# A PROOF OF THE CAFFARELLI CONTRACTION THEOREM VIA ENTROPIC REGULARIZATION

MAX FATHI, NATHAEL GOZLAN, AND MAXIME PROD'HOMME

Abstract. We give a new proof of the Caffarelli contraction theorem, which states that the Brenier optimal transport map sending the standard Gaussian measure onto a uniformly log-concave probability measure is Lipschitz. The proof combines a recent variational characterization of Lipschitz transport map by the second author and Juillet with a convexity property of optimizers in the dual formulation of the entropy-regularized optimal transport (or Schrödinger) problem.

## 1. Introduction

The aim of the paper is to give a new proof of the celebrated Caffarelli contraction theorem [3, 4], which states that the Brenier optimal transport map sending the standard Gaussian measure on $\mathbb{R}^d$, denoted by $\gamma_d$ in all the paper, onto a probability measure $\nu$ having a log-concave density with respect to $\gamma_d$ is a contraction. More precisely, let us recall the generalized version of Caffarelli's theorem:

**Theorem 1** (Caffarelli [3, 4])**.** *For any probability measures $\mu$ and $\nu$ respectively of the form $\mu(dx) = e^{V(x)}\gamma_d(dx)$ and $\nu(dx) = e^{-W(x)}\gamma_d(dx)$ with $V$ and $W$ convex functions, and further assuming $\mu$ has a finite second moment and $\nu$ is compactly supported, there exists a continuously differentiable and convex function $\phi : \mathbb{R}^d \to \mathbb{R}$ such that $\nabla\phi$ is 1-Lipschitz and $\nu = \nabla\phi_{\#}\mu$.*

Caffarelli's original result was only stated for the important particular case where $\mu$ is the Gaussian measure $\gamma_d$ (i.e. $V = 0$), but his proof readily extends to this more general setting [32]. Note that the assumption that $\nu$ is compactly supported can be removed via approximation. See [43, Corollary 5.21] for details. Note that in all the paper we allow convex function to take the value $+\infty$.

This result plays an important role in the Functional Inequality literature, as it enables to transfer geometric inequalities such as Log-Sobolev or Gaussian Isoperimetric inequalities from the Gaussian measure to probability measures with a uniformly log-concave density. See [10, 25, 26, 38] for some applications of Theorem 1 to functional inequalities. It has also been used to derive deficit estimates in functional inequalities [14, 11]. Crucial for such applications is the dimension-free nature of the bound, to preserve the dimension-independent estimates that arise from these functional inequalities, and which are at the center of their applications in statistics for example. More recently, there have been some extensions, such as dimension-free Sobolev estimates [31, 32] and variants for compactly-supported perturbations of the Gaussian measure [7].

Caffarelli's original proof relies on the formulation of Brenier maps as solutions to a Monge-Ampère equation, and uses maximum principle-type estimates. In particular, it does not actually exploit the fact that $\nabla\phi$ is an optimal transport map. This is also the case for the other proofs [32, 29]. Our purpose here is to provide a different proof that does directly exploit ideas from optimal transport theory.

*Date*: April 6, 2020.
1991 *Mathematics Subject Classification.* 60A10, 49J55, 60G42.
*Key words and phrases.* Optimal transport, Caffarelli contraction Theorem, Schrödinger bridges.

In this paper, we develop an approach based on a variational characterization of Lipschitz regularity of optimal transport maps obtained by the second author and Juillet in [23]. To recall this result, we need to introduce some notations and definitions. We will denote by $\mathcal{P}(\mathbb{R}^d)$ the set of Borel probability measures on $\mathbb{R}^d$ and by $\mathcal{P}_k(\mathbb{R}^d)$, $k \geqslant 1$, the subset of $\mathcal{P}(\mathbb{R}^d)$ of probability measures having a finite moment of order $k$. The quadratic Kantorovich distance $W_2$ is defined for all $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ as follows:

$$W_2^2(\mu,\nu) = \inf_{\pi \in C(\mu,\nu)} \int |x-y|^2 \, \pi(dxdy),$$

where $|\cdot|$ denotes in all the paper the standard Euclidean norm and $C(\mu,\nu)$ is the set of couplings between $\mu$ and $\nu$, that is to say the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ such that $\pi(A \times \mathbb{R}^d) = \mu(A)$ and $\pi(\mathbb{R}^d \times B) = \nu(B)$, for all Borel sets $A, B$ of $\mathbb{R}^d$. Finally, if $\eta_1, \eta_2 \in \mathcal{P}_1(\mathbb{R}^d)$, one says that $\eta_1$ is dominated by $\eta_2$ for the convex order if $\int f \, d\eta_1 \leqslant \int f \, d\eta_2$ for all convex function $f : \mathbb{R}^d \to \mathbb{R}$. In this case, we write $\eta_1 \leqslant_c \eta_2$. With these notations in hand, the variational characterization of [23] reads as follows:

**Theorem 2.** *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ ; the following assertions are equivalent:*

(i) *There exists a continuously differentiable and convex function $\phi : \mathbb{R}^d \to \mathbb{R}$ such that $\nabla\phi$ is 1-Lipschitz and $\nu = \nabla\phi_{\#}\mu$,*
(ii) *For all $\eta \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\eta \leqslant_c \nu$,*

$$W_2(\nu,\mu) \leqslant W_2(\eta,\mu).$$

In other words the Brenier map between $\mu$ and $\nu$ is a contraction if and only if $\nu$ is the closest point to $\mu$ among all probability measures dominated by $\nu$ in the convex order. We will give an alternative proof of this theorem in Section 4 in the particular case where the support of $\nu$ is convex (which is enough for our purpose here) using Kantorovich duality and variational arguments.

Our strategy to recover Theorem 1 is thus to show the following monotonicity property of the $W_2$ distance: if $\mu$ and $\nu$ satisfy the assumptions of Theorem 1, it holds

$$W_2(\nu,\mu) \leqslant W_2(\eta,\mu), \qquad \forall \eta \leqslant_c \nu. \tag{1}$$

For that purpose, we will establish a similar inequality at an information theoretic level replacing $W_2$ by the so called entropic transport cost $\mathcal{T}_H^\varepsilon$ (presented in details in the next section) that is defined in terms of the minimization of the relative entropy between $\pi$ and a reference measure $R^\varepsilon$ involving some small noise parameter $\varepsilon > 0$. We will prove the following monotonicity property of the entropic cost:

$$\mathcal{T}_H^\varepsilon(\nu,\mu) \leqslant \mathcal{T}_H^\varepsilon(\eta,\mu) \tag{2}$$

for all $\eta \leqslant_c \nu$ with a finite Shannon information. As observed by Mikami [36] and extensively developed by Léonard [33, 34] the zero noise limit of $\varepsilon\mathcal{T}_H^\varepsilon$ is $\frac{1}{2}W_2^2$. Thus letting $\varepsilon \to 0$ in (2) will give (1).

As mentioned above, the Caffarelli contraction theorem has a lot of applications in the field of geometric and functional inequalities. We refer the interested reader to [19, 20, 9, 8] for some direct applications of entropic costs, Schrödinger bridges and entropic interpolations in this field.

The paper is organized as follows. Section 2 introduces entropic transport costs and the main results of the paper. Section 3 gives proofs of these results. Section 4 presents the alternative proof of Theorem 2.

## 2. Entropic transport costs and main results

2.1. **Entropic costs and their zero-noise limit.** Consider the classical Ornstein-Uhlenbeck process $(Z_t)_{t\geqslant 0}$ on $\mathbb{R}^d$, defined by the following stochastic differential equation:

$$dZ_t = \frac{-1}{2} Z_t\, dt + dW_t, \qquad t \geqslant 0,$$

where $(W_t)_{t\geqslant 0}$ is a standard $d$ dimensional Brownian motion and $Z_0 \sim \gamma_d$. As it is well known, the process $Z$ admits the following explicit representation

$$Z_t = Z_0 e^{-t/2} + e^{-t/2}\int_0^t e^{s/2}\, dW_s, \qquad t \geqslant 0.$$

The joint law of $(Z_0, Z_\varepsilon)$ will be denoted by $R^\varepsilon$. It is therefore given by

$$R^\varepsilon = \mathrm{Law}\left(X, Xe^{-\varepsilon/2} + \sqrt{1-e^{-\varepsilon}}Y\right),$$

with $X, Y$ two independent standard Gaussian random vectors on $\mathbb{R}^d$. In other words,

$$R^\varepsilon(dxdy) = \gamma_d(dx) r_x^\varepsilon(dy),$$

where $x \mapsto r_x^\varepsilon$ is the probability kernel defined by $r_x^\varepsilon = \mathcal{N}(xe^{-\varepsilon/2}, (1-e^{-\varepsilon})I_d)$.

Recall that the relative entropy of a probability measure $\alpha$ with respect to another probability measure $\beta$ on some measurable space $(\mathcal{X}, \mathcal{A})$ is defined by

$$H(\alpha|\beta) = \int \log\left(\frac{d\alpha}{d\beta}\right)\, d\alpha,$$

if $\alpha$ is absolutely continuous with respect to $\beta$. If this is not the case, one sets $H(\alpha|\beta) = +\infty$. The relative entropy is a non-negative quantity that vanishes only when the two probability measures are equal, this is why it is often called Kullback-Leibler distance (even though it is not a true distance).

**Definition 3** (Entropic transport cost)**.** *For all probability measures $\mu, \nu$ on $\mathbb{R}^d$, the entropic transport cost associated to $R^\varepsilon$ is defined by*

$$\mathcal{T}_H^\varepsilon(\mu,\nu) = \inf_{\pi \in C(\mu,\nu)} H(\pi|R^\varepsilon).$$

As shown by Mikami, Léonard and others [36, 33] the zero noise limit of $\varepsilon\mathcal{T}_H^\varepsilon$ is $\frac{1}{2}W_2^2$. At a heuristic level, this phenomenon can be easily understood from the following identities:

$$\begin{aligned}
\varepsilon H(\pi|R^\varepsilon) &= \varepsilon \int \log\left(\frac{d\pi}{dx}\right)\, d\pi - \varepsilon\int \log\left(\frac{dR^\varepsilon}{dx}\right)\, d\pi \\
&= \varepsilon \int \log\left(\frac{d\pi}{dx}\right)\, d\pi + \frac{\varepsilon}{2(1-e^{-\varepsilon})}\int |y - e^{-\varepsilon/2}x|^2\, \pi(dxdy) + \frac{\varepsilon}{2}\int |x|^2\, \mu(dx) + c(\varepsilon),
\end{aligned}$$

where $c(\varepsilon) \to 0$ (and is independent of $\mu, \nu, \pi$). So for small $\varepsilon$, minimizing $\pi \mapsto H(\pi|R^\varepsilon)$ amounts to minimizing $\pi \mapsto \frac{1}{2}\int |x-y|^2\, \pi(dxdy)$.

In the sequel we will use the following result, which can be easily deduced from a general convergence theorem due to Carlier, Duval, Peyré and Schmitzer [5, Theorem 2.7]. We will say that a probability measure $\eta$ is of finite (Shannon) entropy if it is absolutely continuous with respect to the Lebesgue measure and if $\int \log\left(\frac{d\eta}{dx}\right)\, d\eta$ is finite. Note that, if $\eta \in \mathcal{P}_2(\mathbb{R}^d)$, then it is of finite entropy if and only if $H(\eta|\gamma_d) < +\infty$.

**Theorem 4** (Carlier et al. [5])**.** *Suppose that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ are of finite entropy. Then, it holds*

$$\varepsilon\mathcal{T}_H^\varepsilon(\mu,\nu) \to \frac{1}{2}W_2^2(\mu,\nu) \qquad \textit{as } \varepsilon \to 0.$$

We state now a technical lemma that will be needed to apply Theorem 4 in our framework:

**Lemma 5.** *If $\mu$ and $\nu$ satisfy the assumptions of Theorem 1, then they are of finite entropy.*

*Proof of Lemma 5.* Let us first show that the probability measure $\mu$ has finite entropy. Since $\mu$ has a finite second moment, it is enough to show that $H(\mu|Leb) < +\infty$, which amounts to show that $V$ is $\mu$ integrable. Since $V$ is bounded from below by some affine function, it is clear that $[V]_-$ is $\mu$ integrable. Moreover, since the convex function $V$ is such that $\int e^{V(x)}\,\gamma_d(dx) = 1$, this implies according to [22, Lemma 2.1] that $[V]_+(x) \leqslant \frac{|x|^2}{2}$, for all $x \in \mathbb{R}^d$, and so $[V]_+$ is also $\mu$ integrable. Similarly, to see that $H(\nu|\gamma_d) < +\infty$, it is enough to show that $W$ is $\nu$ integrable. On the one hand, $\int [W]_+\,d\nu = \int \left[\log(e^{-W})e^{-W}\right]_-\,d\gamma_d \leqslant \frac{1}{e}$. On the other hand, $[W]_-$ is $\nu$ integrable since $W$ is bounded from below by some affine function. □

2.2. **Entropic cost in the framework of Caffarelli theorem.** As explained above, the key step in our approach consists in showing that on the set of probability measures dominated by $\nu$ in the convex order, the closest point to $\mu$ for the entropic cost distance is $\nu$ itself (when $\nu$ satisfies the assumptions of Theorem 1).

**Theorem 6.** *Let $\mu$ and $\nu$ satisfy the assumptions of Theorem 1. Additionally assume that $V$ is bounded from below. If $\eta$ is of finite entropy and such that $\eta \leqslant_c \nu$, then for all $\varepsilon > 0$*

$$\mathcal{T}_H^{\varepsilon}(\mu,\eta) \geqslant \mathcal{T}_H^{\varepsilon}(\mu,\nu).$$

Let us admit Theorem 6 for a moment and complete the proof of Theorem 1, which will also use the following Lemma:

**Lemma 7.** *Let $\nu(dx) = e^{-W(x)}\,\gamma_d(dx)$ with $W : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ convex and $\eta \leqslant_c \nu$. Assume furthermore that $\nu$ has compact support. Define, for all $\theta \in (0,\pi/2)$,*

$$\nu_\theta = \mathrm{Law}(X\cos\theta + Z\sin\theta) \qquad \text{and} \qquad \eta_\theta = \mathrm{Law}(Y\cos\theta + Z\sin\theta),$$

*where $X \sim \nu$, $Y \sim \eta$ and $Z$ independent of $X, Y$ and such that the law $\alpha$ of $Z$ is given by $\alpha(dz) = \frac{1}{C}\mathbf{1}_B\,\gamma_d(dz)$ where $B$ is the Euclidean unit ball and $C$ a normalizing constant. Then, for all $\theta \in (0,\pi/2)$,*

1. *the probability $\nu_\theta$ has a density of the form $e^{-W_\theta}$ with respect to $\gamma_d$, with $W_\theta : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ convex,*
2. *the probability measures $\nu_\theta$ and $\eta_\theta$ are compactly supported,*
3. *it holds $\eta_\theta \leqslant \nu_\theta$,*
4. *the probability $\eta_\theta$ has finite entropy.*

*Proof.* The density $f_\theta$ of $\nu_\theta$ is given by

$$f_\theta(x) = \frac{1}{C'}\int_B e^{-W\left(\frac{x-\sin\theta y}{\cos\theta}\right)} e^{-\frac{|x-\sin\theta y|^2}{2\cos^2\theta}} e^{-\frac{|y|^2}{2}}\,dy,$$

where $C'$ is a normalizing constant. A simple calculation shows that

$$e^{\frac{|x|^2}{2}} f_\theta(x) = \frac{1}{C'}\int_B e^{-W\left(\frac{x-\sin\theta y}{\cos\theta}\right)} e^{-\frac{|\sin\theta x - y|^2}{2\cos^2\theta}}\,dy$$

and, according to Prekopa Theorem [39], the right hand side is log-concave, which completes the proof of Item (1). The proofs of Items (2) and (3) are straightforward and left to the reader. The density of $\eta_\theta$ is $g_\theta(x) = \frac{1}{C}\int e^{-\frac{|\cos\theta y - x|^2}{2\sin^2\theta}}\mathbf{1}_B\left(\frac{\cos\theta y - x}{\sin\theta}\right)\,\eta(dy)$ and so $g_\theta \leqslant \frac{1}{C}$. On the other hand, $g_\theta \log g_\theta \geqslant -1/e$. Since the support of $\eta_\theta$ is compact, one sees that $g_\theta \log g_\theta$ is integrable and so $\eta_\theta$ has finite entropy. □

We shall now show give the proof of Theorem 1:

*Proof of Theorem 1.* Let us temporarily assume that $V$ is bounded from below. According to Lemma 5, $\mu$ and $\nu$ have finite entropy. So using Theorem 4, one concludes by letting $\varepsilon \to 0$ that for all compactly supported probability measures $\nu$ of the form $\nu(dx) = e^{-W(x)}\gamma_d(dx)$, with $W : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ convex, it holds

$$W_2(\mu, \nu) \leqslant W_2(\mu, \eta)$$

for all $\eta$ of finite entropy and such that $\eta \leqslant_c \nu$. Now, fix some compactly supported $\nu_0$ of the form $\nu_0(dx) = e^{-W_0(x)}\gamma_d(dx)$, with $W_0 : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ convex and let us show that the inequality (1) holds for any $\eta \leqslant_c \nu_0$. Take $\eta \leqslant_c \nu_0$ and define, for all $\theta \in (0, \pi/2)$,

$$\nu_\theta = \mathrm{Law}(X\cos\theta + Z\sin\theta) \qquad \text{and} \qquad \eta_\theta = \mathrm{Law}(Y\cos\theta + Z\sin\theta),$$

where $X \sim \nu_0$, $Y \sim \eta$ and $Z$ is independent of $X$ and $Y$ and has density $\frac{1}{C}\mathbf{1}_B(x)e^{-\frac{|x|^2}{2}}$, where $B$ is the Euclidean unit ball. According to Lemma 7, $\nu_\theta$ is compactly supported and of the form $e^{-W_\theta}\gamma_d$, with $W_\theta$ convex, $\eta_\theta$ is of finite entropy and $\eta_\theta \leqslant_c \nu_\theta$. Therefore, it holds $W_2(\mu, \nu_\theta) \leqslant W_2(\mu, \eta_\theta)$. Letting $\theta \to 0$ gives $W_2(\mu, \nu_0) \leqslant W_2(\mu, \eta)$ for all $\eta \leqslant_c \nu_0$, which, according to Theorem 2, completes the proof when $V$ is bounded from below.

Finally, let us remove the assumption that $V$ is bounded from below. Since $V$ is convex, it is bounded from below by some affine function. Thus there exists some $a \in \mathbb{R}^d$ such that $x \mapsto V(x) + a\cdot x$ is bounded from below. Consider the probability measure $\tilde{\mu}$ defined as the push forward of $\mu$ under the translation $x \mapsto x + a$. An easy calculation shows that the density of $\tilde{\mu}$ with respect to $\gamma_d$ is $Ce^{V(x-a)+a\cdot(x-a)}$, with $C$ a normalizing constant, and so $\tilde{\mu}$ satisfies our assumptions. Therefore, there exists a continuously differentiable convex function $\tilde{\phi} : \mathbb{R}^d \to \mathbb{R}$ such that $\nabla\tilde{\phi}$ is 1-Lipschitz and $\nu = \nabla\tilde{\phi}_\#\tilde{\mu}$. Setting $\phi(x) = \tilde{\phi}(x + a)$, $x \in \mathbb{R}^d$, one gets $\nu = \nabla\phi_\#\mu$ which completes the proof. □

Before proving Theorem 6, let us informally explain why one can guess the statement is easier to prove at the level of entropic cost than directly for the Wasserstein distance. If we consider the plain relative entropy, we have the variational formula

$$H(\rho|\mu) = \sup \int f\, d\rho - \log \int e^f\, d\mu,$$

where the supremum runs over the set of functions $f$ such that $\int e^f\, d\mu < +\infty$. Moreover, equality is achieved for $f = \log(d\rho/d\mu)$. Hence, taking $f = -(V + W)$, gives

$$H(\rho|\mu) \geqslant \int -(V + W)\, d\rho \geqslant \int -(V + W)\, d\nu = H(\nu|\mu)$$

as soon as $\rho \leqslant_c \nu$. So this trivial bound hints at the fact that comparison is easier for entropies when we have a log concavity condition on the relative density.

To prove Theorem 6, we need to know more about the optimal coupling $\pi$ for $\mathcal{T}_H^\varepsilon(\mu, \nu)$. The following is classical in entropic transport literature and goes back to the study of the so called Schrödinger bridges [41].

**Proposition 8.** *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ be such that $H(\mu|\gamma_d) < +\infty$ and $H(\nu|\gamma_d) < +\infty$.*

(1) *There exists a unique coupling $\pi^\varepsilon \in C(\mu, \nu)$ such that*

$$\mathcal{T}_H^\varepsilon(\mu, \nu) = H(\pi^\varepsilon|R^\varepsilon) < +\infty.$$

(2) *There exist two measurable functions $f^\varepsilon, g^\varepsilon : \mathbb{R}^d \to \mathbb{R}^+$ such that $\log f^\varepsilon \in L^1(\mu)$, $\log g^\varepsilon \in L^1(\nu)$ and*

$$\pi^\varepsilon(dxdy) = f^\varepsilon(x)g^\varepsilon(y)R^\varepsilon(dxdy).$$

*Sketch of proof.* (1) We equip the set $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with the usual topology of narrow convergence. For this topology, the function $\pi \mapsto H(\pi|R^\varepsilon)$ is lower-semicontinuous and the set $C(\mu,\nu)$ is compact. Therefore, the function $H(\,\cdot\,|R^\varepsilon)$ attains its minimum at some point $\pi^\varepsilon$ of $C(\mu,\nu)$. It is easily checked that the coupling $\pi_0 = \mu \otimes \nu$ is such that $H(\pi_0|R^\varepsilon) < +\infty$, so $H(\pi^\varepsilon|R^\varepsilon) < +\infty$. Uniqueness comes from the strict convexity of $H(\,\cdot\,|R^\varepsilon)$. For the proof of (2) we refer to [12, Corollary 3.2]. In the special case where $\mu$ and $\nu$ satisfy our log-convexity/concavity assumptions we will give a self-contained proof in Section 3. □

In the setting of Theorem 1, it turns out that much more can be said about the functions $f$ and $g$. This is explained in the following result, which seems of independent interest.

**Theorem 9.** *With the same notation as in Proposition 8, let $\mu$ be a probability measure of the form $\mu(dx) = e^{V(x)}\gamma_d(dx)$ with a finite second moment and $\nu$ be a compactly supported probability measure on $\mathbb{R}^d$ of the form $\nu(dx) = e^{-W(x)}\gamma_d(dx)$, with $V, W$ convex and $V$ bounded from below. There exist a* log-convex *function $f^\varepsilon : \mathbb{R}^d \to [1,+\infty)$ and a* log-concave *function $g^\varepsilon : \mathbb{R}^d \to [0,+\infty)$ such that the unique optimal coupling $\pi^\varepsilon \in \Pi(\mu,\nu)$ is of the form $\pi^\varepsilon(dxdy) = f^\varepsilon(x)g^\varepsilon(y)\,R^\varepsilon(dxdy)$. Moreover, the function $\log f^\varepsilon$ is integrable with respect to $\mu$ and the function $\log g^\varepsilon$ is integrable with respect to $\nu$ and it holds*

$$\mathcal{T}_H^\varepsilon(\mu,\nu) = H(\pi^\varepsilon|R^\varepsilon) = \int \log f^\varepsilon\,d\mu + \int \log g^\varepsilon\,d\nu.$$

We now give a brief heuristic explanation as to why one can expect this statement to imply the Caffarelli contraction theorem. Informally, from the convergence of the entropic cost to the Wasserstein distance, we expect from the dual formulation that $\varepsilon \log f$ converges to $|x|^2/2 - \varphi$ (up to some additive constant), where $\varphi$ is a potential giving rise to the optimal transport map $T = \nabla\varphi$. However convexity is preserved by pointwise convergence, so we expect $|x|^2/2 - \varphi$ to also be convex. But this is equivalent to $\nabla\varphi$ being 1-Lipschitz, since the eigenvalues of the Hessian of $\varphi$ must then be bounded by 1. Theorem 2 will allow us to avoid having to prove convergence of $\varepsilon \log f$ to a Kantorovich potential.

Section 3 is essentially devoted to the proof of Theorem 9. With Theorem 9 in hand, the proof of Theorem 6 becomes almost straightforward:

*Proof of Theorem 6.* Recall the following duality inequality for the relative entropy : if $\alpha, \beta$ are two probability measures on a measurable space $(\mathcal{X}, \mathcal{A})$ such that $H(\alpha|\beta) < +\infty$, then for any measurable function $h : \mathcal{X} \to \mathbb{R}$ such that $\int e^h\,d\beta < +\infty$, it holds $\int [h]_+\,d\alpha < +\infty$ and

$$H(\alpha|\beta) \geqslant \int h\,d\alpha - \log\left(\int e^h\,d\beta\right) \tag{3}$$

Let $\pi \in C(\mu,\eta)$ be a coupling between $\mu$ and some probability $\eta \leqslant_c \nu$ such that $H(\pi|R^\varepsilon) < +\infty$ ; applying the inequality above to $\alpha = \pi$, $\beta = R^\varepsilon$ and $h(x,y) = \log\left(f^\varepsilon(x)g^\varepsilon(y)\right)$, $x, y \in \mathbb{R}^d$ gives

$$\begin{aligned}
H(\pi|R^\varepsilon) &\geqslant \int \log f^\varepsilon(x) + \log g^\varepsilon(y)\,\pi(dxdy)\\
&= \int \log f^\varepsilon(x)\,\mu(dx) + \int \log g^\varepsilon(y)\,\eta(dy)\\
&\geqslant \int \log f^\varepsilon(x)\,\mu(dx) + \int \log g^\varepsilon(y)\,\nu(dy)\\
&= H(\pi^\varepsilon|R^\varepsilon) = \mathcal{T}_H^\varepsilon(\mu,\nu),
\end{aligned}$$

where the second inequality comes from the fact that $\log g^\varepsilon$ is a *concave* function and $\eta \leqslant_c \nu$. Optimizing over $\pi$, gives the inequality $\mathcal{T}_H^\varepsilon(\mu,\eta) \geqslant \mathcal{T}_H^\varepsilon(\mu,\nu)$ and completes the proof. □

To conclude this section, we mention some perspectives. The most natural question is whether this scheme of proof can be adapted to establish a version of Caffarelli's theorem in other settings than $\mathbb{R}^d$, such as on manifolds or in free probability [24]. See [38] for some motivations in analysis and geometry. See [21] for a study of Schrödinger's problem in a wider geometric setting. Another question is about integrated or non-local quantitative regularity estimates, such as those in [31, 32]. The role of 1-Lipschitz bounds in Theorem 2 is very specific, we do not know if there is an analogue of that equivalence adapted to other types of regularity bounds. However, it could be possible to prove stable a priori bounds on $\varepsilon \log f^\varepsilon$ and pass to the limit. Of particular interest is whether we can establish integrated gradient bounds for non-uniformly convex potentials, since such estimates can still be used to establish Poincaré inequalities [37, 30]. Finally, [14] proves a rigidity/stability result for the Caffarelli contraction theorem, and it would be interesting to find a way to improve the quantitative bounds.

## 3. Proof of the main results

This section contains the material needed to prove Theorem 6. The ideas developed here are adapted from a paper by Fortet [16]. We warmly thank Christian Léonard for mentioning us this paper and explaining to us the ingredients of Fortet's proof. Fortet's work was recently revisited in [15, 35].

We will denote by $P^\varepsilon$ the Ornstein-Uhlenbeck semi-group at time $\varepsilon$ defined for all non-negative measurable function $\psi$ by

$$P^\varepsilon \psi(x) = \mathbb{E}[\psi(Z_\varepsilon)|Z_0 = x] = \frac{1}{(2\pi)^{d/2}} \frac{1}{(1-e^{-\varepsilon})^{d/2}} \int_{\mathbb{R}^d} \psi(y + e^{-\varepsilon/2}x) e^{-\frac{|y|^2}{2(1-e^{-\varepsilon})}}\, dy, \qquad x \in \mathbb{R}^d.$$

Suppose that $f^\varepsilon, g^\varepsilon$ are measurable non-negative functions such that $\pi^\varepsilon(dxdy) = f^\varepsilon(x) g^\varepsilon(y) R^\varepsilon(dxdy)$ belongs to $C(\mu,\nu)$. Then, writing the marginals condition, one sees that $f^\varepsilon$ and $g^\varepsilon$ are related to each other by the identities: for all $x, y \in \mathbb{R}^d$

$$f^\varepsilon(x) P^\varepsilon g^\varepsilon(x) = e^{V(x)} \qquad \text{and} \qquad g^\varepsilon(y) P^\varepsilon f^\varepsilon(y) = e^{-W(y)}. \tag{4}$$

These relations suggest to introduce the functional $\Phi^\varepsilon$ defined as follows: for all measurable function $h : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$,

$$\Phi^\varepsilon(h) = V - \log\left( P^\varepsilon \left( e^{-W} \frac{1}{P^\varepsilon(e^h)} \right) \right).$$

With this notation, a couple $(f^\varepsilon, g^\varepsilon)$ satisfies (4) if and only if $g^\varepsilon = e^{-W} \frac{1}{P^\varepsilon(f^\varepsilon)}$ and $f^\varepsilon = e^{h^\varepsilon}$ with $h^\varepsilon$ such that

$$h^\varepsilon = \Phi^\varepsilon(h^\varepsilon).$$

This fixed point equation suggests that the unknown function $h^\varepsilon$ could be obtained as the limit when $n \to +\infty$ of a sequence $(h_n)_{n \geqslant 0}$ satisfying the recursive scheme

$$h_{n+1} = \Phi^\varepsilon(h_n), \qquad n \geqslant 0 \tag{5}$$

and initialized with some function $h_0$. This fixed point scheme is actually at the core of the use of Sinkhorn's algorithm to numerically approximate optimal transport via entropic regularization [13, 2].

**Remark 10.** *Note that we allow the argument $h$ to which we apply $\Phi^\varepsilon$ to take the value $+\infty$. The heat flow is nonetheless well defined, since we only apply it to $e^h$, which is nonnegative. Of course, as soon as $h$ takes infinite value on a set of positive mass, $\Phi^\varepsilon(h)$ takes value $+\infty$ everywhere, and the function that takes value $+\infty$ everywhere is a trivial fixed point. Part of the difficulty in the proof of Theorem 15 below will be to show that the fixed point we obtain is not the trivial one, but is finite everywhere.*

The convexity of $h^\varepsilon$ can then be established if we can initiate this fixed point scheme (5) with some convex initial data $h_0$, thanks to the following key result:

**Lemma 11.** *If $h : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex, then $\Phi^\varepsilon(h)$ is also convex.*

*Proof.* This property is inherited from the following classical properties of $P^\varepsilon$ :

- If $f$ is log-convex, then $P^\varepsilon(f)$ is log-convex. This simply follows from Hölder inequality.
- If $g$ is log-concave, then $P^\varepsilon(g)$ is log-concave. This follows from the fact that the set of log-concave functions is stable under convolution which is a well known consequence of Prekopa Theorem [39].

□

The line of reasoning sketched above is essentially the one adopted in the proof of Theorem 9, except that the recurrence scheme (5) needs to be properly modified in order to force its convergence (this modification is the same as the one proposed by Fortet in [16]).

**Remark 12.** *In the compact setting, the map $\Phi^\varepsilon$ is actually a contraction with respect to a well-chosen metric, see for example* [18, Lemma 1] *or* [6] *(following the earlier ideas of* [17] *in the discrete setting). This would ensure that the fixed point must belong to any stable subspace. Here, we work in a noncompact setting ($\mu$ has non-compact support) and it seems the map is globally only 1-Lipschitz at that level of generality. One could however expect that it remains a contraction on a suitable stable subspace of convex functions.*

**Remark 13.** *A natural question is whether our scheme of proof can be used directly at the level of the Kantorovich dual formulation of optimal transport, rather than on the regularized version. The answer seems to be no, as in the limit while the minimizers in the dual formulation of entropic transport, suitably rescaled, converge to the Kantorovich potentials, the fixed point problem becomes degenerate in the limit, and only selects so-called c-convex functions (with the cost here being the quadratic distance), so we lose uniqueness. Indeed, there is no known fixed point scheme similar to* (5) *for Kantorovich potentials, which is why Sinkhorn's algorithm is only used to numerically approximate the regularized problem* [13].

Before moving on to the proof, let us present two other essential properties of $\Phi^\varepsilon$.

**Lemma 14.**

(1) *The map $\Phi^\varepsilon$ is monotone: $h \leqslant k \Rightarrow \Phi^\varepsilon(h) \leqslant \Phi^\varepsilon(k)$.*
(2) *For any measurable $h : \mathbb{R}^d \to \mathbb{R}$, it holds*

$$\int \exp\left(h(x) - \Phi^\varepsilon(h)(x)\right) d\mu \leqslant 1,$$

*with equality if $h$ is bounded from above.*

*Proof.* The first point is straightforward. Let us prove the second point. Since the operator $P^\varepsilon$ is symmetric in $L^2(\gamma_d)$, for any function $h : \mathbb{R}^d \to \mathbb{R}$ it holds

$$\int e^{h\wedge a - \Phi^\varepsilon(h)}\, d\mu = \int e^{h\wedge a} P^\varepsilon\left(e^{-W}\frac{1}{P^\varepsilon(e^h)}\right) d\gamma_d = \int P^\varepsilon\left(e^{h\wedge a}\right) e^{-W}\frac{1}{P^\varepsilon(e^h)}\, d\gamma_d.$$

Letting $a \to +\infty$, one gets by monotone convergence $\int e^{h-\Phi^\varepsilon(h)}\, d\mu = \int_{\{P^\varepsilon(e^h)<+\infty\}} e^{-W}\, d\gamma_d$ which gives the claim. □

The existence of a coupling of the desired form can be established under more general conditions on $\mu$ and $\nu$:

**Theorem 15.** *Let $\mu$ be a probability measure of the form $\mu(dx) = e^{V(x)}\gamma_d(dx)$ with $V : \mathbb{R}^d \to \mathbb{R}$ convex and bounded from below, and let $\nu$ be a probability measure on $\mathbb{R}^d$ of the form $\nu(dx) = e^{-W(x)}\gamma_d(dx)$, with $W : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ a convex function such that $\{W < -m\}$ is bounded for $m = \inf_{\mathbb{R}^d} V \leqslant 0$. There exist a* log-convex *function $f^\varepsilon : \mathbb{R}^d \to [1, +\infty)$ and a* log-concave *function $g^\varepsilon : \mathbb{R}^d \to [0, +\infty)$ such that the measure $\pi^\varepsilon$ defined by $\pi^\varepsilon(dxdy) = f^\varepsilon(x)g^\varepsilon(y)\, R^\varepsilon(dxdy)$ belongs to $C(\mu, \nu)$.*

*Proof of Theorem 15.* Let us show that there exists a convex function $\bar{h} : \mathbb{R}^d \to \mathbb{R}^+$ such that $\Phi^\varepsilon(\bar{h}) = \bar{h}$. Then, defining $f^\varepsilon = e^{\bar{h}}$ and $g^\varepsilon = e^{-V}/P^\varepsilon(f^\varepsilon)$, we see that $f^\varepsilon$ is log-convex, $g^\varepsilon$ is log-concave (we use again the fact that $P^\varepsilon$ preserves log-convexity) and satisfy (4).

Let us define by induction the sequence $(h_n)_{n\geqslant 0}$ as follows: $h_0 = 0$ and for all $n \geqslant 0$

$$h_{n+1} = \left[\Phi^\varepsilon(h_n)\right]_+ \wedge n. \tag{6}$$

By construction, note that $h_n$ takes values in $[0, n-1]$. Let us show by induction that the sequence $(h_n)_{n\geqslant 0}$ is non-decreasing. First observe that $h_1 = 0 = h_0$ and so in particular $h_0 \leqslant h_1$. According to Item (1) of Lemma 14, the operator $\Phi^\varepsilon$ is non-decreasing. Therefore, if $h_{n+1} \geqslant h_n$ for some $n \geqslant 0$, then

$$h_{n+2} = \left[\Phi^\varepsilon(h_{n+1})\right]_+ \wedge (n+1) \geqslant \left[\Phi^\varepsilon(h_n)\right]_+ \wedge (n+1) \geqslant \left[\Phi^\varepsilon(h_n)\right]_+ \wedge n = h_{n+1}.$$

Let us denote by $h_\infty$ the pointwise limit of $h_n$ as $n \to \infty$. The function $h_\infty$ takes values in $\mathbb{R}^+ \cup \{+\infty\}$. Let us show that $h_\infty$ solves the following fixed point equation

$$h_\infty = \left[\Phi^\varepsilon(h_\infty)\right]_+. \tag{7}$$

Indeed, by monotone convergence, $P^\varepsilon(e^{h_n}) \to P^\varepsilon(e^{h_\infty})$. Then, by dominated convergence,

$$P^\varepsilon\left(e^{-W}\frac{1}{P^\varepsilon(e^{h_n})}\right) \to P^\varepsilon\left(e^{-W}\frac{1}{P^\varepsilon(e^{h_\infty})}\right)$$

which implies that $h_n \to \left[\Phi^\varepsilon(h_\infty)\right]_+$ and gives (7). Note that the use of the dominated convergence theorem is justified by the fact that $P^\varepsilon(e^{h_n}) \geqslant 1$ since $h_n \geqslant 0$ and the fact that $e^{-W}$ is easily seen to bounded from above.

We shall now prove that $\Phi^\varepsilon(h_\infty)(x) < +\infty$ for all $x \in \mathbb{R}^d$. We shall proceed by contradiction, and show that if $h_\infty$ took an infinite value somewhere, then our fixed point procedure would then terminate in a finite number of steps. This would imply by contradiction that $h_\infty$ cannot take an infinite value, since after a finite number of steps the function $h_n$ is necessarily bounded from above, by definition. So let us assume, that there exists some $x_o \in \mathbb{R}^d$ such that $\Phi^\varepsilon(h_\infty)(x_o) = +\infty$. This easily implies that $P^\varepsilon(e^{h_\infty}) = +\infty$ almost everywhere, which in turn implies that $\Phi^\varepsilon(h_\infty) \equiv \infty$. Since $h_\infty \geqslant \Phi^\varepsilon(h_\infty)$, one concludes also that $h_\infty \equiv +\infty$. Now let us show that there exists $n_0$ such that for all $n \geqslant n_0$

$$\inf_{x\in\mathbb{R}^d} \Phi^\varepsilon(h_n)(x) \geqslant 0. \tag{8}$$

For any $x \in \mathbb{R}^d$, it holds (denoting by $C = (2\pi)^{d/2}(1-e^{-\varepsilon})^{d/2}$ and by $m = \inf_{\mathbb{R}^d} V$)

$$\begin{aligned}
P^\varepsilon\left(e^{-W}\frac{1}{P^\varepsilon(e^{h_n})}\right)(x) &= \frac{1}{C}\int_{\mathbb{R}^d} e^{-W(y)}\frac{1}{P^\varepsilon(e^{h_n})}(y)e^{-\frac{|y-e^{-\varepsilon/2}x|^2}{2(1-e^{-\varepsilon})}}\,dy \\
&= \frac{1}{C}\int_{\{W\leqslant -m\}} e^{-W(y)}\frac{1}{P^\varepsilon(e^{h_n})}(y)e^{-\frac{|y-e^{-\varepsilon/2}x|^2}{2(1-e^{-\varepsilon})}}\,dy \\
&+ \frac{1}{C}\int_{\{W>-m\}} e^{-W(y)}\frac{1}{P^\varepsilon(e^{h_n})}(y)e^{-\frac{|y-e^{-\varepsilon/2}x|^2}{2(1-e^{-\varepsilon})}}\,dy \\
&\leqslant \frac{1}{C}\int_{\{W\leqslant -m\}} e^{-\frac{|y-e^{-\varepsilon/2}x|^2}{2(1-e^{-\varepsilon})}}\,dy \max_{z\in\{W\leqslant -m\}} e^{-W(z)}\frac{1}{P^\varepsilon(e^{h_n})}(z). \\
&+ \frac{e^m}{C}\int_{\{W>-m\}} e^{-\frac{|y-e^{-\varepsilon/2}x|^2}{2(1-e^{-\varepsilon})}}\,dy,
\end{aligned}$$

where we used the fact that $P^\varepsilon(e^{h_n}) \geqslant 1$, since $h_n \geqslant 0$. The sequence of functions $\frac{1}{P^\varepsilon(e^{h_n})}$ is a non-increasing sequence of continuous functions converging to 0. Therefore, according to Dini's Theorem, the convergence is uniform on the compact set $K = \overline{\{W<0\}}$. Since $W$ is convex, it is bounded from below on $K$. Therefore, there exists $n_0$ such that $\max_{z\in K} e^{-W(z)}\frac{1}{P^\varepsilon(e^{h_n})}(z) \leqslant e^m$ for all $n \geqslant n_0$. Plugging this inequality into the inequality above, one easily gets (8). Now, according to (8), there exists some $n_o$ such that $\Phi^\varepsilon(h_{n_0}) \geqslant 0$. Therefore $h_{n_o+1} = \Phi^\varepsilon(h_{n_o}) \wedge n_o \leqslant \Phi^\varepsilon(h_{n_o+1})$, since the sequence $(h_n)$ is increasing and the operator $\Phi^\varepsilon$ is monotone. Since $h_{n_0+1}$ is bounded, Item (2) of Lemma 14 yields $\int e^{h_{n_o+1}-\Phi^\varepsilon(h_{n_o+1})}d\mu = 1$, which implies that $h_{n_o+1} = \Phi^\varepsilon(h_{n_o+1})$. Therefore, $h_\infty = h_{n_o+1}$, which necessarily contradicts the fact that $h_\infty \equiv +\infty$, since by definition $h_{n_o+1}$ is bounded from above. So there is a contradiction, and therefore $h_\infty$ must be finite everywhere.

At this point, one could show that $h_\infty$ is actually a fixed point of $\Phi^\varepsilon$, and not just a solution to (7), but this is not going to be necessary for us.

Now let $k_0 = h_\infty^{**}$ be the convex regularization of $h_\infty$ (which is well defined since $h_\infty$ is bounded from below). By definition $k_0 \leqslant h_\infty$ and since $h_\infty \geqslant 0$, it holds $k_0 \geqslant 0$. Define by induction $(k_n)_{n\geqslant 1}$ by $k_{n+1} = \max(\Phi^\varepsilon(k_n); k_0)$. Since according to Lemma 11 $\Phi^\varepsilon$ preserves convexity and $k_0$ is convex, $k_n$ is convex for all $n$. The sequence $k_n$ is non-decreasing and satisfies $k_n \leqslant h_\infty$ for all $n$. Therefore, $k_n$ converges pointwise to some $k_\infty$, which is also convex and finite valued. Reasoning as above one sees that $k_\infty = \max(\Phi^\varepsilon(k_\infty); k_0)$ and so in particular $k_\infty \geqslant \Phi^\varepsilon(k_\infty)$. According to Item (2) of Lemma 14, it holds $\int e^{k_\infty - \Phi^\varepsilon(k_\infty)}\,d\mu \leqslant 1$. Since $k_\infty \geqslant \Phi^\varepsilon(k_\infty)$, the function $e^{k_\infty-\Phi^\varepsilon(k_\infty)}$ is bounded from below by 1. Therefore, $k_\infty = \Phi^\varepsilon(k_\infty)$ almost everywhere. But since both functions are convex, equality actually holds everywhere, and $k_\infty$ is a convex fixed point of $\Phi^\varepsilon$. Setting $\bar{h} = k_\infty$ completes the proof.

Due to the uniqueness of fixed points of the Schrödinger problem up to multiplicative constants [35, Proposition 5.1], solutions to the fixed point equation $h = \Phi^\varepsilon(h)$ are unique up to an additive constant. Therefore, a posteriori we actually have $k_\infty = h_\infty$, and thus the original function $h_\infty$ was actually itself convex, and the sequence $(k_n)$ is constant. □

*Proof of Theorem 9.* First, let us note that $H(\mu\otimes\nu|R^\varepsilon) < +\infty$. Since $\mu$ and $\nu$ have finite second moment, this is easily seen to be equivalent to $H(\mu|\gamma_d) < +\infty$ and $H(\nu|\gamma_d) < +\infty$, which is true according to Lemma 5. According to Theorem 15, there exists a coupling $\pi^\varepsilon(dxdy) = f^\varepsilon(x)g^\varepsilon(y)\,R^\varepsilon(dxdy) \in C(\mu,\nu)$ such that $f^\varepsilon$ is log-convex and $g^\varepsilon$ is log-concave. It remains to show that this coupling is optimal for $\mathcal{T}_H^\varepsilon(\mu,\nu)$. Since $P^\varepsilon f^\varepsilon(y)g^\varepsilon(y) = e^{-W(y)}$, $y\in\mathbb{R}^d$, one sees that $\log g^\varepsilon(y) = -W(y) - \log P^\varepsilon f^\varepsilon(y)$. The function $\log P^\varepsilon f^\varepsilon$ is bounded continuous on the support of $\nu$ and $W$ is integrable with respect to $\nu$. Therefore, $\log g^\varepsilon$ is integrable with respect to $\nu$.

On the other hand, since $\log f^\varepsilon \geqslant 0$, the integral $\int \log f^\varepsilon \, d\mu$ makes sense in $[0, +\infty]$. Let $\pi \in C(\mu, \nu)$ be a coupling such that $H(\pi|R^\varepsilon) < +\infty$ (this set is non empty, since it contains $\mu \otimes \nu$). Applying Inequality (3) with $\alpha = \pi$, $\beta = R^\varepsilon$ and $h(x, y) = \log f^\varepsilon(x) + \log g^\varepsilon(y)$, $x, y \in \mathbb{R}^d$, gives

$$H(\pi|R^\varepsilon) \geqslant \int \log f^\varepsilon(x) + \log g^\varepsilon(y) \, \pi(dxdy) = \int \log f^\varepsilon(x) \, \mu(dx) + \int \log g^\varepsilon(y) \, \nu(dy),$$

which shows that $\log f^\varepsilon$ is integrable with respect to $\mu$. A simple calculation shows that

$$H(\pi^\varepsilon|R^\varepsilon) = \int \log f^\varepsilon(x) \, \mu(dx) + \int \log g^\varepsilon(y) \, \nu(dy),$$

which shows its optimality. □

## 4. Variational proof of Theorem 2

The goal of this section is to give an alternative proof of Theorem 2. The original proof in [23] uses a weak version of optimal transport as an intermediary, but we give here a new proof relying only on the variational problem solved by the Brenier map. However, we need to restrict the proof to the case where $\nu$ is absolutely continuous with respect to Lebesgue (and $\mu$ too but this is hardly a restriction when we assume that a Brenier map exists), with its support being convex. Note that for the purpose of proving the Caffarelli contraction theorem, these assumptions are not a restriction.

Let us recall some classical facts about quadratic transport that will be used in the proof (see [42, 43] for proofs and more general statements). If $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the quadratic transport cost $\frac{1}{2}W_2^2$ admits the following dual formulation due to Kantorovich:

$$\frac{1}{2}W_2^2(\mu, \nu) = \sup_{\varphi, \psi} \left\{ \int \frac{|x|^2}{2} - \varphi(x) \, \mu(dx) + \int \frac{|y|^2}{2} - \psi(y) \, \nu(dy) \right\}, \tag{9}$$

where the supremum runs over couples of *convex conjugate* functions $(\varphi, \psi)$, that is to say that $\varphi, \psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ are convex, lower semi-continuous and such that $\psi = \varphi^*$ and $\varphi = \psi^*$, where we recall that the Legendre transform $h^*$ of a function $h : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$h^*(y) = \sup_{x \in \mathbb{R}^d} \{x \cdot y - h(x)\}, \qquad \forall y \in \mathbb{R}^d.$$

A classical result in optimal transport tells moreover that the supremum in (9) is always attained. If $(\varphi, \psi)$ is such a dual optimizer, we will say that $\varphi$ (resp. $\psi$) is a transport potential from $\mu$ to $\nu$ (resp. $\nu$ to $\mu$). This terminology is justified by the fact that if $\mu$ is absolutely continuous with respect to Lebesgue measure, then according to Brenier theorem, if $\varphi$ is a transport potential from $\mu$ to $\nu$, the map $\nabla\varphi$ (which is well defined $\mu$ almost surely) is (the $\mu$ almost surely unique) optimal transport map between $\mu$ and $\nu$, *i.e* $\nu = \nabla\varphi_{\#}\mu$ and $W_2^2(\mu, \nu) = \int |x - \nabla\varphi(x)|^2 \, \mu(dx)$. Finally, if $\nu$ is also absolutely continuous with respect to Lebesgue measure, then $\nabla\psi$ is the optimal transport map between $\nu$ and $\mu$ and it holds

$$\nabla\psi \circ \nabla\varphi(x) = x \qquad \text{and} \qquad \nabla\psi \circ \nabla\varphi(y) = y$$

for $\mu$ almost every $x$ and $\nu$ almost every $y$.

For reader's convenience, let us reformulate Theorem 2 in a slightly different form and with the extra assumptions mentioned above :

**Theorem 16.** *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ be absolutely continuous with respect to Lebesgue measure and suppose that $\nu$ has a convex support. The following are equivalent:*

*(i) There exists a transport potential $\varphi : \mathbb{R}^d \to \mathbb{R}$ from $\mu$ to $\nu$ which is continuously differentiable on $\mathbb{R}^d$ and such that $\nabla\phi$ is 1-Lipschitz on $\mathbb{R}^d$,*

(*ii*) *There exists a transport potential* $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *from* $\nu$ *to* $\mu$ *such that the function* $\mathbb{R}^d \to \mathbb{R} \cup \{+\infty\} : x \mapsto \psi(x) - \frac{|x|^2}{2}$ *is convex.*

(*iii*) *For all* $\eta \in \mathcal{P}_2(\mathbb{R}^d)$ *such that* $\eta \leqslant_c \nu$,

$$W_2(\nu, \mu) \leqslant W_2(\eta, \mu).$$

The equivalence between $(i)$ and $(ii)$ uses the following classical fact of convex analysis:

$$\varphi \ \mathcal{C}^1 \text{ and } \nabla\varphi \text{ 1-Lipschitz} \iff \varphi^* - \frac{|\cdot|^2}{2} \text{ is convex,}$$

see [27, Theorem E 4.2.1] or [23, Lemma 2.1]. It turns out that condition $(ii)$ will be easier to handle than condition $(i)$.

*Proof of Theorem 16,* $(ii) \Rightarrow (iii)$. Assume $\psi$ is a transport potential from $\nu$ to $\mu$ such that $\psi - \frac{|\cdot|^2}{2}$ is convex and let $\varphi = \psi^*$. For any $\eta \leqslant_c \nu$, it holds

$$\begin{aligned}
\frac{1}{2}W_2^2(\mu,\nu) &= \int \frac{|x|^2}{2} - \varphi(x)\,\mu(dx) + \int \frac{|y|^2}{2} - \psi(y)\,\nu(dy) \\
&\leqslant \int \frac{|x|^2}{2} - \varphi(x)\,\mu(dx) + \int \frac{|y|^2}{2} - \psi(y)\,\eta(dy) \\
&\leqslant \frac{1}{2}W_2^2(\mu,\eta),
\end{aligned}$$

where the first equality comes from the optimality of $(\varphi, \psi)$, the second inequality from the fact that $\eta \leqslant_c \nu$ and the concavity of $\frac{|\cdot|^2}{2} - \psi$ and the third inequality from Kantorovich duality. □

For any probability measure $\rho \in \mathcal{P}_2(\mathbb{R}^d)$, we will denote from now on as

$$C_\rho := \{\eta \in \mathcal{P}_2(\mathbb{R}^d) : \eta \leqslant_c \rho\}$$

the set of all probability measures which are dominated by $\rho$ in the convex order. Note that this set is geodesically convex in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$.

In order to prove the converse implication $(iii) \Rightarrow (ii)$, we will proceed by contradiction and show that if $(ii)$ is not true then one can construct a competitor $\eta \in C_\nu$ with a smaller Wasserstein distance to $\mu$. For that purpose, we will use the following simple localization lemma.

**Lemma 17.** *Let* $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ *and suppose that* $T : \mathbb{R}^d \to \mathbb{R}^d$ *is an optimal transport map from* $\mu$ *to* $\nu$. *Let* $A \subset \mathbb{R}^d$ *be a Borel set such that* $\nu(A) > 0$, *define* $\nu_A$ *as the renormalized restriction of* $\nu$ *to* $A$*:* $\nu_A := \frac{\nu \llcorner A}{\nu(A)}$ *and* $\mu_A$ *as the renormalized restriction of* $\mu$ *to* $T^{-1}(A)$*:* $\mu_A := \frac{\mu \llcorner T^{-1}(A)}{\nu(A)}$. *Suppose that there exists* $\eta_A \in C_{\nu_A}$ *such that* $W_2(\mu_A, \eta_A) < W_2(\mu_A, \nu_A)$. *Then the probability measure* $\eta$ *defined by* $\eta = \nu \llcorner A^c + \nu(A)\eta_A$ *is such that* $\eta \in C_\nu$ *and* $W_2(\mu, \eta) < W_2(\mu, \nu)$.

*Proof.* It is clear that $\eta \leqslant_c \nu$. Let us show that $\eta$ is closer to $\mu$ than $\nu$. For the sake of simplicity, we will assume that there exists an optimal transport map $S$ between $\mu_A$ and $\eta_A$. According to [43, Theorem 4.6], the map $T$ is still the optimal transport map from $\mu_A$ to $\nu_A$. The map $R$ defined by $R(x) = S(x)$ if $x \in T^{-1}(A)$ and $R(x) = T(x)$ if $x \in T^{-1}(A)^c$ is a transport map between $\mu$ and $\eta$ (not necessarily optimal) and it holds:

$$\begin{aligned} W_2^2(\mu,\eta) &\leqslant \int |x-R(x)|^2\,\mu(dx) \\ &= \mu(T^{-1}(A))\int |x-S(x)|^2\,\mu_A(dx) + \int_{T^{-1}(A)^c} |x-T(x)|^2\,\mu(dx) \\ &< \mu(T^{-1}(A))W_2^2(\mu_A,\nu_A) + \int_{T^{-1}(A)^c} |x-T(x)|^2\,\mu(dx) \\ &= \mu(T^{-1}(A))\int |x-T(x)|^2\,\mu_A(dx) + \int_{T^{-1}(A)^c} |x-T(x)|^2\,\mu(dx) \\ &= W_2^2(\mu,\nu). \end{aligned}$$

□

Before completing the proof of Theorem 16, let us state a lemma about strongly convex functions. Given a convex function $f$, we will denote $\mathrm{dom}(f) = \{x \in \mathbb{R}^d : f(x) < +\infty\}$ the domain of $f$.

**Lemma 18.** *Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a lower semi-continuous convex function such that* $\mathrm{dom}(f)$ *has a non-empty interior. The function $f$ is such that $f - \frac{|\cdot|^2}{2}$ is convex if and only if for all $x, y \in \mathrm{int}\,\mathrm{dom}(f)$ where $f$ is differentiable it holds*

$$(\nabla f(x) - \nabla f(y)) \cdot (x-y) \geqslant |x-y|^2. \tag{10}$$

Note that if $f$ is continuously differentiable, the conclusion of the lemma is straightforward.

*Proof of Lemma 18.* Let us denote by $D = \mathrm{dom}(f)$. According to [28, Theorem 6.1.2], $f - \frac{|\cdot|^2}{2}$ is convex if and only if for all $x, y \in D$, it holds

$$(x-y)\cdot(u-v) \geqslant |x-y|^2, \qquad \forall u \in \partial f(x),\ \forall v \in \partial f(y), \tag{11}$$

where, we recall that for any $x \in D$, we denote by $\partial f(x)$ the sub-gradient of the convex function $f$ at point $x$ which is defined as the set of all vectors $u \in \mathbb{R}^d$ such that $f(z) \geqslant f(x) + u \cdot (z-x)$, for all $z \in \mathbb{R}^d$. We recall also that when $f$ is differentiable at $x$ (which is true for Lebesgue almost every $x$ in the interior of $D$) then $\partial f(x) = \{\nabla f(x)\}$. Therefore, if $f$ satisfies (11) it satisfies (10).

Let us show the converse. According to [40, Theorem 25.6 ], for any $a \in D$, it holds

$$\partial f(a) = \mathrm{Conv}(S(a)) + R(a),$$

where $R(a)$ is the normal cone to $D$ at $a$, i.e

$$R(a) = \{h \in \mathbb{R}^d : h \cdot (z-a) \leqslant 0,\ \forall z \in D\}$$

and $S(a)$ is the set of vectors $u$ such that there exists a sequence of points $a_k \in \mathrm{int}\,D$ where $f$ is differentiable such that $a_k \to a$ and $\nabla f(a_k) \to u$ as $k \to \infty$. Let $x, y \in D$ and $u \in \partial f(x)$ and $v \in \partial f(y)$, with decomposition $u = u' + h$, $v = v' + k$ with $u' \in \mathrm{Conv}(S(x))$, $v' \in \mathrm{Conv}(S(y))$, $h \in R(x)$ and $k \in R(y)$. Since $(x-y)\cdot(h-k) \geqslant 0$, it is enough to show that $(x-y)\cdot(u'-v') \geqslant |x-y|^2$. By convexity, it is enough to prove that $(x-y)\cdot(u'-v') \geqslant |x-y|^2$ for all $u' \in S(x)$ and $v' \in S(y)$. If $a_k$ and $b_k$ are sequences converging to $x$ and $y$ respectively in such a way $\nabla f(a_k) \to u'$ and $\nabla f(b_k) \to v'$, then according to (10), it holds

$$(\nabla f(a_k) - \nabla f(b_k)) \cdot (a_k - b_k) \geqslant |a_k - b_k|^2$$

and letting $k \to \infty$ gives the desired inequality. □

We can now move on to the final part of the proof of Theorem 16.

*Proof of Theorem 16, $(iii) \Rightarrow (ii)$.* Assume that $(ii)$ does not hold, that is to say that whenever $\psi$ is a transport potential from $\nu$ to $\mu$ then the function $\psi - \frac{|\cdot|^2}{2}$ is not convex. First we want to make sure that the convexity problem occurs on the support of $\nu$, denoted by $\mathrm{Spt}(\nu)$ in what follows. Take $\tilde{\psi}$ an arbitrary transport potential from $\nu$ to $\mu$. Then, since $\mathrm{Spt}(\nu)$ is closed and convex, the function $\psi$ defined by $\psi(y) = \tilde{\psi}(y)$ if $y \in \mathrm{Spt}(\nu)$ and $\psi(y) = +\infty$ otherwise is still convex and lower semi-continuous. Defining $\varphi := \psi^*$, one easily sees that $\varphi \leqslant \tilde{\psi}^*$ and so $(\varphi, \psi)$ is a dual optimizer. In all what follows we will deal with this special potential $\psi$.

Since $(ii)$ does not hold, Lemma 18 applied with $f = \psi$ shows that there exist two points $x_0, y_0$ in the interior of $\mathrm{Spt}(\nu)$ where $f$ is differentiable and such that

$$(x_0 - y_0) \cdot (\nabla f(x_0) - \nabla f(y_0)) < |x_0 - y_0|^2.$$

By continuity of $r \mapsto \nu(B_r(z))$, for $z \in \mathrm{Spt}(\nu)$, we can find two functions $r_\varepsilon \to 0$ and $s_\varepsilon \to 0$ as $\varepsilon \to 0$ such that for all $\varepsilon > 0$, $\nu(B_{r_\varepsilon}(x_0)) = \nu(B_{s_\varepsilon}(y_0))$. We then define $A_\varepsilon$ to be the union of these two (disjoint) balls:

$$A_\varepsilon := B_{r_\varepsilon}(x_0) \cup B_{s_\varepsilon}(y_0).$$

Then $\nu_{A_\varepsilon}$ converges weakly to $\frac{1}{2}\delta_{x_0} + \frac{1}{2}\delta_{y_0}$, as $\varepsilon \to 0$. Let $b_\varepsilon = \int y \, \nu_{A_\varepsilon}(dy)$ be the barycenter of $A_\varepsilon$ with respect to $\nu_{A_\varepsilon}$. Then we have that

$$\lim_{\varepsilon \to 0} b_\varepsilon = b := \frac{x_0 + y_0}{2}.$$

In order to construct a competitor, let us collapse the mass of $\nu_{A_\varepsilon}$ towards $b_\varepsilon$ using the displacement interpolant between $\nu_{A_\varepsilon}$ and $\delta_{b_\varepsilon}$

$$\rho_t^\varepsilon := [(1-t)\mathrm{Id} + t b_\varepsilon]_\# \nu_{A_\varepsilon}.$$

Then it is easily seen that for all $t \in [0,1]$, $\rho_t^\varepsilon \leqslant_c \nu_{A_\varepsilon}$. Note that we must go towards $b_\varepsilon$, instead of $b$ directly, in order to stay in $C_{\nu_{A_\varepsilon}}$. As $\varepsilon \to 0$, this will not make a difference.

Let us now compute $W_2(\mu_{A_\varepsilon}, \rho_t^\varepsilon)$, where $\mu_{A_\varepsilon}$ is defined as in Lemma 17, and show that for $t$ and $\varepsilon$ small enough it is strictly less than $W_2(\mu_{A_\varepsilon}, \nu_{A_\varepsilon})$. Note that $\rho_t^\varepsilon$ is the image of $\mu_{A_\varepsilon}$ under the map $(1-t)\nabla\varphi + t b_\varepsilon$ which is clearly the gradient of the convex function $x \mapsto (1-t)\varphi(x) + t b_\varepsilon \cdot x$ and is thus optimal. Therefore,

$$\begin{aligned} W_2^2(\mu_{A_\varepsilon}, \rho_t^\varepsilon) &= \int |x - (1-t)\nabla\varphi(x) - t b_\varepsilon|^2 \, \mu_{A_\varepsilon}(dx) \\ &= W_2^2(\mu_{A_\varepsilon}, \nu_{A_\varepsilon}) + 2t \int (x - \nabla\varphi(x)) \cdot (\nabla\varphi(x) - b_\varepsilon) \, \mu_{A_\varepsilon}(dx) + t^2 \int |\nabla\varphi(x) - b_\varepsilon|^2 \, \mu_{A_\varepsilon}(dx). \end{aligned}$$

So, for any fixed $\varepsilon > 0$, the derivative at $t = 0$ is thus given by

$$\frac{d}{dt}_{|t=0} W_2^2(\mu_{A_\varepsilon}, \rho_t^\varepsilon) = 2 \int (x - \nabla\varphi(x)) \cdot (\nabla\varphi(x) - b_\varepsilon) \, \mu_{A_\varepsilon}(dx)$$

=This formula was obtained in [1, Proposition 7.3.6], and is also a particular case of [43, Theorem 23.9] which gives the time derivative of the Wasserstein distance along general curves of probability measures. Now, our goal is to show that the quantity calculated above is negative for all $\varepsilon$ small enough. Since $\mu_{A_\varepsilon} = \nabla\psi_\# \nu_{A_\varepsilon}$ and $\nabla\varphi \circ \nabla\psi(y) = y$ for $\nu_{A_\varepsilon}$ almost every $y$, one gets

$$\frac{d}{dt}_{|t=0} W_2^2(\mu_{A_\varepsilon}, \rho_t^\varepsilon) = 2 \int (\nabla\psi(y) - y) \cdot (y - b_\varepsilon) \, \nu_{A_\varepsilon}(dy).$$

To conclude, we use the following continuity property of the subgradient : if $\psi(x) < +\infty$, then for any $\delta > 0$, there exists $r > 0$ such that if $z \in B_r(x)$ then $\partial\psi(z) \subset \partial\psi(x) + B_\delta(0)$ (see [28, Theorem

6.2.4]). Since $\partial\psi(x_0) = \{\nabla\psi(x_0)\}$ and $\partial\psi(y_0) = \{\nabla\psi(y_0)\}$, it follows easily that

$$\begin{aligned}\lim_{\varepsilon\to 0} \frac{d}{dt}_{|t=0} W_2^2(\mu_{A_\varepsilon}, \rho_t^\varepsilon) &= (\nabla\psi(x_0) - x_0)\cdot(x_0 - b) + (\nabla\psi(y_0) - y_0)\cdot(y_0 - b)\\ &= (\nabla\psi(x_0) - \nabla\psi(y_0))\cdot(x_0 - y_0) - |x_0 - y_0|^2 < 0.\end{aligned}$$

Therefore, for $\varepsilon$ and $t$ small enough $W_2(\mu_{A_\varepsilon}, \rho_t^\varepsilon) < W_2(\mu_{A_\varepsilon}, \nu_{A_\varepsilon})$, which according to Lemma 17 shows that there exists $\eta \in C_\nu$ such that $W_2(\mu, \eta) < W_2(\mu, \nu)$ and completes the proof. $\square$

**Acknowledgments**: MF and NG were supported by ANR-11-LABX-0040-CIMI within the program ANR-11-IDEX- 0002-02. MF and MP were supported by Projects MESA (ANR-18-CE40-006) and EFI (ANR-17-CE40-0030) of the French National Research Agency (ANR). NG is supported by a grant of the Simone and Cino Del Duca Foundation. We thank Christian Léonard for pointing out to us Fortet's work and sharing a preliminary version of [35] explaining it, Gabriel Peyré for his lectures on numerical optimal transport in March 2019 in Toulouse, and Franck Barthe, Nicolas Juillet and Michel Ledoux for useful discussions. We also thank an anonymous referee for his valuable comments, and for a suggestion that simplified the proof of Theorem 15.

## References

1. Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. MR 2401600 (2009h:49002)
2. Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré, *Iterative Bregman projections for regularized transportation problems*, SIAM J. Sci. Comput. **37** (2015), no. 2, A1111–A1138. MR 3340204
3. Luis A. Caffarelli, *Monotonicity properties of optimal transportation and the FKG and related inequalities*, Comm. Math. Phys. **214** (2000), no. 3, 547–563. MR 1800860
4. ______, *Erratum: "Monotonicity of optimal transportation and the FKG and related inequalities" [Comm. Math. Phys.* **214** *(2000), no. 3, 547–563; MR1800860 (2002c:60029)]*, Comm. Math. Phys. **225** (2002), no. 2, 449–450. MR 1889232
5. Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer, *Convergence of entropic schemes for optimal transport and gradient flows*, SIAM J. Math. Anal. **49** (2017), no. 2, 1385–1418. MR 3635459
6. Y. Chen, T. Georgiou, and M. Pavon, *Entropic and displacement interpolation: a computational approach using the Hilbert metric*, SIAM J. Math. Analysis **76** (2016), no. 6, 2375–2396.
7. Maria Colombo, Alessio Figalli, and Yash Jhaveri, *Lipschitz changes of variables between perturbations of log-concave measures*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **17** (2017), no. 4, 1491–1519. MR 3752535
8. Giovanni Conforti, *A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost*, Probab. Theory Related Fields **174** (2019), no. 1-2, 1–47. MR 3947319
9. Giovanni Conforti and Luigia Ripani, *Around the entropic Talagrand inequality*, arXiv preprint arXiv:1809.02062 (2018).
10. Dario Cordero-Erausquin, *Some applications of mass transport to Gaussian-type inequalities*, Arch. Ration. Mech. Anal. **161** (2002), no. 3, 257–269. MR 1894593
11. Thomas A. Courtade, Max Fathi, and Ashwin Pananjady, *Quantitative stability of the entropy power inequality*, IEEE Trans. Inform. Theory **64** (2018), no. 8, 5691–5703. MR 3832330
12. I. Csiszár, *I-divergence geometry of probability distributions and minimization problems*, Ann. Probability **3** (1975), 146–158. MR 0365798
13. M. Cuturi, *Sinkhorn distances: Lightspeed computation of optimal transport.*, in Advances in Neural Information Processing Systems (NIPS) (2013), no. 26, 2292–2300.
14. Guido De Philippis and Alessio Figalli, *Rigidity and stability of Caffarelli's log-concave perturbation theorem*, Nonlinear Anal. **154** (2017), 59–70. MR 3614644
15. Montacer Essid and Michele Pavon, *Traversing the Schrödinger Bridge Strait: Robert Fortet's Marvelous Proof Redux*, J. Optim. Theory Appl. **181** (2019), no. 1, 23–60. MR 3921394
16. Robert Fortet, *Résolution d'un système d'équations de M. Schrödinger*, J. Math. Pures Appl. **19** (1940), 83–105. MR 0002692
17. Joel Franklin and Jens Lorenz, *On the scaling of multidimensional matrices*, Linear Algebra Appl. **114/115** (1989), 717–735. MR 986904

18. A. Genevay, *Entropy-regularized optimal transport for machine learning*, Ph.D. thesis, Université Paris-Dauphine, 2019.
19. Ivan Gentil, Christian Léonard, and Luigia Ripani, *About the analogy between optimal transport and minimal entropy*, Ann. Fac. Sci. Toulouse Math. (6) **26** (2017), no. 3, 569–601. MR 3669966
20. Ivan Gentil, Christian Léonard, Luigia Ripani, and Luca Tamanini, *An entropic interpolation proof of the HWI inequality*, Stochastic Processes and their Applications (2019).
21. N. Gigli and L. Tamanini, *Benamou-Brenier and duality formulas for the entropic cost on $RCD^*(k,n)$ spaces*, arxiv preprint (2018).
22. N. Gozlan, M. Madiman, C. Roberto, and P. M. Samson, *Deviation inequalities for convex functions motivated by the Talagrand conjecture*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **457** (2017), no. Veroyatnost' i Statistika. 25, 168–182. MR 3723580
23. Nathael Gozlan and Nicolas Juillet, *On a mixture of Brenier and Strassen theorems*, Proc. Lond. Math. Soc. (3) **120** (2020), no. 3, 434–463. MR 4008375
24. A. Guionnet and D. Shlyakhtenko, *Free monotone transport*, Invent. Math. **197** (2014), no. 3, 613–661. MR 3251831
25. Gilles Hargé, *A particular case of correlation inequality for the Gaussian measure*, Ann. Probab. **27** (1999), no. 4, 1939–1951. MR 1742895
26. ———, *Inequalities for the Gaussian measure and an application to Wiener space*, C. R. Acad. Sci. Paris Sér. I Math. **333** (2001), no. 8, 791–794. MR 1868955
27. J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of convex analysis*, Grundlehren Text Editions, Springer-Verlag, Berlin, 2001.
28. Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal, *Convex analysis and minimization algorithms. I*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 305, Springer-Verlag, Berlin, 1993, Fundamentals. MR 1261420
29. Young-Heon Kim and Emanuel Milman, *A generalization of Caffarelli's contraction theorem via (reverse) heat flow*, Math. Ann. **354** (2012), no. 3, 827–862. MR 2983070
30. Bo'az Klartag, *Poincaré inequalities and moment maps*, Ann. Fac. Sci. Toulouse Math. **22** (2013), no. 1, 1–41.
31. A. V. Kolesnikov, *Global Hölder estimates for optimal transportation*, Mat. Zametki **88** (2010), no. 5, 708–728. MR 2868394
32. ———, *On Sobolev regularity of mass transport and transportation inequalities*, Theory Probab. Appl. **57** (2013), no. 2, 243–264. MR 3201654
33. Christian Léonard, *From the Schrödinger problem to the Monge-Kantorovich problem*, J. Funct. Anal. **262** (2012), no. 4, 1879–1920. MR 2873864
34. ———, *A survey of the Schrödinger problem and some of its connections with optimal transport*, Discrete Contin. Dyn. Syst. **34** (2014), no. 4, 1533–1574. MR 3121631
35. ———, *Revisiting Fortet's proof of existence of a solution to the Schrödinger system*, ArXiv preprint arXiv:1904.13211 (2019).
36. Toshio Mikami, *Monge's problem with a quadratic cost by the zero-noise limit of h-path processes*, Probab. Theory Related Fields **129** (2004), no. 2, 245–260. MR 2063377
37. Emanuel Milman, *On the role of convexity in isoperimetry, spectral gap and concentration*, Invent. Math. **177** (2009), no. 1, 1–43.
38. ———, *Spectral estimates, contractions and hypercontractivity*, J. Spectr. Theory **8** (2018), no. 2, 669–714. MR 3812812
39. András Prékopa, *On logarithmic concave measures and functions*, Acta Sci. Math. (Szeged) **34** (1973), 335–343. MR 0404557
40. R. Tyrrell Rockafellar, *Convex analysis*, Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J., 1970. MR 0274683
41. E. Schrödinger, *Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique*, Ann. Inst. H. Poincaré **2** (1932), no. 4, 269–310. MR 1508000
42. C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.
43. Cédric Villani, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new. MR 2459454

MF : Institut de Mathématiques de Toulouse, Université Paul Sabatier 118, route de Narbonne F-31062 Toulouse Cedex 9

*Email address*: max.fathi@math.univ-toulouse.fr

NG : Université Paris Descartes, MAP5, UMR 8145, 45 rue des Saints Pères, 75270 Paris Cedex 06

*Email address*: natael.gozlan@parisdescartes.fr

MP : Institut de Mathématiques de Toulouse, Université Paul Sabatier 118, route de Narbonne F-31062 Toulouse Cedex 9

*Email address*: maxime.prodhomme@math.univ-toulouse.fr