcal{F}(h(\mu)\sqrt{n}\,|\nabla\theta\cdot u|)\,d\mu.$$

Back to (2.7), integrating the previous inequality with respect to $d\sigma(u)$, using that $\mathcal{F}$ is convex and (2.5), we find

$$\begin{aligned}\int \operatorname{tr}(\mathcal{F}(D^2\theta))\,d\mu &\ge \frac{1}{3\times 4^3\times 8}\int \mathcal{F}\Big(h(\mu)\sqrt{n}\int_{S^{n-1}}|\nabla\theta\cdot u|\,d\sigma(u)\Big)\,d\mu\\ &\ge c\int \mathcal{F}(h(\mu)|\nabla\theta|)\,d\mu.\end{aligned}$$

This completes the proof of Lemma 2.2, modulo the two facts above that we now prove.

Proof of Fact 2.3 Let us first collect some straightforward properties of $\mathcal{F}$, or equivalently, in view of (2.3), of $\min(s^2, |s|)$. These functions commute with power functions. In particular, we shall use that

$$\forall s \ge 0,\qquad \mathcal{F}(\sqrt{s}) \le \sqrt{\mathcal{F}(s)} \le 2\mathcal{F}(\sqrt{s}). \tag{2.8}$$

Note that $\mathcal{F}(2s) \le 4\mathcal{F}(s)$. Observe also that for a finite family $s_1,\dots,s_k \ge 0$, we have

$$\sum_{i\le k}\mathcal{F}(s_i) \ge \frac{1}{4}\mathcal{F}\Big(\sqrt{\sum_{i\le k}s_i^2}\Big). \tag{2.9}$$

Indeed, if we denote by $s_{\max}$ the largest number, we see using (2.3) that

$$\sum_{i\le k}\mathcal{F}(s_i) \ge \frac{1}{4}\sum_{i\le k}\min(s_i, s_i^2) \ge \frac{1}{4}\min(s_{\max}, s_{\max}^2)\sum_{i\le k}\Big(\frac{s_i}{s_{\max}}\Big)^2.$$

Then distinguishing between $s_{\max} \le 1$, and $s_{\max} \ge 1$, a case for which we replace it by using $s_{\max} \le \sqrt{\sum_{i\le k}s_i^2}$, we find

$$\sum_{i\le k}\mathcal{F}(s_i) \ge \frac{1}{4}\min\Big(\sum_{i\le k}s_i^2, \sqrt{\sum_{i\le k}s_i^2}\Big) \ge \frac{1}{4}\mathcal{F}\Big(\sqrt{\sum_{i\le k}s_i^2}\Big).$$

Let us mention that inequality (2.9) will be the one responsible for the loss of a factor $\sqrt{n}$ in the case of product measures.

Back to the proof of Fact 2.4, let us notice that $\mathcal{F}(A) \geq \mathcal{F}(|A|)$, since $\mathcal{F}(s_i) \geq \mathcal{F}(|s_i|)$ for any eigenvalue s_i of A. Denote

$$H := |A| = \sqrt{A^*A};$$

it is a nonnegative symmetric matrix. Let $(e_1, \dots, e_n)$ be an orthonormal basis of eigenvectors of A. Then

$$\operatorname{tr}(\mathcal{F}(H)) = \sum_{i\leq n} |\mathcal{F}(H)e_i| = \sum_{i\leq n} \mathcal{F}(|He_i|) \geq \frac{1}{2}\sum_{i\leq n}\sqrt{\mathcal{F}}(|He_i|^2),$$

where we used (2.8). Let us mention in passing that using the convexity of $\mathcal{F}$, we can establish more generally that for for any $u \in S^{n-1}$ we have $|\mathcal{F}(H)u| \geq \frac{1}{2}\sqrt{\mathcal{F}(|Hv|^2)}$.

From this, using the fact that $\sqrt{\mathcal{F}}$ is <u>concave</u> on $\mathbb{R}^+$, then (2.8) again, and finally (2.9), we find that

$$\begin{aligned}
\operatorname{tr}(\mathcal{F}(H)) &\geq \frac{1}{2}\sum_{i\leq n}\sqrt{\mathcal{F}}\big(|He_i|^2\big) = \frac{1}{2}\sum_{i\leq n}\sqrt{\mathcal{F}}\Big(n\int_{S^{n-1}}(He_i\cdot u)^2\,d\sigma(u)\Big)\\
&\geq \frac{1}{2}\int_{S^{n-1}}\sum_{i\leq n}\sqrt{\mathcal{F}}(n(He_i\cdot u)^2)\,d\sigma(u) \geq \frac{1}{2}\int_{S^{n-1}}\sum_{i\leq n}\mathcal{F}(\sqrt{n}|He_i\cdot u|)\,d\sigma(u)\\
&\geq \frac{1}{8}\int_{S^{n-1}}\mathcal{F}\Big(\sqrt{n}\sqrt{\sum_{i\leq n}|He_i\cdot u|^2}\Big)\,d\sigma(u) = \frac{1}{8}\int_{S^{n-1}}\mathcal{F}\big(\sqrt{n}|Hu|\big)\,d\sigma(u).
\end{aligned}$$

To conclude, use that $|Hu|^2 = H^2u\cdot u = A^2u\cdot u = |Au|^2$. ∎

Proof of Fact 2.4 By scaling the metric, we can assume that $h(\mu) = 1$. More precisely, we can change the scalar produce $x\cdot y$ into $h(\mu)^{-1}x\cdot y$, which changes the gradient accordingly in (1.9) and (2.6).

Denote by m a μ-median of f. By a standard argument, it is enough to prove that

$$4\int \mathcal{F}\big(|f-m|\big)\,d\mu \leq 3\times 4^2\int \mathcal{F}(|\nabla f|)\,d\mu.$$

Indeed, since $\mathcal{F}(2t) \leq 4\mathcal{F}(t)$ and since $\mathcal{F}$ is convex increasing on $\mathbb{R}^+$, we have for any function g with μ-median m_g:

$$\begin{aligned}
\int \mathcal{F}\big(\big|g - \textstyle\int g\,d\mu\big|\big)\,d\mu &\leq 2\int \mathcal{F}(|g-m_g|)\,d\mu + 2\mathcal{F}\big(\big|\textstyle\int(g-m_g)\,d\mu\big|\big)\\
&\leq 4\int \mathcal{F}(|g-m_g|)\,d\mu.
\end{aligned}$$

The same kind of argument shows that one can use a median m_g instead of the mean, in the definition (1.9). Indeed, for any $g\in L^1(\mu)$ with median m_g, we have

$$\int |g-m_g|\,d\mu \leq \int \big|g - \textstyle\int g\,d\mu\big|\,d\mu + \big|\textstyle\int g\,d\mu - m_g\big|.$$

We can assume that $m_g \geq \int g\,d\mu$ (otherwise use $-g$), and by the definition of m_g and by Markov's inequality we have

$$\frac{1}{2} \leq \mu\big(\{g\geq m_g\}\big) \leq \mu\big(\{|g - \textstyle\int g\,d\mu| \geq m_g - \int g\,d\mu\}\big) \leq \frac{1}{m_g - \int g\,d\mu}\displaystyle\int \big|g-\textstyle\int g\,d\mu\big|\,d\mu,$$

https://doi.org/10.4153/CJM-2016-046-3 Published online by Cambridge University Press

and so $\left|m_g - \int g\,d\mu\right| \le 2\int \left|g - \int g\,d\mu\right| d\mu$. Therefore, we have

$$\int |g - m_g|\,d\mu \le 3\int \left|g - \int g\,d\mu\right| d\mu \le 3\int |\nabla g|\,d\mu.$$

Given our f with μ-median m, let us introduce the (continuous) function g such that

$$g(x) = \begin{cases} \mathcal{F}(|f(x) - m|) & \text{if } f(x) \ge m, \\ -\mathcal{F}\big(|f(x) - m|\big) & \text{if } f(x) < m. \end{cases}$$

Since $\mathcal{F} \ge 0$, the function g has zero μ-median. Therefore,

$$\int \mathcal{F}(|f - m|)\,d\mu = \int |g|\,d\mu \le 3\int |\nabla g|\,d\mu. \tag{2.10}$$

We will now use an argument inspired by [20]. Let us observe that for every $s \in \mathbb{R}^+$, $t \in [0,1]$ (this is the only good choice to estimate the Legendre transform of $\mathcal{F}$), we have

$$st \le 4\mathcal{F}(s) + \frac{1}{16}t^2.$$

Indeed, for $s \ge 1$ the inequality is obvious, since $4\mathcal{F}(s) \ge s \ge st$, and for $s < 1$, use that $4\mathcal{F}(s) \ge s^2$ to complete the square. Since $\mathcal{F}' \in [0,1]$ on $\mathbb{R}^+$, we have

$$\begin{aligned}\int |\nabla g|\,d\mu = \int \mathcal{F}'(|f - m|)\,|\nabla f|\,d\mu &\le 4\int \mathcal{F}(|\nabla f|)\,d\mu + \frac{1}{16}\int \mathcal{F}'(|f - m|)^2\,d\mu \\ &\le 4\int \mathcal{F}(|\nabla f|)\,d\mu + \frac{1}{4}\int \mathcal{F}(|f - m|)\,d\mu,\end{aligned}$$

where the second inequality follows from $\mathcal{F}'(s)^2 \le 4\mathcal{F}(s)$ for every $s \in \mathbb{R}^+$ (this can be seen, for instance, by computing $(4\mathcal{F} - \mathcal{F}'^2)'(s) = 2s(1 + 2s + 2s^2)/(1 + s)^3 \ge 0$). Plugging this into (2.10), we find

$$\frac{1}{4}\int \mathcal{F}(|f - m|)\,d\mu \le 3 \times 4\int \mathcal{F}(|\nabla f|)\,d\mu,$$

which gives the desired inequality. ■

This completes the proof of Lemma 2.2 and Theorem 1.3. It only remains to treat the cases of equality in Proposition 1.1.

Determination of equality cases in Proposition 1.1 The idea is that if equality holds in Proposition 1.1, and if ν and μ_V have the same barycenter (a situation that can be imposed by translating ν, provided we know that translation preserves equality cases), then we can apply Theorem 1.3 and conclude that $W_1(\mu_V, \nu) = 0$, which implies $\nu = \mu_V$. Oddly enough, the converse also requires some work; even the fact that there is equality when $\nu = \mu_V$ is not straightforward, and actually requires the convexity of V.

We will prove the stronger result of Remark 1.2. Given a vector $u \in \mathbb{R}^n$, let us denote by $T_u\nu$ the translation by u of the probability ν; if $d\nu(x) = F(x)\,dx$, then $dT_u\nu(x) = F(x - u)\,dx$. The following lemma is essential as it establishes the translation invariance of the inequality under study.

Lemma 2.5 (Translation invariance) *With the notation of Proposition 1.1, we have, for any probability ν and any vector $u \in \mathbb{R}^n$, that*

$$H(T_u \nu \| \mu_V) - \mathcal{W}_{\mathbf{c}_V}(\mu_V, T_u \nu) = H(\nu \| \mu_V) - \mathcal{W}_{\mathbf{c}_V}(\mu_V, \nu).$$

Proof To simplify the notation, we can assume that $\int e^{-V} = 1$, and also that $d\nu(x) = f(x) d\mu_V(x) = f(x) e^{-V(x)} dx$. To treat the transportation term, we will need the following observation.

Fact 2.6 *Given $u \in \mathbb{R}^n$, introduce $\widetilde{u} = (0, u) \in \mathbb{R}^{2n}$. Let μ and ν be two probability measures on $\mathbb{R}^n$ and let $\mathbf{c}_V$ be the cost from Proposition 1.1. If π is a $\mathbf{c}_V$-optimal coupling for (μ, ν), then $T_{\widetilde{u}}\pi$ is a $\mathbf{c}_V$-optimal coupling for $(\mu, T_u \nu)$.*

Let us prove this fact. The coupling condition is clear, so we only need to check that $T_{\widetilde{u}}\pi$ is $\mathbf{c}_V$-optimal when π is. Equivalently, by the characterization of optimality in terms of cyclical monotony (see [32, Chapter 5]), it suffices to check that the support of $T_{\widetilde{u}}\pi$ is $\mathbf{c}_V$-cyclically monotone when the support of π is $\mathbf{c}_V$-cyclically monotone. Let $(x_1, y_1), \dots, (x_k, y_k)$ be arbitrary points of $\mathbb{R}^{2n}$, with the convention that $(x_{k+1}, y_{k+1}) := (x_1, y_1)$. We have

$$\begin{aligned}\sum_{i=1}^k \mathbf{c}_V(x_i, y_{i+1} + u) - \sum_{i=1}^k \mathbf{c}_V(x_i, y_i + u) &= -\sum_{i=1}^k \nabla V(x_i) \cdot (y_{i+1} - y_i) \\ &= \sum_{i=1}^k \mathbf{c}_V(x_i, y_{i+1}) - \sum_{i=1}^k \mathbf{c}_V(x_i, y_i),\end{aligned}$$

which shows that, indeed, the support of $T_{\widetilde{u}}\pi$ is $\mathbf{c}_V$-cyclically monotone if and only if the support of π is $\mathbf{c}_V$-cyclically monotone.

With this fact in hand, let us finish the proof of Lemma 2.5. Let π be a $\mathbf{c}_V$-optimal coupling for (μ_V, ν). Then by the previous Fact we have that

$$\begin{aligned}\mathcal{W}_{\mathbf{c}_V}(\mu_V, T_u \nu) - \mathcal{W}_{\mathbf{c}_V}(\mu_V, \nu) &= \iint \big[\mathbf{c}_V(x, y + u) - \mathbf{c}_V(x, y) \big] \, d\pi(x) \\ &= \int \big[V(y + u) - V(y) \big] \, d\nu(y),\end{aligned}$$

where we used that $\iint \nabla V(x) \cdot u \, d\pi(x, y) = \int \nabla V(x) \cdot u \, e^{-V(x)} \, dx = 0$.

The entropic terms are easier to analyse. Since $dT_u \nu(x) = f(x - u) e^{-V(x-u)} \, dx = f(x - u) e^{-V(x-u) + V(x)} d\mu_V(x)$, we have

$$\begin{aligned}&H(T_u \nu \| \mu_V) - H(\nu \| \mu_V) \\ &\quad = \int \big[\log(f(x - u)) - V(x - u) + V(x) \big] f(x - u) e^{-V(x-u)} \, dx \\ &\qquad - \int \log f(x) \log(f(x)) e^{-V(x)} \, dx \\ &\quad = \int \big[- V(x) + V(x + u) \big] \, d\nu(x).\end{aligned}$$

By subtracting the previous two equations, we obtain the conclusion of Lemma 2.5. ∎

Next, the role of the convexity of V can be summarized as follows.

Lemma 2.7 *Let $V:\mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function satisfying (1.1) and let $\mathbf{c}_V$ be the cost given by (1.2), which is well defined for almost every $x \in \mathbb{R}^n$. If there exists an absolutely continuous probability measure μ with support $\mathbb{R}^n$ such that $\mathcal{W}_{\mathbf{c}_V}(\mu,\mu) = 0$, then V is convex on $\mathbb{R}^n$. Conversely, if V is convex, then $\mathcal{W}_{\mathbf{c}_V}(\mu,\mu) = 0$ for every absolutely continuous probability measure μ.*

Proof Since $\mathbf{c}(x,x) = 0$, if $\mathcal{W}_{\mathbf{c}}(\mu,\mu) = 0$, then it means that the image π of μ by the map $x \to (x,x)$ is an optimal coupling, and therefore its support is $\mathbf{c}_V$-cyclically monotone. By the assumption on μ, this implies that for (almost) all $x, y \in \mathbb{R}^n$ we have $\mathbf{c}_V(x,y) + \mathbf{c}_V(y,x) \geq \mathbf{c}_V(x,x) + \mathbf{c}_V(y,y)$, which can be rewritten as

$$\big(\nabla V(y) - \nabla V(x)\big)\cdot(y-x) \geq 0. \tag{2.11}$$

For a locally Lipschitz function (therefore also $W^{1,1}_{\mathrm{loc}}$), this property implies that V is convex. Indeed, we can consider $V_\epsilon = V * \eta_\epsilon$ where $\eta_\epsilon(x) = \epsilon^{-n}\eta(x/\epsilon)$ is an approximation of the identity in $\mathbb{R}^n$, with η compactly supported. Then the property (2.11) passes to V_ϵ, which is now smooth, so that this property holds at every (x,y) and this implies that V_ϵ is convex on $\mathbb{R}^n$, because it implies that the restriction of V_ϵ to any affine line has a nondecreasing derivative. We conclude by using that V_ϵ converges to V, point-wise as $\epsilon \to 0$.

Conversely, the fact that $\mathbf{c}_V(x,x) = 0$ implies that $\mathcal{W}_{\mathbf{c}_V}(\mu,\mu) \leq 0$ for any absolutely continuous probability measure μ. Since $\mathbf{c}_V \geq 0$ when V is convex, we get in this case that $\mathcal{W}_{\mathbf{c}_V}(\mu,\mu) = 0$. (One can also verify that when V is convex, the set $\{(x,x); x \in \mathbb{R}^n\}$ is $\mathbf{c}_V$-cyclically monotone.) ∎

We now have all the ingredients for the study of equality cases. If there is equality in (1.3) for some ν, then by Lemma 2.5 there must be equality for any translated measure $T_u\nu$, $u \in \mathbb{R}^n$. But for $u := -\int x\,d\nu + \int x\,d\mu_V$, we have the centering condition $\int x\,dT_u\nu = \int x\,d\mu_V(x)$, and so we must have that $W_1(\mu_V, T_u\nu) = 0$, that is, $T_u\nu = \mu_V$ or equivalently $\nu = T_{-u}\mu_V$. This shows that for equality to hold, ν must be a translate of μ_V. But this in turn implies, again by Lemma 2.5, that there is also equality for $\nu = \mu_V$. Since $H(\mu_V\|\mu_V) = 0$, we must have $\mathcal{W}_{\mathbf{c}_V}(\mu_V,\mu_V) = 0$. By Lemma 2.7, this implies that V is convex.

Conversely, if V is convex, there is equality for $\nu = \mu$, because Lemma 2.7 ensures that $\mathcal{W}_{\mathbf{c}_V}(\mu_V,\mu_V) = 0$, and by Lemma 2.5 we then also have equality for any translate of μ_V. ∎

3 Variance Brascamp–Lieb Inequalities

It is well known that linearization of transportation type inequalities give Poincaré type inequalities. One often uses the dual infimal convolution inequality (1.5) to perform the linearization, but one can do it also directly from the transportation inequality. The procedure for linearizing the Wasserstein distance is standard, especially in the framework of the so-called "Otto calculus" (see, for instance, [28]). It is also known that only the local behavior of the cost matters for linearizing a transport inequality (see, for instance, [18, Section 8.3]. However we did not find a reference for the precise situation studied here, and so we include the following statement for completeness.

Lemma 3.1 *Let* $\mathbf{c}\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ *be a function such that* $\mathbf{c}(y,y) = 0$ *and* $\mathbf{c}(x,y) \geq \delta_0|x-y|^2$ *for every* $x, y \in \mathbb{R}^n$, *for some* $\delta_0 > 0$. *Assume furthermore that for every* y *there exists a nonnegative symmetric operator* H_y *for which*

$$\mathbf{c}(y+h,y) = \frac{1}{2}H_y h \cdot h + |h|^2 o(1)$$

uniformly in y *on compact sets when* $h \to 0$.

Then if μ *is a probability measure on* $\mathbb{R}^n$ *and* g *is a* C^1 *compactly supported function with* $\int g\,d\mu = 0$, *we have*

$$\liminf_{\varepsilon\to 0} \frac{1}{\varepsilon^2}\mathcal{W}_{\mathbf{c}}(\mu,(1+\varepsilon g)\,d\mu) \geq \frac{1}{2}\frac{\left(\int g f\,d\mu\right)^2}{\int H^{-1}\nabla f\cdot\nabla f\,d\mu}$$

for any C^1 *compactly supported function* f.

Proof Given a (bounded) function F on $\mathbb{R}^n$, we introduce its infimal convolution (1.6) associated with our cost $\mathbf{c}$, which satisfies: for every $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$, $Q_{\mathbf{c}}(F)(y) - F(x) \leq \mathbf{c}(x,y)$. It then follows from the definition of $\mathcal{W}_{\mathbf{c}}$ that

$$\mathcal{W}_{\mathbf{c}}(\mu,\nu) \geq \int Q_{\mathbf{c}}(F)\,d\nu - \int F\,d\mu.$$

In our situation where $\nu = (1+\varepsilon g)\,d\mu$ (ε small enough and later tending to 0), we pick $F = \varepsilon f$ with f of class C^1 compactly supported and $\int f\,d\mu = 0$. Let us write

$$Q_c(\varepsilon f)(y) = \inf_x\left\{\varepsilon f(x) + \mathbf{c}(x,y)\right\} = \inf_h\left\{\varepsilon f(y+h) + \mathbf{c}(y+h,y)\right\}.$$

For any given y, let $h_\varepsilon = h_{\varepsilon,y}$ be a point where this infimum is achieved. Since the function f is Lipschitz, of constant $M > 0$ say, we have by our assumption on the cost that

$$\varepsilon f(y) - \varepsilon M|h_\varepsilon| + \delta_0|h_\varepsilon|^2 \leq \varepsilon f(y+h_\varepsilon) + c(y+h_\varepsilon,y) \leq \varepsilon f(y)$$

so that

$$|h_\varepsilon| \leq \frac{M}{\delta_0}\varepsilon.$$

In other words, h_ε tends to zero like ε uniformly in y. Also, since f is continuous compactly supported, we can find (because the cost is nonnegative and large when points are far-apart) a bounded open set Ω, which contains the support of f such that $Q_{\mathbf{c}}(\varepsilon f)(y) \geq 0$ for every $y \in \mathbb{R}^n \setminus \Omega$. Consequently, we have

$$\mathcal{W}_{\mathbf{c}}(\mu,(1+\varepsilon g)\,d\mu) \geq \int Q_{\mathbf{c}}(\varepsilon f)\,(1+\varepsilon g)\,d\mu \geq \int_\Omega Q_{\mathbf{c}}(\varepsilon f)\,(1+\varepsilon g)\,d\mu.$$

We have, uniformly for y in the bounded set Ω,

$$Q_{\mathbf{c}}(\varepsilon f)(y) = \varepsilon f(y+h_\varepsilon) + c(y+h_\varepsilon,y) = \varepsilon f(y) + \varepsilon\nabla f(y)\cdot h_\varepsilon + \frac{1}{2}H_y h_\varepsilon\cdot h_\varepsilon + o(\varepsilon^2),$$

and so

$$Q_{\mathbf{c}}(\varepsilon f)(y) \geq \varepsilon f(y) - \varepsilon^2\frac{1}{2}H_y^{-1}\nabla f(y)\cdot\nabla f(y) + o(\varepsilon^2).$$

https://doi.org/10.4153/CJM-2016-046-3 Published online by Cambridge University Press

After multiplying by $(1+\varepsilon g)$, we can integrate on Ω using that the $o(\varepsilon^2)$ is uniform in y:

$$\frac{1}{\varepsilon^2}\mathcal{W}_{\mathbf{c}}(\mu,(1+\varepsilon g)\,d\mu) \ge \int_\Omega f\,g\,d\mu - \frac{1}{2}\int_\Omega H_y^{-1}\nabla f\cdot\nabla f\,d\mu + o(1).$$

This implies, using that Ω contains the support of f, that

$$\liminf_{\varepsilon\to 0}\frac{1}{\varepsilon^2}\mathcal{W}_{\mathbf{c}}(\mu,(1+\varepsilon g)\,d\mu) \ge \int f\,g\,d\mu - \frac{1}{2}\int H_y^{-1}\nabla f\cdot\nabla f\,d\mu.$$

The result follows by homogeneity (replacing f by λf and optimizing). ∎

The linearization given in Lemma 3.1 shows that the Brascamp–Lieb inequality follows immediately from Proposition 1.1 for $\nu = (1+\varepsilon g)d\mu_V$ with $\int g d\mu_V = 0$, when $\varepsilon \to 0$. Indeed, without loss of generality, we can assume that $D^2V \ge 2\delta_0$ (by adding a small $\delta_0|x|^2$ and later making $\delta_0 \to 0$) so that the cost verifies also $\mathbf{c}_V(x,y) \ge \delta_0|x-y|^2$. Moreover, if V is C^2, we see from the definition (1.2) of the cost that, when $h\to 0$,

$$\mathbf{c}_V(y+h,y) - \frac{1}{2}D^2V(y)h\cdot h = \int_0^1\Big[D^2V(y+(1-t)h) - D^2V(y)\Big]h\cdot h\,(1-t)\,dt = |h|^2 o(1),$$

where the $o(1)$ is uniform in y on compact sets, since D^2V is uniformly continuous on compact sets. On the other hand, if g is a C^1 compactly supported function with $\int g\,d\mu_V = 0$, we have

$$H\big((1+\varepsilon g)d\mu_V|\mu_V\big) = \frac{1}{2}\varepsilon^2\int g^2\,d\mu_V + o(\varepsilon^2).$$

So we find, by applying Proposition 1.1 with $\nu = (1+\varepsilon g)d\mu_V$ and Lemma 3.1 with the choice $f = g$ (which is the optimal one in the present situation), at the limit, that

$$\frac{1}{2}\frac{\big(\int g^2\,d\mu_V\big)^2}{\int (D^2V(x))^{-1}\nabla g\cdot\nabla g\,d\mu_V} \le \frac{1}{2}\int g^2\,d\mu_V,$$

which is the Brascamp–Lieb inequality (1.7).

Let us apply the same procedure with inequality (1.11) in Theorem 1.3, the crucial point being that $\mathcal{N}(h(\mu_V)\,|(y+h)-y|)$ behaves like $h(\mu_V)^2|h|^2$ for h small. So the cost satisfies, for $h\to 0$,

$$\begin{aligned}\widetilde{\mathbf{c}}_V(y+h,y) &= \mathbf{c}_V(y+h,y) + c\,\mathcal{N}(h(\mu_V)\,|h|) = \mathbf{c}_V(y+h,y) + c\,h(\mu_V)^2|h|^2 + o(h^2)\\ &= \frac{1}{2}\Big[D^2V(y)h\cdot h + 2c\,h(\mu_V)^2\,\mathrm{Id}\Big]h\cdot h + o(h^2),\end{aligned}$$

where c is a numerical constant. The same argument as before for $\nu = (1+\varepsilon g)d\mu_V$ shows that if g is a C^1 compactly supported function with $\int g\,d\mu_V = 0$ and

$$\int xg(x)\,d\mu_V(x) = 0,$$

we have

$$\int g^2\,d\mu_V \le \int \big(D^2V + c\,h(\mu_V)^2\,\mathrm{Id}\big)^{-1}\nabla g\cdot\nabla g\,d\mu_V,$$

as claimed in Proposition 1.5.

Finally, let us derive Proposition 1.6. With the notation of the proposition, for given g, denote $g_0 := g - \nabla V \cdot u_0 - c_0$. It is readily checked by elementary calculus that for every vector u_0 and constant c_0 (so not only for the ones we have picked), if $g = g_0 + \nabla V \cdot u_0 + c_0$,

$$\begin{aligned} A &:= \int (D^2V(x))^{-1} \nabla g(x) \cdot \nabla g(x) \, d\mu_V(x) - \mathrm{Var}_{\mu_V}(g) \\ &= \int (D^2V(x))^{-1} \nabla g_0(x) \cdot \nabla g_0(x) \, d\mu_V(x) - \mathrm{Var}_{\mu_V}(g_0). \end{aligned}$$

Next, for our choice of u_0 and c_0 observe that $\int g_0 \, d\mu_V = 0$ and

$$\int x \, g_0(x) \, d\mu(x) = 0,$$

since, in the standard basis, writing $x_j = x \cdot e_j$ for $j = 1, \ldots, n$, we have

$$\int x_j \nabla V(x) \cdot u_0 \, d\mu_V(x) = \int e_j \cdot u_0 \, d\mu_V = e_j \cdot u_0 = \int x_j (g(x) - c_0) \, d\mu_V(x).$$

So by Proposition 1.5, we find

$$\begin{aligned} A &\geq \int (D^2V)^{-1} \nabla g_0 \cdot \nabla g_0 \, d\mu_V - \int \big(D^2V + c\lambda(\mu_V) \, \mathrm{Id} \big)^{-1} \nabla g_0 \cdot \nabla g_0 \, d\mu_V \\ &= c\lambda(\mu_V) \int (D^2V)^{-1} \big(D^2V + c\lambda(\mu_V) \, \mathrm{Id} \big)^{-1} \nabla g_0 \cdot \nabla g_0 \, d\mu_V. \end{aligned}$$

From this bound, we can proceed in two different ways. First, we can use a uniform lower bound and combine it with the Brascamp–Lieb inequality

$$\begin{aligned} A &\geq \frac{c\lambda(\mu_V)}{\sup_x \lambda_{\max}(x) + c\lambda(\mu_V)} \int (D^2V)^{-1} \nabla g_0 \cdot \nabla g_0 \, d\mu_V \\ &\geq \frac{c\lambda(\mu_V)}{\sup_x \lambda_{\max}(x) + c\lambda(\mu_V)} \int |g_0|^2 \, d\mu_V. \end{aligned}$$

Otherwise, using again that $D^2V \leq \lambda_{\max} \, \mathrm{Id}$, we can use Hölder's inequality, to arrive at

$$A \geq \frac{\big(\int |\nabla g_0| \, d\mu_V \big)^2}{\int \lambda_{\max} (\lambda_{\max} + c\lambda(\mu_V)) \, d\mu_V}.$$

But (1.10) implies that $\int |\nabla g_0| \, d\mu_V \geq c\sqrt{\lambda(\mu_V)} \int |g_0| \, d\mu_V$ for some numerical constant $c > 0$. This ends the proof of Proposition 1.6. ■

Acknowledgments We would like to thank Bernard Maurey for useful observations on our manuscript, and the anonymous referee for several insightful questions and observations that led to improved statements. We would like also to express our deep gratitude to Emanuel Milman for his most sharp reading of the first version of our preprint. He pointed out a fatal mistake in the use we made of his results; this led us to rewrite the main statements and their proofs.

References

[1] F. Barthe and D. Cordero-Erausquin, *Invariances in variance estimates*. Proc. Lond. Math. Soc. (3) **106**(2013), 33–64. http://dx.doi.org/10.1112/plms/pds011

[2] F. Barthe and A. V. Kolesnikov, *Mass transport and variants of the logarithmic Sobolev inequality*. J. Geom. Anal. **18**(2008), no. 4, 921–979. http://dx.doi.org/10.1007/s12220-008-9039-6

[3] G. Blower, *The Gaussian isoperimetric inequality and transportation*. Positivity **7**(2003), 203–224. http://dx.doi.org/10.1023/A:1026242611940

[4] S. G. Bobkov, N. Gozlan, C. Roberto, and P.-M. Samson, *Bounds on the deficit in the logarithmic Sobolev inequality*. J. Funct. Anal. **267**(2014), no. 11, 4110–4138. http://dx.doi.org/10.1016/j.jfa.2014.09.016

[5] S. G. Bobkov and C. Houdré, *Some connections between isoperimetric and Sobolev-type inequalities*. Mem. Amer. Math. Soc. **129**(1997), no. 616. http://dx.doi.org/10.1090/memo/0616

[6] S. G. Bobkov and M. Ledoux, *From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities*. Geom. Funct. Anal. **10**(2000), no. 5, 1028–1052. http://dx.doi.org/10.1007/PL00001645

[7] F. Bolley, I. Gentil, and A. Guillin, *Dimensional improvements of the logarithmic Sobolev, Talagrand and Brascamp-Lieb inequalities*. arxiv:1507.01086

[8] H. J. Brascamp and E. H. Lieb, *On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an applicationto the diffusion equation.* J. Funct. Anal. **22**(1976), no. 4, 366–389. http://dx.doi.org/10.1016/0022-1236(76)90004-5

[9] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions.* Comm. Pure Appl. Math. **44**(1991), no. 4, 375–417. http://dx.doi.org/10.1002/cpa.3160440402

[10] D. Cordero-Erausquin, *Some applications of mass transport to Gaussian-type inequalities.* Arch. Ration. Mech. Anal. **161**(2002), no. 3, 257–269. http://dx.doi.org/10.1007/s002050100185

[11] D. Cordero-Erausquin and N. Gozlan, *Transport proofs of weighted Poincaré inequalities for log-concave distributions*. Bernouilli **23**(2017), no. 1, 134–158. http://dx.doi.org/10.3150/15-BEJ739

[12] D. Cordero-Erausquin and B. Klartag, *Moment measures*. J. Funct. Anal. **268**(2015), no. 12, 3834–3866. http://dx.doi.org/10.1016/j.jfa.2015.04.001

[13] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions.* Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992.

[14] M. Fathi, E. Indrei, and M. Ledoux, *Quantitative logarithmic Sobolev inequalities and stability estimates*. Discrete Contin. Dyn. Syst. **36**(2016), no. 12, 6835–6853. http://dx.doi.org/10.3934/dcds.2016097

[15] A. Figalli, F. Maggi, and A. Pratelli, *A refined Brunn-Minkowski inequality for convex sets*. Ann. Inst. H. Poincaré Anal. Non Linéaire **26**(2009), 2511–2519. http://dx.doi.org/10.1016/j.anihpc.2009.07.004

[16] G. De Philippis and A. Figalli, *The Monge-Ampère equation and its link to optimal transportation*. Bull. Amer. Math. Soc. (N.S.) **51**(2014), no. 4, 527–580. http://dx.doi.org/10.1090/S0273-0979-2014-01459-4

[17] I. Gentil, *From the Prékopa-Leindler inequality to modified logarithmic Sobolev inequality*. Ann. Fac. Sci. Toulouse Math. (6) **17**(2008), 291–308. http://dx.doi.org/10.5802/afst.1184

[18] N. Gozlan and C. Léonard, *Transport inequalities. A survey*. Markov Process. Related Fields **16**(2010), no. 4, 635–736.

[19] G. Hargé, *Reinforcement of an inequality due to Brascamp and Lieb*. J. Funct. Anal. **254**(2008), no. 2, 267–300. http://dx.doi.org/10.1016/j.jfa.2007.07.019

[20] B. Klartag, *Concentration of measures supported on the cube*. Israel J. Math. **203**(2014), no. 1, 59–80. http://dx.doi.org/10.1007/s11856-013-0072-1

[21] M. Ledoux, *The concentration of measure phenomenon*. Mathematical Surveys and Monographs, 89, American Mathematical Society, Providence, RI, 2001.

[22] ———, *Spectral gap, logarithmic Sobolev constant, and geometric bounds.* In: Surveys in differential Geometry, IX, Int. Press, Somerville, MA, 2004, pp. 219–240. http://dx.doi.org/10.4310/SDG.2004.v9.n1.a6

[23] R. J. McCann, *Existence and uniqueness of monotone measure-preserving maps.* Duke Math. J. **80**(1995), 309–323. http://dx.doi.org/10.1215/S0012-7094-95-08013-2

[24] ———, *A convexity principle for interacting gases.* Adv. Math. **128**(1997), no. 1, 153–179. http://dx.doi.org/10.1006/aima.1997.1634

[25] E. Milman, *On the role of convexity in isoperimetry, spectral gap and concentration*. Invent. Math. **177**(2009), no. 1, 1–43. http://dx.doi.org/10.1007/s00222-009-0175-9

[26] E. Milman and A. V. Kolesnikov, *Poincaré and Brunn-Minkowski inequalities on weighted Riemannian manifolds with boundary*. arxiv:1310.2526

[27] C. Mooney, *Partial regularity for singular solutions to the Monge-Ampère equation*. Comm. Pure Appl. Math. **68**(2015), no. 6, 1066–1084. http://dx.doi.org/10.1002/cpa.21534

[28] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*. J. Funct. Anal. **173**(2000), 361–400. http://dx.doi.org/10.1006/jfan.1999.3557

[29] R. T. Rockafellar, *Convex analysis*. Princeton Mathematical Series, 28, Princeton University Press, Princeton, NJ, 1970.

[30] M. Talagrand, *Transportation cost for Gaussian and other product measures*. Geom. Funct. Anal. **6**(1996), 587–600. http://dx.doi.org/10.1007/BF02249265

[31] L. Veysseire, *A harmonic mean bound for the spectral gap of the Laplacian on Riemannian manifolds*. C. R. Math. Acad. Sci. Paris **348**(2010), 1319–1322. http://dx.doi.org/10.1016/j.crma.2010.10.015

[32] C. Villani, *Optimal transport. Old and new*. Springer-Verlag, Berlin, 2009.

Institut de Mathématiques de Jussieu,, Université Pierre et Marie Curie - Paris 6,, 75252 Paris Cedex 05, France
e-mail: dario.cordero@imj-prg.fr