



HAL
open science

Global Regularity Estimates for Optimal Transport via Entropic Regularisation

Nathael Gozlan, Maxime Sylvestre

► **To cite this version:**

Nathael Gozlan, Maxime Sylvestre. Global Regularity Estimates for Optimal Transport via Entropic Regularisation. 2025. hal-04894959

HAL Id: hal-04894959

<https://hal.science/hal-04894959v1>

Preprint submitted on 17 Jan 2025

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

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



Distributed under a Creative Commons Attribution 4.0 International License

GLOBAL REGULARITY ESTIMATES FOR OPTIMAL TRANSPORT VIA ENTROPIC REGULARISATION

NATHAEL GOZLAN AND MAXIME SYLVESTRE

ABSTRACT. We develop a general approach to prove global regularity estimates for quadratic optimal transport using the entropic regularisation of the problem.

1. INTRODUCTION

The aim of the paper is to obtain general global regularity estimates for the Brenier optimal transport map between two probability measures on \mathbb{R}^n , in the spirit of the celebrated Caffarelli's contraction theorem, recalled below:

Theorem 1.1 (Caffarelli [5, 6]). *Let $\mu(dx) = e^{-V(x)}dx$, $\nu(dy) = e^{-W(y)}dy$ be two probability measures on \mathbb{R}^n such that $\text{dom}V = \mathbb{R}^n$ and $\text{dom}W$ is convex with non empty interior. Further assume that V, W are twice continuously differentiable on the interior of their domains and satisfy*

$$\nabla^2 V \leq \alpha_V Id, \quad \nabla^2 W \geq \beta_W Id,$$

with $\alpha_V, \beta_W > 0$. Then the optimal transport map for the quadratic transport problem from μ to ν is $\sqrt{\alpha_V/\beta_W}$ -Lipschitz.

Here, and in all the paper, optimality refers to the quadratic transport problem on \mathbb{R}^n . Given two probability measures μ, ν on \mathbb{R}^n with finite second moments, recall that the quadratic transport cost between μ and ν is the quantity

$$(1) \quad W_2^2(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int \|y - x\|^2 \pi(dxdy),$$

where $\|\cdot\|$ denotes the standard Euclidean norm on \mathbb{R}^n and $\Pi(\mu, \nu)$ the set of all couplings between μ and ν that is the set of all probability measures π on $\mathbb{R}^n \times \mathbb{R}^n$ such that π has marginals μ and ν . A map $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is called optimal for this transport problem, whenever $\nu = T_{\#}\mu$ and

$$\int \|T(x) - x\|^2 \mu(dx) = W_2^2(\mu, \nu),$$

which means that the coupling $\pi(dxdy) = \mu(dx)\delta_{T(x)}(dy)$ induced by T achieves the minimal value in (1). According to Brenier's Theorem [3], if μ is absolutely continuous with respect to Lebesgue, then there is a μ almost surely unique transport map which is of the form $T = \nabla\phi$ where ϕ is a convex function on \mathbb{R}^n .

While in general studying the regularity of $\nabla\phi$ is a delicate question [4], the interest of Caffarelli's result recalled above is to furnish simple conditions on the potentials V and W assuring that the optimal map is globally Lipschitz with an explicit and dimension free Lipschitz constant. This result has also found several applications in the field of geometric and functional inequalities [23, 24, 12, 32]. Namely, Caffarelli's contraction theorem applies in particular when $\mu = \gamma_n$ is the standard Gaussian probability measure and ν has a log concave density with respect to γ_n , and implies in this case

Date: January 17, 2025.

2020 Mathematics Subject Classification. 60A10, 49J55, 60G42.

Key words and phrases. Optimal transport, Entropy regularized Optimal Transport, Caffarelli contraction Theorem.

N.G. is supported by a grant of the Agence nationale de la recherche (ANR), Grant ANR-23-CE40-0017 (Project SOCOT).

that the Brenier transport map sending γ_n onto ν is 1-Lipschitz. This can be used for instance to transfer functional inequalities (Poincaré, Log-Sobolev, Gaussian Isoperimetric inequalities) known for γ_n to all of its log-concave perturbations with the same dimension free constant as γ_n . In a slightly different spirit, Caffarelli's contraction result can also be used to control the deficit in some functional inequalities [15, 14]. Caffarelli's contraction theorem has been extended to perturbations of log concave densities in [11] and, more recently, to $1/d$ -concave densities [9]. Construction of Lipschitz (but not necessarily optimal) transport maps between regular probability measures is also an active field of research, mainly motivated by geometric and functional inequalities. See for instance [25, 20, 30, 31].

The original proof of Theorem 1.1 relies upon the Monge-Ampère equation satisfied by the function ϕ such that $T = \nabla\phi$ and in particular the regularity theory for such equations. Recently, Fathi, Prod'homme and the first author have proposed in [19] a different proof of Theorem 1.1, based on the entropic regularisation of the quadratic transport problem, that circumvents the use of Monge-Ampère equations. A simpler implementation of this idea has then been proposed by Chewi and Pooladian [10]. Both proofs have in common to involve geometric tools related to the functional version of the Brunn-Minkowski inequality due to Prekopa-Leindler [35, 28, 36] (the stability of the set of log-concave functions under the Ornstein-Uhlenbeck semi-group in the case of [19] or the Poincaré type Brascamp-Lieb inequality in the case of [10]). In this paper, we propose a new adaptation of the entropic regularisation approach of [19, 10] that enables us to obtain other explicit global regularity results for the Brenier map. We recover in particular the Hölder regularity result by Kolesnikov [26]. In this new twist of the entropic regularisation approach, the Prekopa-Leindler inequality will play a plain role.

Before presenting into more details this entropic regularization approach, let us give a flavor of the kind of global regularity estimates this method can reach. One of the main result of the paper is the following (which is a direct consequence of Corollary 4.3): suppose that $\mu(dx) = e^{-V(x)} dx$ and $\nu(dy) = e^{-W(y)} dy$ are two probability measures on \mathbb{R}^n with finite second moments associated to potentials $V : \mathbb{R}^n \rightarrow \mathbb{R}$ and $W : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ such that

$$(2) \quad V((1-t)x_0 + tx_1) + t(1-t)\sigma(\|x_1 - x_0\|) \geq (1-t)V(x_0) + tV(x_1), \quad \forall x_0, x_1 \in \mathbb{R}^n, \forall t \in [0, 1]$$

and

$$(3) \quad W((1-t)y_0 + ty_1) + t(1-t)\rho(\|y_1 - y_0\|) \leq (1-t)W(y_0) + tW(y_1) \quad \forall y_0, y_1 \in \mathbb{R}^n, \forall t \in [0, 1],$$

where $\sigma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and $\rho : \mathbb{R}_+ \rightarrow \mathbb{R} \cup \{+\infty\}$ are two measurable functions, then the Brenier transport map T that takes μ on ν satisfies

$$(4) \quad \|T(x) - T(y)\| \leq \frac{2}{\|x - y\|} \int_0^{\|x-y\|} (\rho^{**})^{-1}(\sigma(s)) ds, \quad \forall x \neq y,$$

where $h^*(v) = \sup_{u \geq 0} \{uv - h(u)\}$, $v \geq 0$, is the monotone conjugate of a function $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ and $(\rho^{**})^{-1}$ is the generalized inverse of the non-decreasing function ρ^{**} defined by

$$(\rho^{**})^{-1}(s) = \sup\{u \geq 0 : \rho^{**}(u) \leq s\}, \quad s \geq 0.$$

Under the assumptions of Theorem 1.1, (2) holds with $\sigma(u) = \alpha_V \frac{u^2}{2}$ and (3) with $\rho(u) = \beta_W \frac{u^2}{2}$, and so the bound (4) exactly yields

$$\|T(x) - T(y)\| \leq \sqrt{\frac{\alpha_V}{\beta_W}} \|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

A function V satisfying (2) is called σ -smooth and a function W satisfying (3) is called ρ -convex. These classes of functions have been introduced in [38, 39] (see also [1] and [40]) and can be characterized in terms of growth properties of the gradient (see Section 2.3 for some explicit examples). For instance, if V satisfies (2) with $\sigma(u) = \alpha_V u^p$, and $\rho(u) = \beta_W u^q$, with $p \leq 2 \leq q$, then, one obtains from (4) that the Brenier map T from μ to ν is Hölder of exponent p/q :

$$\|T(x) - T(y)\| \leq \frac{2q}{p+q} \left(\frac{\alpha_V}{\beta_W} \right)^{\frac{1}{q}} \|x - y\|^{\frac{p}{q}}, \quad \forall x, y \in \mathbb{R}^n.$$

This latter bound recovers a result by Kolesnikov [27].

According to a classical result (recalled in Proposition 11 below), if ϕ is a convex function assumed to be σ_ϕ -smooth for some function $\sigma_\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, then it holds

$$\|\nabla\phi(x) - \nabla\phi(y)\| \leq 2 \frac{\sigma_\phi(\|x - y\|)}{\|x - y\|}, \quad \forall x \neq y.$$

(and this last inequality implies back the σ_ϕ -smoothness up to constant, if σ_ϕ is convex). Therefore, in order to prove that the Brenier map $T = \nabla\phi$ pushing μ onto ν satisfies the bound (4) it is enough to prove that the convex function ϕ is such that

$$(5) \quad \phi((1-t)x_0 + tx_1) + \bar{\sigma}(\|x_1 - x_0\|) \geq (1-t)\phi(x_0) + t\phi(x_1), \quad \forall x_0, x_1 \in \mathbb{R}^n, \forall t \in [0, 1],$$

for the modulus of smoothness $\bar{\sigma}$ given by

$$\bar{\sigma}(u) = \int_0^u (\rho^{**})^{-1}(\sigma(s)) ds, \quad u \geq 0.$$

A key feature of the bound (5) is that it is stable under pointwise convergence. This will play a crucial role in the entropic regularisation method that we shall now present. First of all, let us informally recall the entropy regularized transport problem. Given two probability measures μ, ν on \mathbb{R}^n having finite second moments, the entropic regularized transport cost between μ and ν is given, for all $\epsilon > 0$, by

$$\mathcal{C}_\epsilon(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int \frac{1}{2} \|x - y\|^2 d\pi + \epsilon H(\pi | \mu \otimes \nu),$$

where the relative entropy of π with respect to $\mu \otimes \nu$, is given by $H(\pi | \mu \otimes \nu) = \int \log \frac{d\pi}{d(\mu \otimes \nu)} d\pi$, if $\pi \ll \mu \otimes \nu$ (and is $+\infty$ otherwise). It is now well known [8], that $\mathcal{C}_\epsilon(\mu, \nu) \rightarrow \frac{1}{2} W_2^2(\mu, \nu)$ as $\epsilon \rightarrow 0$. Contrary to the classical transport problem (1), the minimizer π^ϵ of the entropic regularized transport problem (which exists and is unique) can be easily computed from μ, ν using a fixed point algorithm. More precisely, the optimal coupling π_ϵ is of the form

$$\pi_\epsilon(dx dy) = e^{\frac{\langle x, y \rangle - \phi_\epsilon(x) - \psi_\epsilon(y)}{\epsilon}} \mu(dx) \nu(dy),$$

with $(\phi_\epsilon, \psi_\epsilon)$ a couple of convex functions solution of the following system

$$\begin{aligned} \phi_\epsilon(x) &= \epsilon \log \left(\int e^{\frac{\langle x, y \rangle - \psi_\epsilon(y)}{\epsilon}} \nu(dy) \right), \quad \forall x \in \mathbb{R}^n \\ \psi_\epsilon(y) &= \epsilon \log \left(\int e^{\frac{\langle x, y \rangle - \phi_\epsilon(x)}{\epsilon}} \mu(dx) \right), \quad \forall y \in \mathbb{R}^n. \end{aligned}$$

Moreover, the so-called entropic Kantorovich potential ϕ_ϵ converges μ -almost everywhere (along to some sequence ϵ_k) to the Kantorovich potential ϕ such that $T = \nabla\phi$ [34]. To get the desired inequality (5) for ϕ , we prove it for ϕ_ϵ and then we let $\epsilon \rightarrow 0$. In view of the above system of equations, the main problem is to understand the effect of the Laplace transform on σ -smooth and ρ -convex functions. Let us introduce the operator \mathcal{L}_ϵ acting on measurable functions $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ as follows:

$$\mathcal{L}_\epsilon f(x) = \epsilon \log \left(\int e^{\frac{\langle x, y \rangle - f(y)}{\epsilon}} dy \right), \quad \forall x \in \mathbb{R}^n.$$

The second main contributions of this paper are Propositions 3.1 and 3.2 and can be summarized as follows: if $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, then

- (i) if f is σ -smooth, then $\mathcal{L}_\epsilon f$ is σ^* -convex,
- (ii) if g is ρ -convex, then $\mathcal{L}_\epsilon g$ is ρ^* -smooth.

Note that, as $\epsilon \rightarrow 0$, the rescaled log-Laplace operator \mathcal{L}_ϵ converges to the Fenchel-Legendre transform. Letting $\epsilon \rightarrow 0$ in (i) and (ii), one recovers the well known property [1] that the sets of σ -smooth and ρ -convex functions are in duality with respect to the Fenchel-Legendre conjugation (see Proposition 2.5 for a precise statement). While Property (i) is a simple consequence of Hölder inequality, the proof of Property (ii) requires the use of the Prekopa-Leindler inequality. Using (i) and (ii), the fixed point system of equations satisfied by $(\phi_\epsilon, \psi_\epsilon)$ can be used to obtain an estimate

of the modulus of smoothness of ϕ_ϵ that does not depend on ϵ , which eventually allows to get a global regularity estimate for $\nabla\phi$ as explained above.

The paper is organized as follows. In Section 2, we recall the definitions and properties of σ -smooth and ρ -convex functions. We also introduce directional variants of these notions, leading in particular to anisotropic regularity properties on the gradient. The section is completed by a collection of concrete examples. Section 3 deals with the proof of Items (i) and (ii) above. A directional variant is also considered. Section 4 implements the entropic regularization method for global regularity of optimal transport to prove (4) as well as directional variants of this result. Finally, Section 5 presents some applications of the preceding results and their links with the existing literature.

Acknowledgments. The second author would like to thank Guillaume Carlier for many interesting discussions during the preparation of this work.

2. MODULUS OF CONVEXITY AND MODULUS OF SMOOTHNESS

2.1. Modulus of convexity/smoothness. In order to extend regularity results on the hessian of (convex) functions we will look at quantification of convexity or default of convexity which will imply lower/upper bounds on the hessian if it exists. Thus we turn to the notion of modulus of convexity/smoothness studied in [38, 39, 1]. We refer to [40, Chapter 3, Section 3.5] for a synthetic presentation of these notions. These notions of moduli of convexity and smoothness are close to the notions of semi concavity/convexity [7].

In all what follows, $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ will be a function whose domain $\text{dom}f = \{x \in \mathbb{R}^n : f(x) < +\infty\}$ is assumed to be convex. For any $x_0, x_1 \in \mathbb{R}^n$ and $t \in]0, 1[$, introduce the mean deviation of f as

$$M_t^f(x_0, x_1) = (1-t)f(x_0) + tf(x_1) - f((1-t)x_0 + tx_1).$$

This is a well defined quantity in the following two cases:

- $x_0, x_1 \in \text{dom}f$, in which case, by convexity of the domain, $M_t^f(x_0, x_1) \in \mathbb{R}$
- $(1-t)x_0 + tx_1 \in \text{dom}f$, in which case $M_t^f(x_0, x_1) \in \mathbb{R} \cup \{+\infty\}$.

Following [38, 39, 1], we introduce now the *modulus of convexity* of f as

$$\rho_f(r) = \inf \left\{ \frac{1}{t(1-t)} M_t^f(x_0, x_1) \mid t \in]0, 1[, \|x_0 - x_1\| = r, x_0, x_1 \in \text{dom}f \right\}, \quad r \geq 0$$

and the *modulus of smoothness* of f as

$$\sigma_f(r) = \sup \left\{ \frac{1}{t(1-t)} M_t^f(x_0, x_1) \mid t \in]0, 1[, \|x_0 - x_1\| = r, x_0, x_1 \text{ s.t. } (1-t)x_0 + tx_1 \in \text{dom}f \right\}, \quad r \geq 0,$$

where, in all the paper, $\|\cdot\|$ denotes the standard Euclidean norm on \mathbb{R}^n . Note that by definition $\rho_f(0) = \sigma_f(0) = 0$. Note also that, for $r > 0$, $\rho_f(r) \in \mathbb{R} \cup \{\pm\infty\}$ and takes the value $+\infty$ if r is greater than the diameter of $\text{dom}f$. For $r > 0$, $\sigma_f(r) \in \mathbb{R} \cup \{\pm\infty\}$ and $\sigma_f(r) = -\infty$ for some $r > 0$ if and only if the domain of f is empty.

The function ρ_f is the greatest function $\rho : \mathbb{R}_+ \rightarrow \mathbb{R} \cup \{\pm\infty\}$ such that

$$(6) \quad f((1-t)x_0 + tx_1) + t(1-t)\rho(\|x_1 - x_0\|) \leq (1-t)f(x_0) + tf(x_1), \quad \forall t \in]0, 1[, \forall x_0, x_1 \in \text{dom}f,$$

and the function σ_f is the smallest function $\sigma : \mathbb{R}_+ \rightarrow \mathbb{R} \cup \{\pm\infty\}$ such that

$$(7) \quad f((1-t)x_0 + tx_1) + t(1-t)\sigma(\|x_1 - x_0\|) \geq (1-t)f(x_0) + tf(x_1), \quad \forall t \in]0, 1[, \forall x_0, x_1 \in \mathbb{R}^n \text{ s.t. } (1-t)x_0 + tx_1 \in \text{dom}f.$$

When (6) is satisfied for some function ρ , we say that f is ρ -convex, and when (7) is satisfied for some function σ , we say that f is σ -smooth. In particular, the function f is convex if and only if $\rho_f \geq 0$. The finiteness of σ_f ensures that f has full domain as shown in the following simple result (see also Lemma 2.3 of [1] and Proposition 3.5.2 of [40]).

Lemma 2.1. *If $\text{dom}f$ is non-empty and convex and $\sigma_f(r_o) < +\infty$ for some $r_o > 0$, then $\text{dom}f = \mathbb{R}^n$.*

For the sake of completeness, we include the short proof of this elementary fact .

Proof. Note that σ_f admits the following slightly different expression:

$$(8) \quad \sigma_f(r) = \sup \left\{ \frac{(1-t)f(x-try) + tf(x+(1-t)ry) - f(x)}{t(1-t)} \mid t \in]0, 1[, x \in \text{dom } f, \|y\| = 1 \right\}.$$

In particular, if $\sigma_f(r_0) < +\infty$ and $x \in \text{dom } f$, then $B(x, r_0) \subset \text{dom } f$, denoting by $B(x_0, r)$ the closed ball of radius r_0 centered at x . In other words, $\text{dom } f + B(0, r_0) \subset \text{dom } f$. Iterating, yields that $\text{dom } f = \mathbb{R}^n$. \square

In the applications, we will only consider functions f for which $\sigma_f > 0$ on \mathbb{R}_+^* , because of the following lemma.

Lemma 2.2. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be such that $\sigma_f(r_0) \leq 0$ for some $r_0 > 0$, then the function e^{-f} is not integrable.*

Remark 2.3. *It follows immediately from the lemma that, if $\mu(dx) = e^{-V(x)} dx$ is a probability measure on \mathbb{R}^n such that $\sigma_V < +\infty$, then one also has $\sigma_V > 0$ on \mathbb{R}_+^* .*

Proof. Suppose that $\sigma_f(r_0) \leq 0$ for some $r_0 > 0$. Take $u \in \mathbb{R}^n$ such that $\|u\| = r_0$; then we get

$$e^{-f(x+tu)} \leq e^{-(1-t)f(x)} e^{-tf(x+u)}, \quad \forall x \in \mathbb{R}^n.$$

So integrating with respect to x and applying Hölder inequality yields

$$\int e^{-f} \leq \int e^{-(1-t)f(x)} e^{-tf(x+u)} dx \leq \left(\int e^{-f} dx \right)^{1-t} \left(\int e^{-f} dx \right)^t = \int e^{-f} dx.$$

If $\int e^{-f} < +\infty$, then, by the equality case in Hölder inequality, one gets that there is some constant $\lambda > 0$ such that $e^{-f(x)} = \lambda e^{-f(x+u)}$ for almost every $x \in \mathbb{R}^n$. Integrating with respect to x gives a contradiction. Thus e^{-f} is not integrable. \square

Moreover note that the σ -smoothness of f implies a control on the growth at infinity.

Lemma 2.4. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a σ -smooth function with $\sigma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. Let $x \in \mathbb{R}^n$ such that f is differentiable at x . Then for any $y \in \mathbb{R}^n$ we have*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \sigma(\|y - x\|)$$

Proof. Let $x, y \in \mathbb{R}^n$ and for $t \in]0, 1[$ set $x_t = (1-t)x + ty$. By σ -smoothness of f we have

$$f(y) - f(x) - \frac{f(x_t) - f(x)}{t} \leq (1-t)\sigma(\|y - x\|)$$

Letting $t \rightarrow 0$ grants the result. \square

Recall that if $h : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, then h^* denotes its Legendre transform defined by

$$h^*(y) = \sup_{x \in \mathbb{R}^n} \{ \langle x, y \rangle - h(x) \}, \quad y \in \mathbb{R}^n.$$

In the case of a function $\alpha : \mathbb{R}_+ \rightarrow \mathbb{R} \cup \{\pm\infty\}$, α^* denotes the monotone conjugate of α , defined by

$$\alpha^*(u) = \sup_{t \geq 0} \{ tu - \alpha(t) \}, \quad u \geq 0.$$

The following result shows that ρ -convexity and σ -smoothness are properties dual from each other:

Proposition 2.5. *Let $f, g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be two functions with non-empty convex domains and $\sigma, \rho : \mathbb{R}_+ \rightarrow \mathbb{R} \cup \{\pm\infty\}$.*

- (i) *If f is σ -smooth, then f^* is σ^* -convex.*
- (ii) *If g is ρ -convex, then g^* is ρ^* -smooth.*

The proof of Proposition 2.5 for convex functions can be found in [1, Proposition 2.6] or [40, Proposition 3.5.3]. In the general case, the proof is exactly the same. The argument is recalled, for the sake of completeness, in the proof of the more general Proposition 2.9 below.

In this paper, we will mainly be interested in the situation where f is a convex function. In this case, the following result gathers known properties of σ_f and ρ_f .

Proposition 2.6. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper convex and lower semicontinuous function.*

(a) *The modulus $\rho_f : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ is such that*

$$\rho_f(ct) \geq c^2 \rho_f(t), \quad \forall c \geq 1, \forall t \geq 0,$$

and is, in particular, non-decreasing.

(b) *The modulus $\sigma_f : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ is convex, non-decreasing and lower semicontinuous and it holds $\sigma_f = (\rho_{f^*})^*$.*

(c) *Either $\text{dom}\sigma_f = \{0\}$ or $\text{dom}\sigma_f = \mathbb{R}_+$.*

For the sake of completeness, we include a sketch of proof of Proposition 2.6 (we refer to [40] or [1] for detailed arguments).

Proof. Item (a) was first proved in [38] ; a proof can be found in [40, Proposition 3.5.1]. See also Item (a) of Proposition 2.10 for a proof in a more general context. Let us prove (b). Since f is a proper convex and lower semicontinuous function, it satisfies $f = f^{**}$. Thus, applying (i) and (ii) of Proposition 2.5, we first conclude that f is σ_f^{**} convex. Thus, by minimality of σ_f , $\sigma_f^{**} \geq \sigma_f$. Since the inequality $\sigma_f^{**} \leq \sigma_f$ is always true, one gets that $\sigma_f^{**} = \sigma_f$, which proves that σ_f is convex and lower semicontinuous (a property that can also be derived directly from (8)). On the other hand, by (ii) of Proposition 2.5, f is also $(\rho_{f^*})^*$ -smooth, and so $\sigma_f \leq (\rho_{f^*})^*$. Since f^* is σ_f^* -convex, one gets, by maximality of ρ_{f^*} , that $\rho_{f^*} \geq \sigma_f^*$ and so, taking the conjugate, $\sigma_f = \sigma_f^{**} \geq (\rho_{f^*})^*$, which proves that $\sigma_f = (\rho_{f^*})^*$ (and in particular σ_f is non-decreasing as a monotone conjugate). Now, let us sketch the proof of (c). By conjugating the super-homogeneity property of convexity modulus (Item (a)), one sees that σ_f satisfies the following subhomogeneity property: for all $r \geq 0$ and $c \in [0, 1]$, it holds $\sigma_f(cr) \geq c^2 \sigma_f(r)$. Therefore, if $\sigma_f(u) < +\infty$ then $\sigma_f(2u) < +\infty$ which yields the claim. \square

Let us conclude this section by recalling the following interesting consequence of σ -smoothness, that will play a crucial role in Section 4. Recall that a subgradient of f at some point $x \in \text{dom}f$ is a vector $a \in \mathbb{R}^n$ such that

$$(9) \quad f(y) \geq f(x) + \langle a, y - x \rangle, \quad \forall y \in \mathbb{R}^n.$$

We denote by $\partial f(x)$ the set of all subgradients of f at x . Note that $\partial f(x)$ may be empty. Whenever f is convex and x belongs to the relative interior of the domain of f , it is well known that $\partial f(x) \neq \emptyset$.

Proposition 2.7. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a function with a non-empty convex domain. Then for all $x_0 \neq x_1 \in \text{dom}f$ and $y_0 \in \partial f(x_0)$, $y_1 \in \partial f(x_1)$, it holds*

$$(10) \quad \sigma_f^*(\|y_1 - y_0\|) \leq \sigma_f(\|x_1 - x_0\|)$$

and

$$(11) \quad \|y_1 - y_0\| \leq 2 \frac{\sigma_f(\|x_1 - x_0\|)}{\|x_1 - x_0\|}.$$

Note that (10) implies (11) because $\sigma^*(s) \geq rs - \sigma(r)$. Observe also that if $f = \frac{\|\cdot\|^2}{2}$, then $\sigma_f(r) = \frac{r^2}{2}$ and there is equality in (10) and (11). Proposition 2.7 is usually stated for convex functions, but is true also for non-convex functions (provided the subdifferentials at x_0 and x_1 are assumed non-empty). The arguments of the proof of Proposition 2.7 can be found in [40, Theorem 3.5.6]. We will give a more general result in Proposition 2.11 below, with a complete proof.

2.2. Directional moduli of convexity and smoothness. The notions of modulus of convexity and modulus of smoothness introduced in the last section only depend on the distance separating the endpoints between which the mean deviation is computed. Here, in order to derive results in anisotropic situations, we introduce a generalized version of the moduli which also depends on the direction.

Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a function with convex domain. The *directional modulus of convexity* of f is

$$R_f(d) = \inf \left\{ \frac{1}{t(1-t)} M_t^f(x_0, x_1) \mid t \in]0, 1[, x_1 - x_0 = d, x_0, x_1 \in \text{dom} f \right\}, \quad d \in \mathbb{R}^n$$

and the *directional modulus of smoothness* is

$$S_f(d) = \sup \left\{ \frac{1}{t(1-t)} M_t^f(x_0, x_1) \mid t \in]0, 1[, x_1 - x_0 = d, (1-t)x_0 + tx_1 \in \text{dom} f \right\}, \quad d \in \mathbb{R}^n.$$

By their very definition, the functions R_f and S_f are automatically even:

$$R_f(-d) = R_f(d) \quad \text{and} \quad S_f(-d) = S_f(d), \quad \forall d \in \mathbb{R}^n.$$

As in the non directional case R_f is the greatest even function $R : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ such that

$$(12) \quad f((1-t)x_0 + tx_1) + t(1-t)R(x_1 - x_0) \leq (1-t)f(x_0) + tf(x_1), \quad \forall t \in]0, 1[, \forall x_0, x_1 \in \text{dom} f.$$

If (12) is satisfied, we will say that f is directionally R -convex or R -convex if there is no ambiguity. Note that R is allowed to take negative values, and therefore an R -convex function is not necessarily convex.

Likewise S_f is the smallest even function $S : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ such that

$$(13) \quad f((1-t)x_0 + tx_1) + t(1-t)S(x_1 - x_0) \geq (1-t)f(x_0) + tf(x_1), \quad \forall t \in]0, 1[, \quad \forall x_0, x_1 \in \mathbb{R}^n \text{ s.t. } (1-t)x_0 + tx_1 \in \text{dom} f.$$

If (13) is satisfied, we will say that f is directionally S -smooth or S -smooth if there is no ambiguity. When an S -smooth function f has a non-empty domain, then $S > -\infty$ everywhere. Note that as before the smoothness of a function has a strong implication on its domain.

Lemma 2.8. *If $\text{dom} f$ is non empty and convex and $S_f(v) < +\infty$ for some v , then $\text{dom} f + \text{Vect}(v) \subset \text{dom} f$.*

Proof. Let $x \in \text{dom} f$. Since f is S_f smooth, then for any $t = 1/2$

$$+\infty > f(x) + \frac{1}{4}S_f(v) \geq \frac{1}{2} \left(f(x - \frac{1}{2}v) + f(x + \frac{1}{2}v) \right).$$

Thus $x - \frac{1}{2}v, x + \frac{1}{2}v \in \text{dom} f$ for any x in the domain of f . The convexity of the domain ensures that $\text{dom} f + \frac{1}{2}[-v, v] \subset \text{dom} f$ where $[a, b]$ is the line joining a and b . The result follows by iteration. \square

The results that hold for the moduli of convexity and smoothness still apply to the directional moduli. We will make repeated use of the following proposition which extends Proposition 2.5.

Proposition 2.9. *Let $f, g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be two functions with non-empty convex domains and $S, R : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ be even functions.*

- (i) *If f is S -smooth, then f^* is S^* -convex.*
- (ii) *If g is R -convex, then g^* is R^* -smooth.*

Proof. We begin with the proof of (i). Since the domain of f is non-empty, $S > -\infty$ everywhere. Let us prove that for all $y_0, y_1 \in \text{dom} f^*$ and $t \in]0, 1[$, it holds

$$f^*(y_t) + t(1-t)S^*(y_1 - y_0) \leq (1-t)f^*(y_0) + tf^*(y_1),$$

where $y_t := (1-t)y_0 + ty_1$. By convexity of f^* , one gets $f^*(y_t) < +\infty$, and since the domain of f is non-empty, $f^*(y_t) > -\infty$. Let $\epsilon > 0$; there is x such that

$$-\infty < f^*(y_t) \leq \langle y_t, x \rangle - f(x) + \epsilon.$$

Thus $f(x) < +\infty$ and by S -smoothness of f for $d \in \text{dom}S$ we have

$$f(x) + t(1-t)S(d) \geq (1-t)f(x-td) + tf(x+(1-t)d).$$

Combining the two preceding inequalities grants

$$f^*(y_t) - \epsilon \leq \langle y_t, x \rangle - (1-t)f(x-td) - tf(x+(1-t)d) + t(1-t)S(d).$$

Thus by definition of f^*

$$\begin{aligned} f^*(y_t) - \epsilon &\leq \langle y_t, x \rangle - (1-t)\langle y_0, x-td \rangle - t\langle y_1, x+(1-t)d \rangle \\ &\quad + (1-t)f^*(y_0) + tf^*(y_1) + t(1-t)S(d) \\ &\leq t(1-t)(S(d) - \langle y_1 - y_0, d \rangle) + (1-t)f^*(y_0) + tf^*(y_1). \end{aligned}$$

Finally minimizing over d grants

$$f^*((1-t)y_0 + ty_1) + t(1-t)S^*(y_1 - y_0) - \epsilon \leq (1-t)f^*(y_0) + tf^*(y_1)$$

The result follows by letting $\epsilon \rightarrow 0$.

The proof of (ii) is similar but we include it for completeness. Let us prove that for all $y_0, y_1 \in \mathbb{R}^n$ such that $y_t = (1-t)y_0 + ty_1 \in \text{dom}g^*$, it holds

$$g^*(y_t) + t(1-t)R^*(y_1 - y_0) \geq (1-t)g^*(y_0) + tg^*(y_1).$$

Let $x_0, x_1 \in \text{dom}f$. By R -convexity of g we have

$$(1-t)g(x_0) + tg(x_1) \geq t(1-t)R(x_1 - x_0) + g(x_t)$$

with $x_t = (1-t)x_0 + tx_1$. Thus

$$\begin{aligned} (1-t)(\langle x_0, y_0 \rangle - g(x_0)) + t(\langle x_1, y_1 \rangle - g(x_1)) &\leq (1-t)\langle x_0, y_0 \rangle + t\langle x_1, y_1 \rangle - g(x_t) - t(1-t)R(x_1 - x_0) \\ &\leq t(1-t)(\langle x_1 - x_0, y_1 - y_0 \rangle - R(x_1 - x_0)) + g^*(y_t) \\ &\leq t(1-t)R^*(y_1 - y_0) + g^*(y_t) \end{aligned}$$

where the second and last inequalities hold by definition of g^* and R^* . Finally optimizing over x_0, x_1 grants the result. \square

In case the function f is convex, Proposition 2.6 can also be extended in the following way.

Proposition 2.10. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper convex and lower semicontinuous function.*

(a) *The modulus $R_f : \mathbb{R}^n \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ is such that*

$$R_f(cv) \geq c^2 R_f(v), \quad \forall c \geq 1, \forall v \in \mathbb{R}^n.$$

(b) *The modulus $S_f : \mathbb{R}^n \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ is convex, non-decreasing and lower semicontinuous and it holds $S_f = (R_{f^*})^*$.*

(c) *The domain of S_f is a vector space.*

Proof. We start with the proof of item (a). Let $c \in]1, 2[$ and $v \in \mathbb{R}^n \setminus \{0\}$. First if $R_f(cv) = +\infty$ there is nothing to prove. Otherwise let $\epsilon > 0$, then there are x_0, x_1 and $t \in]0, \frac{1}{2}[$ such that $x_1 - x_0 = cv$ and

$$t(1-t)(R_f(cv) + \epsilon) > M_t^f(x_0, x_1).$$

Set $x_t = (1-t)x_0 + tx_1$ and $x_c = (1-c^{-1})x_0 + c^{-1}x_1$. Notice that $x_c - x_0 = v$ and $x_t = (1-ct)x_0 + ct x_c$. By R_f -convexity of f we have

$$\begin{aligned} M_t^f(x_0, x_1) &= (1-t)f(x_0) + tf(x_1) - ((1-ct)f(x_0) + ct f(x_c)) + M_{ct}^f(x_0, x_c) \\ &\geq ct [(1-c^{-1})f(x_0) + c^{-1}f(x_1) - f(x_c)] + ct(1-ct)R_f(v) \\ &\geq ct(1-c^{-1})c^{-1}R_f(cv) + ct(1-ct)R_f(v) \end{aligned}$$

because $ct \in]0, 1[$. Thus using the first inequality we deduce by letting $\epsilon \rightarrow 0$

$$t(1-t)R_f(cv) \geq ct(1-c^{-1})c^{-1}R_f(cv) + ct(1-ct)R_f(v)$$

Dividing by t and rearranging the terms grants

$$\frac{1-ct}{c}R_f(cv) \geq c(1-ct)R_f(v)$$

which is the wanted inequality. An induction ensures that $R_f(c^n v) \geq (c^n)^2 R_f(v)$ which gives the result. Let us prove (b). Since f is a proper convex and lower semicontinuous function, it satisfies $f = f^{**}$. Thus, applying (i) and (ii) of Proposition 2.9, we first conclude that f is S_f^{**} convex. Thus, by minimality of S_f , $S_f^{**} \geq S_f$. Since the inequality $S_f^{**} \leq S_f$ is always true, one gets that $S_f^{**} = S_f$, which proves that S_f is convex and lower semicontinuous. On the other hand, by (ii) of Proposition 2.9, f is also $(R_{f^*})^*$ -smooth, and so $S_f \leq (R_{f^*})^*$. Since f^* is S_{f^*} -convex, one gets, by maximality of R_{f^*} , that $R_{f^*} \geq S_{f^*}$ and so, taking the conjugate, $S_f = S_{f^*}^{**} \geq (R_{f^*})^*$, which proves that $S_f = (R_{f^*})^*$. Now let us prove (c). First $0 \in \text{dom} S_f$. By the preceding S_f is a convex function thus its domain is convex. Moreover it is even and in order to prove that its domain is a vector space it remains to prove that it is a cone. If $\text{dom} S_f = \{0\}$ there is nothing to prove. Otherwise let $d \in \text{dom} S_f$ and $c > 1$ then since $S_f = R_{f^*}^*$ we have

$$\begin{aligned} S_f(cd) &= \sup_v \langle cd, v \rangle - R_{f^*}(v) \\ &= \sup_v \langle cd, cv \rangle - R_{f^*}(cv) \\ &\leq \sup_v c^2 \langle d, v \rangle - c^2 R_{f^*}(v) = c^2 S_f(d) \end{aligned}$$

where the inequality holds by item (a). This inequality shows that $cd \in \text{dom} S_f$. The case $c \in [0, 1]$ is managed by convexity of the domain. Finally the domain is a convex cone containing 0 thus it is a vector space. \square

We are now ready to state a directional estimate on the gradient of S -smooth functions which is similar to Proposition 2.7 (and implies it). Recall the definition of the subgradient of a function given in (9).

Proposition 2.11. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a function with a non-empty convex domain and $S : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be even. If f is S -smooth, then for all $x_0, x_1 \in \text{dom} f$ and $y_0 \in \partial f(x_0), y_1 \in \partial f(x_1)$, it holds*

$$S^*(y_1 - y_0) \leq S(x_1 - x_0).$$

Note that in Section 4, this result will be applied to a Kantorovich potential, which is naturally convex.

Proof. Let $x_0, x_1 \in \text{dom} f$ and $y_0 \in \partial f(x_0), y_1 \in \partial f(x_1)$. It follows from the definition of a subgradient that, for $i = 0, 1$, $f^*(y_i) = \langle y_i, x_i \rangle - f(x_i) < +\infty$. Thus $y_0, y_1 \in \text{dom} f^*$. Since f is S -smooth, it follows from Proposition 2.9 that the function f^* is S^* -convex. Thus, for any $t \in]0, 1[$,

$$(1-t)f^*(y_0) + tf^*(y_1) \geq t(1-t)S^*(y_1 - y_0) + f^*(y_t),$$

where $y_t = (1-t)y_0 + ty_1$. Moreover, since $y_0 \in \partial f(x_0)$ we have $x_0 \in \partial f^*(y_0)$ (this is always true, even if f is not convex) and thus

$$(1-t)f^*(y_0) + tf^*(y_1) \geq t(1-t)S^*(y_1 - y_0) + f^*(y_0) + t\langle x_0, y_1 - y_0 \rangle.$$

Subtracting $f^*(y_0)$, dividing by t and letting $t \rightarrow 0$ grants

$$f^*(y_1) - f^*(y_0) \geq S^*(y_1 - y_0) + \langle x_0, y_1 - y_0 \rangle.$$

A symmetric statement holds by exchanging y_0, y_1 and x_0, x_1 . Since S^* is even, summing the two inequalities obtained that way gives

$$\langle x_1 - x_0, y_1 - y_0 \rangle \geq 2S^*(y_1 - y_0)$$

and thus

$$S^*(y_1 - y_0) \leq \langle x_1 - x_0, y_1 - y_0 \rangle - S^*(y_1 - y_0) \leq S(x_1 - x_0)$$

which is the desired result. \square

2.3. Examples of R -convex and S -smooth functions. As we have seen in Propositions 2.7 and 2.11, the modulus of smoothness implies some form of regularity of the subgradient of a function (similarly, the modulus of convexity is related to expansivity properties of the subgradient). In this section, we investigate the converse implication and show how regularity of the (sub)-gradient implies estimates on the smoothness/convexity moduli for some classical class of functions.

Quadratic functions. Let $\alpha \in \mathbb{R}$ and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ such that $f(x) = \alpha\|x\|^2$ then

$$\rho_f(r) = \sigma_f(r) = \alpha r^2, \quad R_f(d) = S_f(d) = \alpha\|d\|^2.$$

More generally, if A is a $d \times d$ matrix, then the function $f(x) = \langle x, Ax \rangle$ admits the following moduli

$$\begin{aligned} \rho_f(r) &= r^2 \min \text{Sp} \left(\frac{A + A^T}{2} \right), \quad \sigma_f(r) = r^2 \max \text{Sp} \left(\frac{A + A^T}{2} \right), \quad r \geq 0 \\ R_f(d) &= S_f(d) = \langle d, Ad \rangle, \quad d \in \mathbb{R}^n, \end{aligned}$$

where $\text{Sp}(M)$ denotes the spectrum of a matrix M .

Functions with a bounded hessian. Let f be a twice continuously differentiable function on \mathbb{R}^n ; we have the following equality for any $x_0, x_1 \in \mathbb{R}^n$

$$M_t^f(x_0, x_1) = \int_0^1 \langle x_1 - x_0, \nabla^2 f(x_s)(x_1 - x_0) \rangle \min(s(1-t), t(1-s)) ds$$

where $x_s = x_0 + s(x_1 - x_0)$. Since $\int_0^1 \min(s(1-t), t(1-s)) ds = t(1-t)/2$ we deduce the following moduli

$$\begin{aligned} \rho_f(r) &\geq \frac{1}{2} r^2 \inf_x \min \text{Sp} \nabla^2 f(x), \quad \sigma_f(r) \leq \frac{1}{2} r^2 \sup_x \max \text{Sp} \nabla^2 f(x), \quad r \geq 0 \\ R_f(d) &\geq \frac{1}{2} \inf_x \langle d, \nabla^2 f(x)d \rangle, \quad S_f(d) \leq \frac{1}{2} \sup_x \langle d, \nabla^2 f(x)d \rangle, \quad d \in \mathbb{R}^n. \end{aligned}$$

Note that these bounds on the moduli are non trivial if and only if we have a uniform lower or upper bound on the hessian.

Functions with a uniformly continuous gradient. When one does not have a control on the Hessian but only a uniform continuity estimate for the gradient, one can still obtain explicit bounds on the modulus of smoothness. More precisely, suppose that f is a continuously differentiable function such that ∇f admits a non-decreasing modulus of continuity ω then

$$\sigma_f(r) \leq 2r\omega(r), \quad r \geq 0.$$

Indeed let $x_0, x_1 \in \mathbb{R}^n$ and $t \in]0, 1[$. Set $x_t = (1-t)x_0 + tx_1$, then for $i = 0, 1$,

$$f(x_i) - f(x_t) - \langle \nabla f(x_t), x_i - x_t \rangle = \int_0^1 \langle \nabla f(x_t + s(x_i - x_t)) - \nabla f(x_t), x_i - x_t \rangle ds \leq \|x_i - x_t\| \omega(\|x_i - x_t\|).$$

Multiplying this inequality by t if $i = 1$ and $(1-t)$ if $i = 0$ and summing grants

$$M_t^f(x_0, x_1) \leq t(1-t)\|x_1 - x_0\|(\omega(t\|x_1 - x_0\|) + \omega((1-t)t\|x_1 - x_0\|)) \leq t(1-t)2\|x_1 - x_0\|\omega(\|x_1 - x_0\|)$$

which proves the claim.

For instance, this case encompasses functions $f \in C^{1,\alpha}(\mathbb{R}^n)$, with $0 < \alpha \leq 1$. Indeed, in this case $\omega(r) = \|f\|_{1,\alpha} r^\alpha$, $r \geq 0$, with $\|f\|_{1,\alpha}$ the Hölder constant of ∇f , and so we have

$$\sigma_f(r) \leq 2r^{1+\alpha}\|f\|_{1,\alpha}, \quad r \geq 0.$$

Finally, Lipschitz functions are also encompassed by this case. Indeed we have $\omega(r) = 2L$ with L the Lipschitz constant of the function f .

Radial functions. The following result is due to Vladimirov, Nesterov and Chekanov [38]. It is stated, without proof, in [39]. Since Reference [38] is difficult to find, an elementary proof of this result is given in Appendix A.

Proposition 2.12. *Let $\alpha : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a non-decreasing function such that $\alpha(ct) \geq c\alpha(t)$ for all $t \geq 0$ and $c \geq 1$. Define $A(r) = \int_0^r \alpha(u) du$, $r \geq 0$, and $f_\alpha(x) = A(\|x\|)$, $x \in \mathbb{R}^n$. Then, the function f_α is ρ -convex, with $\rho(r) = 2A(r/2)$, $r \geq 0$.*

Corollary 2.13. *Suppose that $\alpha : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is an increasing continuous function such that $\alpha(ct) \geq c\alpha(t)$ for all $t \geq 0$ and $c \geq 1$. Denote by $\alpha^{-1} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ the converse function of α . Then, the function $(f_\alpha)^*$ is σ -smooth with $\sigma(r) = 2 \int_0^r \alpha^{-1}(u) du$, $r \geq 0$.*

Proof. This follows immediately from Propositions 2.5 and 2.12. \square

3. ENTROPIC LEGENDRE TRANSFORM

As we have seen in the preceding section, the class of R -convex and S -smooth functions are in duality with respect to the classical Legendre transform. The goal of this section is to show a similar correspondence for the *entropic Legendre transform* introduced below. The motivation comes from Section 4 where our goal will be to derive estimates on the moduli of smoothness and convexity of the entropic Kantorovich potentials which are tied together via this entropic Legendre transform.

Given $\epsilon > 0$ and a positive Borel measure m on \mathbb{R}^n , the entropic Legendre transform $\mathcal{L}_{\epsilon, m}(\psi)$ of a measurable function $\psi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ with respect to the measure m is defined by

$$\mathcal{L}_{\epsilon, m}(\psi)(x) = \epsilon \ln \left(\int \exp \left(\frac{\langle x, y \rangle - \psi(y)}{\epsilon} \right) dm(y) \right), \quad x \in \mathbb{R}^n.$$

This function is always well defined in $\mathbb{R} \cup \{\pm\infty\}$, but might have an empty domain. If ψ is such that $\text{dom} \psi$ has a positive m measure, then $\mathcal{L}_{\epsilon, m}(\psi)(x) > -\infty$ for all $x \in \mathbb{R}^n$. Note that for $\epsilon = 1$, $\mathcal{L}_{1, m}(\psi)$ is the log-Laplace of the measure $e^{-\psi} dm$. Thanks to Hölder's inequality $\mathcal{L}_{\epsilon, \mathcal{H}^n}(\psi)$ is convex and thus has a convex domain. It is also lower semicontinuous, thanks to Fatou's lemma. Thus, provided its domain is non empty, $\mathcal{L}_{\epsilon, \mathcal{H}^n}(\psi)$ is a proper convex lower semicontinuous function.

In the limit $\epsilon \rightarrow 0$, the entropic Legendre transform approaches the classical Legendre transform. More precisely, denoting by \mathcal{H}^n the Lebesgue measure on \mathbb{R}^n and assuming that $\mathcal{L}_{\epsilon, \mathcal{H}^n}(\psi)(x) < +\infty$ for all $\epsilon > 0$ small enough, then

$$\mathcal{L}_{\epsilon, \mathcal{H}^n}(\psi)(x) \rightarrow \text{esssup}_{y \in \mathbb{R}^n} \{\langle x, y \rangle - \psi(y)\}$$

as $\epsilon \rightarrow 0$. In particular, if ψ is continuous and has a super linear growth at ∞ , then for all $x \in \mathbb{R}^n$

$$\mathcal{L}_{\epsilon, \mathcal{H}^n}(\psi)(x) \rightarrow \sup_{y \in \mathbb{R}^n} \{\langle x, y \rangle - \psi(y)\} = \psi^*(x)$$

as $\epsilon \rightarrow 0$.

In what follows, we will in particular extend properties (i) and (ii) of Propositions 2.5 and 2.9 to the entropic Legendre transform for $m = \mathcal{H}^n$.

3.1. ψ ρ -convex implies $\mathcal{L}_{\epsilon, \mathcal{H}^n}(\psi)$ ρ^* -smooth. The following result is an entropic analog of property (ii) of Propositions 2.5 and 2.9 (and formally gives back these results by letting $\epsilon \rightarrow 0$). For any $\ell \leq n$, we denote by \mathcal{H}^ℓ the ℓ dimensional Hausdorff measure.

Proposition 3.1.

- (a) *If $\psi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ has a convex domain with positive \mathcal{H}^n measure and is such that $\rho_\psi > -\infty$, then the function $\mathcal{L}_{\epsilon, \mathcal{H}^n}(\psi)$ is ρ_ψ^* -smooth. In other words, $\sigma_{\mathcal{L}_{\epsilon, \mathcal{H}^n}(\psi)} \leq \rho_\psi^*$.*
- (b) *More generally, if $\psi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ has a non-empty convex domain and is directionally R -convex, that is such that*

$$\psi((1-t)y_0 + ty_1) + t(1-t)R(y_1 - y_0) \leq (1-t)\psi(y_0) + t\psi(y_1), \quad \forall y_0, y_1 \in \text{dom} \psi, \forall t \in [0, 1],$$

where $R : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, then denoting by $m = \mathcal{H}_{|L}^\ell$ where L is the affine hull of the domain of ψ and ℓ the dimension of H , the function $\mathcal{L}_{\epsilon,m}(\psi)$ is directionally R^* -smooth which means it satisfies:

$$\mathcal{L}_{\epsilon,m}(\psi)((1-t)x_0 + tx_1) + t(1-t)R^*(x_1 - x_0) \geq (1-t)\mathcal{L}_{\epsilon,m}(\psi)(x_0) + t\mathcal{L}_{\epsilon,m}(\psi)(x_1),$$

for all $x_0, x_1 \in \mathbb{R}^n$ such that $(1-t)x_0 + (1-t)x_1 \in \text{dom}\mathcal{L}_{\epsilon,m}(\psi)$.

The proof of this result will make use of the well known Prekopa-Leindler [35, 36, 28] inequality, that we now recall : if $f_0, f_1, h : L \rightarrow \mathbb{R}_+$ are measurable functions defined on some affine subspace L of \mathbb{R}^n such that, for some $t \in]0, 1[$, it holds

$$h((1-t)y_0 + ty_1) \geq f_0^{1-t}(y_0)f_1^t(y_1), \quad \forall y_0, y_1 \in L$$

then

$$\int_L h \geq \left(\int_L f_0 \right)^{1-t} \left(\int_L f_1 \right)^t,$$

where integration is understood with respect to the Lebesgue measure on L , that is $\mathcal{H}_{|L}^\ell$ with ℓ the dimension of L . This result is usually stated for $L = \mathbb{R}^n$ but its extension to a general affine subspace is classical and straightforward.

Proof. Let us first prove Item (b). Let $x_0, x_1 \in \mathbb{R}^n$ and $y_0, y_1 \in \text{dom}\psi$. For $t \in]0, 1[$ set $x_t = (1-t)x_0 + tx_1$, $y_t = (1-t)y_0 + ty_1$, and assume that $x_t \in \text{dom}\mathcal{L}_{\epsilon,m}(\psi)$. By R -convexity of ψ since $R > -\infty$ we have $y_t \in \text{dom}\psi$ and a direct computation grants the following inequalities

$$\begin{aligned} \langle x_t, y_t \rangle - \psi(y_t) &\geq t(1-t)(R(y_1 - y_0) - \langle x_1 - x_0, y_1 - y_0 \rangle) + (1-t)(\langle x_0, y_0 \rangle - \psi(y_0)) + t(\langle x_1, y_1 \rangle - \psi(y_1)) \\ &\geq -t(1-t)R^*(x_1 - x_0) + (1-t)(\langle x_0, y_0 \rangle - \psi(y_0)) + t(\langle x_1, y_1 \rangle - \psi(y_1)) \end{aligned}$$

Note that the last inequality holds trivially if $R^*(x_1 - x_0) = +\infty$. Therefore, the functions f_0, f_1, h defined for all $y \in L$ by

$$h(y) = \exp\left(\frac{\langle x_t, y \rangle - \psi(y)}{\epsilon}\right), \quad f_0(y) = \exp\left(\frac{\langle x_0, y \rangle - \psi(y)}{\epsilon}\right), \quad f_1(y) = \exp\left(\frac{\langle x_1, y \rangle - \psi(y)}{\epsilon}\right)$$

satisfy

$$h(y_t) \geq \exp\left(\frac{-t(1-t)R^*(x_1 - x_0)}{\epsilon}\right) f_0^{1-t}(y_0)f_1^t(y_1),$$

for all $y_0, y_1 \in \text{dom}\psi$. Note that this inequality is also true if y_0 or y_1 is not in the domain of ψ , indeed the inequality will hold trivially because $f_i(y_i) = 0$ and $h \geq 0$. Therefore by Prekopa-Leindler inequality we have

$$\int_L h \geq \exp\left(\frac{-t(1-t)R^*(x_1 - x_0)}{\epsilon}\right) \left(\int_L f_0 \right)^{1-t} \left(\int_L f_1 \right)^t.$$

Since $m(\text{dom}\psi) > 0$ the integrals are strictly positive. Taking the logarithm and multiplying by ϵ yields

$$\mathcal{L}_{\epsilon,m}(\psi)(x_t) \geq -t(1-t)R^*(x_1 - x_0) + (1-t)\mathcal{L}_{\epsilon,m}(\psi)(x_0) + t\mathcal{L}_{\epsilon,m}(\psi)(x_1),$$

where once again the inequality is vacuous if $R^*(x_1 - x_0) = +\infty$. This holds for any x_0, x_1 such that $x_t \in \text{dom}\mathcal{L}_{\epsilon,m}(\psi)$ and thus item (b) is proven.

To prove (a), observe that if $R(u) = \rho_\psi(\|u\|)$, $u \in \mathbb{R}^n$, then $R^*(v) = \rho_\psi^*(\|v\|)$, $v \in \mathbb{R}^n$, and ψ is directionnally R -convex.

So, according to (b), for all $x_0, x_1 \in \mathbb{R}^n$, it holds

$$\mathcal{L}_{\epsilon,\mathcal{H}^n}(\psi)((1-t)x_0 + tx_1) + t(1-t)\rho_\psi^*(\|x_1 - x_0\|) \geq (1-t)\mathcal{L}_{\epsilon,\mathcal{H}^n}(\psi)(x_0) + t\mathcal{L}_{\epsilon,\mathcal{H}^n}(\psi)(x_1).$$

So, by definition of $\sigma_{\mathcal{L}_{\epsilon,\mathcal{H}^n}(\psi)}$, we get $\rho_\psi^*(r) \geq \sigma_{\mathcal{L}_{\epsilon,\mathcal{H}^n}(\psi)}(r)$ for all $r \geq 0$, which completes the proof. \square

3.2. ψ σ -smooth implies $\mathcal{L}_{\epsilon, \mathcal{H}^n}(\psi)$ σ^* -convex. The following result now establish an entropic analog of property (i) of Propositions 2.5 and 2.9.

Proposition 3.2.

- (a) Let $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function such that $\sigma_\psi > -\infty$; then $\mathcal{L}_{\epsilon, \mathcal{H}^n}(\psi)$ is σ_ψ^* -convex. In other words, $\rho_{\mathcal{L}_{\epsilon, \mathcal{H}^n}(\psi)} \geq \sigma_\psi^*$.
- (b) More generally let $\psi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be such that $K := \text{dom}\psi$ is an affine subspace of \mathbb{R}^n . Moreover assume that ψ is directionally S -smooth i.e.

$$\psi((1-t)y_0 + ty_1) + t(1-t)S(y_1 - y_0) \geq (1-t)\psi(y_0) + t\psi(y_1), \forall t \in [0, 1],$$

for all $y_0, y_1 \in \mathbb{R}^n$ such that $(1-t)y_0 + ty_1 \in \text{dom}\psi$, where $S : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is such that $\text{dom}S \subset \overrightarrow{K}$. Denote by k the dimension of K and set $m = \mathcal{H}_{|K}^k$. Then $\mathcal{L}_{\epsilon, m}(\psi)$ is S^* -convex, that is

$$\mathcal{L}_{\epsilon, m}(\psi)((1-t)x_0 + tx_1) + t(1-t)S^*(x_1 - x_0) \leq (1-t)\mathcal{L}_{\epsilon, m}(\psi)(x_0) + t\mathcal{L}_{\epsilon, m}(\psi)(x_1)$$

for all $x_0, x_1 \in \mathbb{R}^n$.

Proof. As in Proposition 3.1 let us start with the proof of Item (b). Let us denote by K the domain of ψ which is assumed to be an affine subspace of \mathbb{R}^n and let $\overrightarrow{K} = \{x_1 - x_0 : x_0, x_1 \in K\}$ be the associated vector subspace of directions. Let $x_0, x_1 \in \text{dom}\mathcal{L}_{\epsilon, m}(\psi)$ (otherwise there is nothing to prove) and $v \in \overrightarrow{K}$. For $y \in \mathbb{R}^n$, applying the S -convexity of ψ to $y_0 = y$ and $y_1 = y + v$, grants

$$\langle x_t, y + tv \rangle - \psi(y + tv) \leq t(1-t)(S(v) - \langle x_1 - x_0, v \rangle) + (1-t)(\langle x_0, y \rangle - \psi(y)) + t(\langle x_1, y + v \rangle - \psi(y + v)).$$

Note that if $y + tv \notin \text{dom}\psi$ the inequality holds vacuously. Therefore, the functions f_0, f_1, h defined by

$$h(y) = \exp\left(\frac{\langle x_t, y + tv \rangle - \psi(y + tv)}{\epsilon}\right), \quad f_0(y) = \exp\left(\frac{\langle x_0, y \rangle - \psi(y)}{\epsilon}\right), \quad f_1(y) = \exp\left(\frac{\langle x_1, y + v \rangle - \psi(y + v)}{\epsilon}\right)$$

satisfy

$$h(y) \leq \exp\left(\frac{t(1-t)(S(v) - \langle x_1 - x_0, v \rangle)}{\epsilon}\right) f_0^{1-t}(y) f_1^t(y).$$

Thus by Hölder's inequality we have

$$\int h dm \leq \exp\left(\frac{t(1-t)(S(v) - \langle x_1 - x_0, v \rangle)}{\epsilon}\right) \left(\int f_0(y) dm\right)^{1-t} \left(\int f_1(y) dm\right)^t$$

Note that by the invariance of the measure m with respect to the translation by a vector in \overrightarrow{K} we have

$$\epsilon \ln \left(\int h dm\right) = \mathcal{L}_{\epsilon, m}(\psi)(x_t), \quad \epsilon \ln \left(\int f_1(y) dm\right) = \mathcal{L}_{\epsilon, m}(\psi)(x_1).$$

Thus taking the logarithm and multiplying by ϵ in the inequality above grants

$$\mathcal{L}_{\epsilon, m}(\psi)(x_t) \leq (1-t)\mathcal{L}_{\epsilon, m}(\psi)(x_0) + t\mathcal{L}_{\epsilon, m}(\psi)(x_1) + t(1-t)(S(v) - \langle x_1 - x_0, v \rangle)$$

for any $v \in \overrightarrow{K}$. Finally, optimizing over $v \in \overrightarrow{K}$ grants the result (since $\text{dom}S \subset \overrightarrow{K}$ by assumption). The proof of Item (a) follows by setting $S(v) = \sigma_\psi(\|v\|)$ in a similar fashion to Proposition 3.1. \square

3.3. A remark on the Cramér transform. Let X be a random vector with values in \mathbb{R}^n . We recall that the log-Laplace transform of X is the function $\Lambda_X : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ defined by

$$\Lambda_X(x) = \log \mathbb{E}[e^{\langle x, X \rangle}], \quad x \in \mathbb{R}^n.$$

When X has a density of the form $e^{-\psi}$ with respect to Lebesgue, with $\psi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, then

$$\Lambda_X = \mathcal{L}_{1, \mathcal{H}^n}(\psi)$$

The Cramér transform of X , denoted Λ_X^* , is defined as the Legendre transform of Λ_X :

$$\Lambda_X^*(y) = \sup_{x \in \mathbb{R}^n} \{\langle x, y \rangle - \Lambda_X(x)\}, \quad y \in \mathbb{R}^n$$

and plays an important role in Large Deviation Theory, see e.g [16]. Combining the results of Propositions 3.1, 3.2 and 2.5, we obtain the following:

Proposition 3.3. *Let X be a random vector with density $e^{-\psi}$ with respect to Lebesgue.*

- (1) *The function Λ_X^* is ρ_ψ^{**} -convex. In other words, $\rho_{\Lambda_X^*} \geq \rho_\psi^{**}$.*
- (2) *If ψ is finite valued, then Λ_X^* is σ_ψ -smooth. In other words, $\sigma_{\Lambda_X^*} \leq \sigma_\psi$.*

A similar correspondence could be stated for directional moduli.

4. APPLICATION TO OPTIMAL TRANSPORT

Let μ, ν be two probability measures on \mathbb{R}^n with finite second moments. In this section, we consider the entropic regularization of the quadratic optimal transport problem between μ and ν : for all $\epsilon > 0$, let

$$(14) \quad \mathcal{C}_\epsilon(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int \frac{1}{2} \|x - y\|^2 d\pi + \epsilon H(\pi | \mu \otimes \nu),$$

where $\Pi(\mu, \nu)$ is the set of probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ admitting μ, ν as marginals. We refer to e.g. [33] for a detailed introduction to this topic.

According to [8],

$$\mathcal{C}_\epsilon(\mu, \nu) \rightarrow \frac{1}{2} W_2^2(\mu, \nu)$$

as $\epsilon \rightarrow 0$.

For any $\epsilon > 0$, the problem (14) admits a unique minimizer π_ϵ , which is of the form

$$\pi_\epsilon(dx dy) = e^{\frac{f_\epsilon(x) + g_\epsilon(y)}{\epsilon} - \frac{\|x - y\|^2}{2\epsilon}} \mu(dx) \nu(dy),$$

with $f_\epsilon, g_\epsilon : \mathbb{R}^n \rightarrow \mathbb{R}$ solutions of the following system of non linear equations :

$$\begin{aligned} f_\epsilon(x) &= -\epsilon \log \left(\int e^{\frac{g_\epsilon(y) - \|x - y\|^2}{2\epsilon}} \nu(dy) \right), & \forall x \in \mathbb{R}^n \\ g_\epsilon(y) &= -\epsilon \log \left(\int e^{\frac{f_\epsilon(x) - \|x - y\|^2}{2\epsilon}} \mu(dx) \right), & \forall y \in \mathbb{R}^n. \end{aligned}$$

Let us consider the functions $\phi_\epsilon, \psi_\epsilon$ defined by

$$\phi_\epsilon(x) = \frac{1}{2} \|x\|^2 - f_\epsilon(x), \quad x \in \mathbb{R}^n, \quad \text{and} \quad \psi_\epsilon(y) = \frac{1}{2} \|y\|^2 - g_\epsilon(y), \quad y \in \mathbb{R}^n.$$

The couple $(\phi_\epsilon, \psi_\epsilon)$ is now solution of the following system

$$(15) \quad \phi_\epsilon(x) = \epsilon \log \left(\int e^{\frac{\langle x, y \rangle - \psi_\epsilon(y)}{\epsilon}} \nu(dy) \right), \quad \forall x \in \mathbb{R}^n$$

$$(16) \quad \psi_\epsilon(y) = \epsilon \log \left(\int e^{\frac{\langle x, y \rangle - \phi_\epsilon(x)}{\epsilon}} \mu(dx) \right), \quad \forall y \in \mathbb{R}^n,$$

which can be rewritten as

$$\mathcal{L}_{\epsilon, \nu}(\psi_\epsilon) = \phi_\epsilon, \quad \mathcal{L}_{\epsilon, \mu}(\phi_\epsilon) = \psi_\epsilon.$$

In this section, we apply the results of Section 3 to estimate the modulus of smoothness of the convex function ϕ_ϵ . In doing so we will be able to deduce an upper bound on the modulus of smoothness of ϕ the Kantorovich potential of the optimal transport problem ($\epsilon = 0$).

Remark 4.1. *We use here the same entropic regularization as in [10]. In [19], a different entropic regularization of the quadratic transport problem was considered, related to Schrödinger bridges for the Ornstein-Uhlenbeck process (we refer to [29] for a survey on the Schrödinger problem).*

4.1. Case of absolutely continuous measures. Let us first assume that μ and ν are both absolutely continuous with respect to Lebesgue:

$$\mu(dx) = e^{-V(x)} dx \quad \text{and} \quad \nu(dy) = e^{-W(y)} dy$$

with $V : \mathbb{R}^n \rightarrow \mathbb{R}$ and $W : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ (in particular μ has full support). In this case, the system of equations (15), (16) satisfied by $(\phi_\epsilon, \psi_\epsilon)$ is equivalent to

$$\phi_\epsilon = \mathcal{L}_{\epsilon, \mathcal{H}^n}(\psi_\epsilon + \epsilon W), \quad \psi_\epsilon = \mathcal{L}_{\epsilon, \mathcal{H}^n}(\phi_\epsilon + \epsilon V).$$

The following result, based on Propositions 3.1 and 3.2, provides an (ϵ -independent) estimate of the smoothness modulus of ϕ_ϵ in terms of the smoothness modulus of V and the convexity modulus of W .

Theorem 4.2. *Let $\mu(dx) = e^{-V(x)}dx$ and $\nu(dy) = e^{-W(y)}dy$ be two measures on \mathbb{R}^n such that V is σ -smooth with $\text{dom}\sigma = \mathbb{R}^n$ and W such that $\rho_W > -\infty$ and $\rho_W^*(v) < +\infty$ for some $v > 0$. Then it holds for all $\epsilon > 0$*

$$\sigma_{\phi_\epsilon}(r) \leq \int_0^r (\rho_W^{**})^{-1}(\sigma_V(s)) ds, \quad \forall r \geq 0,$$

where

$$(\rho_W^{**})^{-1}(t) = \sup\{s \geq 0 \mid \rho_W^{**}(s) \leq t\}, \quad t \geq 0.$$

Note that $\rho_W^*(v) < +\infty$ for some $v > 0$ is equivalent to ρ_W^{**} is not identically 0.

In the proof below, we will make use of the infimum convolution operator \square defined as follows: for any functions $f, g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$,

$$f \square g(x) = \inf_{y \in \mathbb{R}^n} \{f(y) + g(x - y)\}, \quad x \in \mathbb{R}^n.$$

Proof. As stated above, $\psi_\epsilon = \mathcal{L}_{\epsilon, \mathcal{H}^n}(\phi_\epsilon + \epsilon V)$. Since $\text{dom}V = \mathbb{R}^n$ and $\psi_\epsilon \neq +\infty$ on the non-empty support of ν , we deduce that ψ_ϵ is a proper l.s.c. convex function, thus $\rho_{\psi_\epsilon} \geq 0$. Moreover by hypothesis $\rho_W > -\infty$, this ensures that $\rho_{\psi_\epsilon + \epsilon W} \geq \rho_{\psi_\epsilon} + \epsilon \rho_W > -\infty$. Thus, using Proposition 3.1, we deduce that $\sigma_{\phi_\epsilon} \leq \rho_{\psi_\epsilon + \epsilon W}^* \leq (\rho_{\psi_\epsilon} + \epsilon \rho_W)^*$. Similarly to ψ_ϵ , the function ϕ_ϵ is a convex proper l.s.c. function, thus $\sigma_{\phi_\epsilon} \geq 0$ and by hypothesis $\sigma_V > -\infty$. This allows us to apply Proposition 3.2 to get $\rho_{\psi_\epsilon} \geq (\sigma_{\phi_\epsilon} + \epsilon \sigma_V)^*$. Combining both inequalities above yields

$$\sigma_{\phi_\epsilon} \leq ((\sigma_{\phi_\epsilon} + \epsilon \sigma_V)^* + \epsilon \rho_W)^*$$

because the Legendre transform is non-increasing. Recall that if $f, g : \mathbb{R}_+ \rightarrow \mathbb{R} \cup \{+\infty\}$ are arbitrary functions, then $(f + g)^* \leq f^* \square g^*$ and $f^{**} \leq f$. Thus

$$\sigma_{\phi_\epsilon} \leq (\sigma_{\phi_\epsilon} + \epsilon \sigma_V) \square \epsilon \rho_W^*(\cdot / \epsilon).$$

Let $v > 0$ be such that $\rho_W^*(v) < +\infty$. Applying this inequality at ϵv gives

$$\sigma_{\phi_\epsilon}(\epsilon v) \leq (\sigma_{\phi_\epsilon} + \epsilon \sigma_V) \square \epsilon \rho_W^*(\cdot / \epsilon)(\epsilon v) \leq \epsilon \rho_W^*(v) < +\infty.$$

Therefore $\text{dom}\sigma_{\phi_\epsilon}$ is not reduced to $\{0\}$ and so, according to Item (c) of Proposition 2.6, $\text{dom}\sigma_{\phi_\epsilon} = \mathbb{R}_+$. Take $u, v \geq 0$; evaluating now the inequality at $u + \epsilon v$, one gets

$$\frac{\sigma_{\phi_\epsilon}(u + \epsilon v) - \sigma_{\phi_\epsilon}(u)}{\epsilon} \leq \sigma_V(u) + \rho_W^*(v).$$

Now, by the convexity of σ_{ϕ_ϵ} given by Proposition 2.6, we get

$$v \sigma'_{\phi_\epsilon}(u) - \rho_W^*(v) \leq \sigma_V(u),$$

where σ'_{ϕ_ϵ} denotes say the right derivative of σ_{ϕ_ϵ} . Taking the supremum in v grants

$$\rho_W^{**}(\sigma'_{\phi_\epsilon}(u)) \leq \sigma_V(u)$$

and so, by definition of $(\rho_W^{**})^{-1}$,

$$\sigma'_{\phi_\epsilon}(u) \leq (\rho_W^{**})^{-1}(\sigma_V(u)).$$

Finally integration grants

$$\sigma_{\phi_\epsilon}(r) \leq \int_0^r (\rho_W^{**})^{-1}(\sigma_V(s)) ds.$$

□

Corollary 4.3. *Under the assumptions of Theorem 4.2, the Brenier transport map T sending μ onto ν satisfies the following:*

$$(17) \quad \|T(x) - T(y)\| \leq \frac{2}{\|x - y\|} \int_0^{\|x-y\|} (\rho_W^{**})^{-1}(\sigma_V(s)) ds,$$

for μ almost all $x \neq y \in \mathbb{R}^n$.

Note that in the case where the function $\frac{1}{r} \int_0^r (\rho_W^{**})^{-1}(\sigma_V(s)) ds \rightarrow 0$ as $r \rightarrow 0^+$, the function T is uniformly continuous and the bound (17) holds for all $x \neq y \in \mathbb{R}^n$.

Proof. By Theorem 4.2 ϕ_ϵ is σ -smooth for any $\epsilon > 0$ where $\sigma(r) = \int_0^r (\rho_W^{**})^{-1}(\sigma_V(s)) ds$, $r \geq 0$. According to [34, Theorem 1.1], as $\epsilon \rightarrow 0$, the potential ϕ_ϵ converges in $L^1(\mu)$ towards ϕ the (up to constant unique) Brenier potential for the transport of μ onto ν , up to extraction of a subsequence. Thus, up to extraction of a subsequence, ϕ_ϵ converges μ almost everywhere to ϕ . Since μ has full support this convergence holds almost everywhere. Moreover, for all $\epsilon > 0$, it holds

$$\phi_\epsilon((1-t)x_0 + tx_1) + t(1-t)\sigma(\|x_1 - x_0\|) \geq (1-t)\phi_\epsilon(x_0) + t\phi_\epsilon(x_1), \quad \forall x_0, x_1 \in \mathbb{R}^n, \quad \forall t \in]0, 1[$$

So passing to the limit when $\epsilon \rightarrow 0$, one sees that ϕ satisfies the same inequality almost everywhere. But ϕ has full domain and is convex thus it is continuous and the inequality holds everywhere. In other words, ϕ is σ -smooth. Applying Proposition 2.7 gives that, for all $x \neq y \in \mathbb{R}^n$ and $a \in \partial\phi(x)$, $b \in \partial\phi(y)$,

$$\|b - a\| \leq \frac{2}{\|y - x\|} \int_0^{\|y-x\|} (\rho_W^{**})^{-1}(\sigma_V(s)) ds.$$

Since ϕ is almost everywhere differentiable and $\partial\phi(x) = \{\nabla\phi(x)\} = \{T(x)\}$ at every point of differentiability, this completes the proof. □

4.2. Case of measures supported on affine subspaces. The purpose of this section is twofold: adapt the Theorem 4.2 to directional moduli in order to take into account anisotropies of the measures and use this result to derive regularity estimates of the optimal transport between singular measures (which is a pathological case of anisotropy). Here we assume that there are K, L two affine subspaces of \mathbb{R}^n and two functions V, W defined respectively on K, L such that:

$$\mu(dx) = e^{-V(x)} \mathcal{H}_K(dx) \quad \text{and} \quad \nu(dy) = e^{-W(y)} \mathcal{H}_L(dy),$$

where \mathcal{H}_K and \mathcal{H}_L denote the Lebesgue measures on K, L , that is $\mathcal{H}_K = \mathcal{H}_{|K}^k$ and $\mathcal{H}_L = \mathcal{H}_{|L}^l$ with k, l the respective dimensions of K and L . Similarly to paragraph 4.1, the Schrödinger potentials $(\phi_\epsilon, \psi_\epsilon)$ satisfy :

$$\phi_\epsilon = \mathcal{L}_{\epsilon, \mathcal{H}_L}(\psi_\epsilon + \epsilon W), \quad \psi_\epsilon = \mathcal{L}_{\epsilon, \mathcal{H}_K}(\phi_\epsilon + \epsilon V).$$

In the same fashion we derive an (ϵ -independent) estimate on the directional smoothness modulus of ϕ_ϵ .

Theorem 4.4. *Let K, L be two affine subspaces of \mathbb{R}^n . Let $\mu(dx) = e^{-V(x)} \mathcal{H}_K(dx)$ and $\nu(dy) = e^{-W(y)} \mathcal{H}_L(dy)$ be two probability measures. Assume that V is S -smooth, for some $S : \mathbb{R}^n \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ such that $\text{dom}S = \overrightarrow{K}$ and W is R -convex with $R : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ such that $R^*(v) < +\infty$ for some $v \neq 0$. Further assume that $\text{dom}S \subset \text{Vect}(\text{dom}R^*)$. Then, for all $\epsilon > 0$, the following holds*

$$(18) \quad S_{\phi_\epsilon}(d) \leq \int_0^1 \sup_{R^{**}(p) \leq S(td)} \langle p, d \rangle dt, \quad \forall d \in \mathbb{R}^n.$$

Proof. The proof is essentially the same as the one of Theorem 4.2. Note that ψ_ϵ is a proper l.s.c. convex function. In particular $R_{\psi_\epsilon} \geq 0$ and, by hypothesis, $R_{\psi_\epsilon} + \epsilon R > -\infty$ thus, by Proposition 3.1, $S_{\phi_\epsilon} \leq (R_{\psi_\epsilon} + \epsilon R)^*$. Likewise, ψ_ϵ is a proper l.s.c. convex function and thus $S_{\phi_\epsilon} \geq 0$. Moreover,

$\text{dom}(S_{\phi_\epsilon} + \epsilon S) \subset \overrightarrow{K}$. Since $\phi_\epsilon + \epsilon V$ is $(S_{\phi_\epsilon} + \epsilon S)$ -smooth, Proposition 3.2 ensures $R_{\psi_\epsilon} \geq (S_{\phi_\epsilon} + \epsilon S)^*$. Combining the two inequalities above gives $S_{\phi_\epsilon} \leq ((S_{\phi_\epsilon} + \epsilon S)^* + \epsilon R)^*$, from which we deduce

$$(19) \quad S_{\phi_\epsilon} \leq (S_{\phi_\epsilon} + \epsilon S) \square \epsilon R^*(\cdot / \epsilon).$$

If $v \in \text{dom}R^*$, then (19) yields

$$S_{\phi_\epsilon}(\epsilon v) \leq \epsilon R^*(v) < +\infty.$$

Thus $\epsilon v \in \text{dom}S_{\phi_\epsilon}$. Since, according to Proposition 2.10, $\text{dom}S_{\phi_\epsilon}$ is a vector space, we conclude that $\text{Vect}(\text{dom}R^*) \subset \text{dom}S_{\phi_\epsilon}$.

Let $u \in \text{dom}S_{\phi_\epsilon}$ and $v \in \mathbb{R}^n$, then (19) yields

$$\frac{S_{\phi_\epsilon}(u + \epsilon v) - S_{\phi_\epsilon}(u)}{\epsilon} \leq S(u) + R^*(v)$$

which grants, by convexity of S_{ϕ_ϵ} , that for any $a \in \partial S_{\phi_\epsilon}(u)$, it holds

$$\langle a, v \rangle - R^*(v) \leq S(u).$$

Note that $\partial S_{\phi_\epsilon}(u)$ is not empty since the domain of S_{ϕ_ϵ} is a vector subspace of \mathbb{R}^n . Optimizing over v gives $R^{**}(a) \leq S(u)$, for all $a \in \partial S_{\phi_\epsilon}(u)$. Finally, for any $d \in \text{dom}S_{\phi_\epsilon}$, it holds

$$S_{\phi_\epsilon}(d) = \int_0^1 s'(t) dt$$

with $s'(t) = \lim_{r \rightarrow 0^+} \frac{S_{\phi_\epsilon}((t+r)d) - S_{\phi_\epsilon}(td)}{r}$, the directional derivative of S_{ϕ_ϵ} at td in the direction d . Using the well known identity $s'(t) = \sup_{a \in \partial S_{\phi_\epsilon}(td)} \langle a, d \rangle$, one gets

$$s'(t) \leq \sup_{R^{**}(p) \leq S(td)} \langle p, d \rangle$$

which proves (18) for $d \in \text{dom}S_{\phi_\epsilon}$. Since, by hypothesis, $\text{dom}S \subset \text{Vect}(\text{dom}R^*)$ and we have seen that $\text{Vect}(\text{dom}R^*) \subset \text{dom}S_{\phi_\epsilon}$, we conclude that $\text{dom}S \subset \text{dom}S_{\phi_\epsilon}$. Thus, if $d \notin \text{dom}S_{\phi_\epsilon}$, then for any $t > 0$, $td \notin \text{dom}S_{\phi_\epsilon}$ and so $td \notin \text{dom}S$, in which case

$$\sup_{R^{**}(p) \leq S(td)} \langle p, d \rangle = \sup_{p \in \mathbb{R}^n} \langle p, d \rangle = +\infty$$

for all $t \in (0, 1]$. Thus (18) is also true in this case, which completes the proof. \square

Letting $\epsilon \rightarrow 0$ enables to estimate the directional modulus of smoothness of Kantorovich potentials between μ and ν . Recall that a Kantorovich potential for the transport of μ onto ν is a lower semi continuous convex function $\phi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ such that

$$\int \phi d\mu + \int \phi^* d\nu = \sup_{\pi \in \Pi(\mu, \nu)} \iint \langle x, y \rangle \pi(dx dy).$$

Kantorovich potentials are not always unique. Nevertheless, under the assumptions of Theorem 4.4, Kantorovich potentials are uniquely determined on K (see e.g [2, Proposition B.2]): if ϕ_1 and ϕ_2 are two Kantorovich potentials, then $\phi_1 - \phi_2$ is constant on K . In all what follows, we will always work with the Kantorovich potential ϕ such that $\phi = +\infty$ on $\mathbb{R}^n \setminus K$ (which is thus uniquely determined on \mathbb{R}^n up to constant).

Corollary 4.5. *Under the assumptions of Theorem 4.4, the Kantorovich potential ϕ for the optimal transport problem from μ to ν satisfies*

$$(20) \quad \phi((1-t)x_0 + tx_1) + t(1-t)\bar{S}(x_1 - x_0) \geq (1-t)\phi(x_0) + t\phi(x_1),$$

for all $x_0, x_1 \in \mathbb{R}^n$ and for all $t \in]0, 1[$, where

$$\bar{S}(d) = \int_0^1 \sup_{R^{**}(p) \leq S(td)} \langle p, d \rangle dt, \quad d \in \mathbb{R}^n.$$

Proof. If $\phi((1-t)x_0 + tx_1) = +\infty$ or $\bar{S}(x_1 - x_0) = +\infty$, then (20) obviously holds. Suppose that $\phi((1-t)x_0 + tx_1) < +\infty$ and $\bar{S}(x_1 - x_0) < +\infty$. Note that $\text{dom}\bar{S} \subset \text{dom}S = \vec{K}$. Thus, $(1-t)x_0 + tx_1 \in K$ and $x_1 - x_0 \in \vec{K}$. Since K is an affine subspace, we conclude that $x_0 \in K$ and $x_1 \in K$. So it is enough to prove (20) when $x_0, x_1 \in K$. Following the proof of Corollary 4.3, we conclude with the same limiting argument that, for a fixed $t \in]0, 1[$, ϕ satisfies (20) for μ -almost every x_0, x_1 . However $\text{dom}\phi$ contains the support of μ which is K . Thus ϕ is continuous on K with respect to the relative topology on K and this ensures that (20) holds for any $x_0, x_1 \in K$, and completes the proof. \square

5. EXAMPLES

5.1. Global Hölder estimates. First let us give a proof of the initial contraction result by Caffarelli [5] stated in Theorem 1.1, using the arguments developed in this paper.

Proof of Theorem 1.1. Using the result of Section 2.3 we deduce that

$$\sigma_V(r) = \alpha_V \frac{r^2}{2}, \quad \text{and} \quad \rho_W(r) = \beta_W \frac{r^2}{2}, \quad r \in \mathbb{R}_+.$$

Thus Corollary 4.3 ensures

$$\|\nabla\phi(x) - \nabla\phi(y)\| \leq \frac{2}{\|x-y\|} \int_0^{\|x-y\|} (\rho_W^{**})^{-1}(\sigma_V(s)) ds$$

for any $x, y \in \mathbb{R}^n$. A direct computation grants $(\rho_W^{**})^{-1}(\sigma_V(s)) = \sqrt{\alpha_V/\beta_W} s$ and the result follows by integration. \square

The following generalization of Theorem 1.1, was obtained by Kolesnikov in [26], extending the result to less regular densities and obtaining global Hölder estimates:

Theorem 5.1 (Kolesnikov [26]). *Let $\mu(dx) = e^{-V(x)} dx$ and $\nu(dy) = e^{-W(y)} dy$ be two measures on \mathbb{R}^n where $V, W \in C^1$ are such that $\sigma_V(r) \leq \alpha_V r^p$ and $\rho_W(r) \geq \beta_W r^q$ with $1 \leq p \leq 2 \leq q$. Then, the optimal transport map $\nabla\phi$ from μ to ν satisfies*

$$\|\nabla\phi(x) - \nabla\phi(y)\| \leq \frac{2q}{p+q} \left(\frac{\alpha_V}{\beta_W} \right)^{\frac{1}{q}} \|x-y\|^{\frac{p}{q}}, \quad \forall x, y \in \mathbb{R}^n.$$

Proof. The reasoning is the same as before. Note that $\rho : r \mapsto \beta_W r^q$ is convex therefore $\rho^{**} = \rho$. From this we deduce that $(\rho_W^{**})^{-1}(\sigma_V(s)) \leq \left(\frac{\alpha_V}{\beta_W} \right)^{\frac{1}{q}} s^{\frac{p}{q}}$. A direct application of Corollary 4.3 gives the result. \square

The regularity results presented in Section 4 allow for more freedom in the choice of the smoothness modulus of V . The two following results on the regularity of the optimal transport from a log-Lipschitz measure to a Gaussian and from a Cauchy distribution to a Gaussian demonstrate this. As underlined in Section 2.3, the modulus of smoothness of a given function f is finite when it is Lipschitz. This situation encompasses the transport of the exponential measure to the Gaussian and is the object of the following proposition.

Proposition 5.2. *Let $\mu(dx) = e^{-V(x)} dx$ be a probability measure with finite second moment such that V is L -Lipschitz and $\nu(dy) = e^{-W(y)} dy = \gamma(dy)$ be the standard Gaussian measure on \mathbb{R}^n . Then the optimal transport map $\nabla\phi$ from μ to ν is $\frac{1}{2}$ -Hölder with a Hölder norm equal to $\frac{8\sqrt{2}}{3}L$.*

Proof. As stated in Section 2.3, the smoothness modulus of V satisfies $\sigma_V(r) \leq 4Lr$ and the convexity modulus of W satisfies $\rho_W(r) \geq \frac{r^2}{2}$. Thus by Corollary 4.3 the optimal transport map $\nabla\phi$ from μ to ν satisfies

$$\|\nabla\phi(x) - \nabla\phi(y)\| \leq \frac{2}{\|x-y\|} \int_0^{\|x-y\|} \sqrt{8Ls} ds = \frac{8\sqrt{2}}{3}L \|x-y\|^{\frac{1}{2}}$$

\square

Finally, when $\mu(dx) = e^{-V(x)}dx$ is a Cauchy distribution, then $V(x) = n \ln(1 + \|x\|^2) + C$ with C the normalizing constant. In particular, note that the modulus of smoothness of V is quadratic close to 0 but enjoys a linear behaviour at infinity. Thus, the optimal transport from μ to a Gaussian is Lipschitz at close range and Hölder at long range as demonstrated by the following proposition.

Proposition 5.3. *Let $\mu(dx) = e^{-V(x)}dx$ be a probability measure on \mathbb{R}^n with $V(x) = n \ln(1 + \|x\|^2) + C$ with C the normalizing constant and $\nu(dy) = \gamma(dy)$ be the standard Gaussian measure on \mathbb{R}^n . Then the optimal transport map $\nabla\phi$ from μ to ν satisfies*

$$\|\nabla\phi(x) - \nabla\phi(y)\| \leq \min\left(2\sqrt{3n}\|x - y\|, \frac{8\sqrt{2}}{3}\|x - y\|^{\frac{1}{2}}\right).$$

Proof. As before the result will follow from a direct application of Corollary 4.3. To derive the desired property of the smoothness modulus of V we will use the link between the continuity modulus of the gradient of V and the smoothness modulus of V showed in Section 2.3. We denote by ω the continuity modulus of ∇V . Let $r \in \mathbb{R}_+^*$ and $x, y \in \mathbb{R}^n$ be such that $\|x - y\| = r$. Assume w.l.o.g. that $\|x\| \geq \|y\|$, then

$$\begin{aligned} \|\nabla V(x) - \nabla V(y)\| &= \left\| \frac{2n}{1 + \|x\|^2}x - \frac{2n}{1 + \|y\|^2}y \right\| \\ &\leq \frac{2n}{1 + \|x\|^2}\|x - y\| + \left| \frac{2n}{1 + \|x\|^2} - \frac{2n}{1 + \|y\|^2} \right| \|y\| \\ &\leq 2nr + \frac{2n\langle x - y, y + x \rangle}{(1 + \|x\|^2)(1 + \|y\|^2)} \|y\| \\ &\leq 2nr + \frac{2n\|x - y\|\|y + x\|}{(1 + \|x\|^2)(1 + \|y\|^2)} \|y\| \\ &\leq 2nr + 2nr \frac{\|y\|\|y + x\|}{(1 + \|x\|^2)(1 + \|y\|^2)} \\ &\leq 3nr \end{aligned}$$

because of Cauchy-Schwarz inequality for the third inequality and $\frac{\|y\|\|y+x\|}{(1+\|x\|^2)(1+\|y\|^2)} \leq \frac{1}{2}$. Moreover observe that $\|\nabla V(x)\| \leq n$ thus

$$\omega(r) \leq \min(3nr, 2n).$$

Using Section 2.3, we deduce that $\sigma_V(r) \leq \min(6nr^2, 4nr)$. Since ν is the standard Gaussian measure, Corollary 4.3 ensures

$$\|\nabla\phi(x) - \nabla\phi(y)\| \leq \frac{2}{\|x - y\|} \int_0^{\|x-y\|} \sqrt{\min(12s^2, 8s)} ds \leq \min\left(2\sqrt{3n}\|x - y\|, \frac{8\sqrt{2}}{3}\|x - y\|^{\frac{1}{2}}\right).$$

□

5.2. Anisotropic measures. The next theorem gives an anisotropic extension of Caffarelli's theorem.

Theorem 5.4. *Let $\mu(dx) = e^{-V(x)}dx$ and $\nu(dy) = e^{-W(y)}dy$ be two probability measures on \mathbb{R}^n where $V, W \in C^2$ and are such that $\nabla^2 V \leq A^{-1}$ and $\nabla^2 W \geq B^{-1}$ with A, B two symmetric positive definite matrices. Then the optimal transport map $\nabla\phi$ from μ to ν satisfies*

$$(21) \quad \nabla^2\phi \leq B^{1/2} \left(B^{-1/2} A^{-1} B^{-1/2} \right)^{1/2} B^{1/2}.$$

Equality holds when $\mu = \mathcal{N}(0, A)$ and $\nu = \mathcal{N}(0, B)$.

Equality case in (21) is well known and stems from the fact that the linear map

$$T(x) = B^{1/2} \left(B^{-1/2} A^{-1} B^{-1/2} \right)^{1/2} B^{1/2} x = A^{-1/2} \left(A^{1/2} B A^{1/2} \right)^{1/2} A^{-1/2} x$$

is optimal for the transport between $\mu = \mathcal{N}(0, A)$ and $\nu = \mathcal{N}(0, B)$, see [17, 21].

The conclusion of Theorem 5.4 was first obtained by Valdimarsson [37, Theorem 1.2] in the special case where μ is a Gaussian measure and the matrices A, B commute, in which case the bound simply reads

$$\nabla^2 \phi \leq A^{-1/2} B^{1/2}.$$

In [10], Chewi and Pooladian recovered Valdimarsson's result for a general μ using their entropic regularization method. Here we finally remove the commutation assumption.

It turns out that the ϵ -independent bound provided by Theorem 4.4 is suboptimal in this case. Namely, a simple calculation shows that under the assumptions of Theorem 5.4, Theorem 4.4 yields

$$S_{\phi}(d) \leq \frac{1}{2} \|A^{-1/2} d\| \|B^{1/2} d\|, \quad \forall d \in \mathbb{R}^n,$$

which is suboptimal by Cauchy-Schwarz (say in the commuting case). Note however that this bound coincides with the conclusion of Theorem 5.4 in the case $A^{-1} = B$.

In the proof below, we adapt the reasoning of Theorem 4.4 to obtain an explicit optimal ϵ -dependent bound for the smoothness modulus of the entropic potentials ϕ_{ϵ} and we derive the desired inequality by letting $\epsilon \rightarrow 0$.

Proof. As shown in the proof of Theorem 4.4, the smoothness modulus $S_{\phi_{\epsilon}}$ satisfies the following functional inequality:

$$S_{\phi_{\epsilon}} \leq (S_{\phi_{\epsilon}} + \epsilon S) \square \epsilon R^*(\cdot / \epsilon),$$

with, by assumption on V, W ,

$$S(d) = \frac{1}{2} \langle d, A^{-1} d \rangle, \quad R(d) = \frac{1}{2} \langle d, B^{-1} d \rangle, \quad d \in \mathbb{R}^n.$$

Noticing that $R^*(d) = \frac{1}{2} \langle d, B d \rangle$, $d \in \mathbb{R}^n$, and that $S_{\phi_{\epsilon}}(0) = 0$, we deduce from the functional inequality above that for all integer $N \geq 1$ and sequence of points u_0, u_1, \dots, u_N with $u_0 = d$ and $u_N = 0$ it holds

$$S_{\phi_{\epsilon}}(d) \leq \epsilon \sum_{i=1}^N S(u_i) + \frac{1}{\epsilon} \sum_{i=0}^{N-1} R^*(u_i - u_{i+1}).$$

Denoting by $\ell^2(\mathbb{R}^n)$ the space of sequences $u = (u_i)_{i \geq 0}$ with values in \mathbb{R}^n such that $\sum_{i \geq 0} \|u_i\|^2 < +\infty$, we thus get that

$$S_{\phi_{\epsilon}}(d) \leq \inf_{u \in \ell^2(\mathbb{R}^n), u_0 = d} \left\{ \epsilon \sum_{i=1}^{+\infty} S(u_i) + \frac{1}{\epsilon} \sum_{i=0}^{+\infty} R^*(u_i - u_{i+1}) \right\} := S_{\epsilon}(d).$$

Let us calculate explicitly S_{ϵ} . Since the value of S_{ϵ} is given by a minimization problem involving a strictly convex l.s.c proper and coercive function on a closed convex set of the Hilbert space $\ell^2(\mathbb{R}^n)$, there exists a unique $\bar{u} \in \ell^2(\mathbb{R}^n)$ (depending on d and ϵ) such that $\bar{u}_0 = d$ and

$$S_{\epsilon}(d) = \epsilon \sum_{i=1}^{+\infty} S(\bar{u}_i) + \frac{1}{\epsilon} \sum_{i=0}^{+\infty} R^*(\bar{u}_i - \bar{u}_{i+1}).$$

By the first order condition, we get

$$\epsilon A^{-1} \bar{u}_i + \frac{1}{\epsilon} (2B(\bar{u}_i) - B\bar{u}_{i+1} - B\bar{u}_{i-1}) = 0, \quad i \geq 1.$$

Let $\bar{v}_i = B^{1/2} \bar{u}_i$, $i \geq 1$. Then, the sequence \bar{v} satisfies

$$\bar{v}_{i+1} = (\epsilon^2 M + 2I) \bar{v}_i - \bar{v}_{i-1},$$

with $M = B^{-1/2} A^{-1} B^{-1/2}$. Since M is symmetric, there exists an orthogonal matrix P such that $P^T M P = D$ with D a diagonal matrix. Set $\bar{w}_i = P^T \bar{v}_i$, $i \geq 1$, then

$$\bar{w}_{i+1} = (\epsilon^2 D + 2I) \bar{w}_i - \bar{w}_{i-1}, \quad i \geq 1,$$

and $\bar{w}_0 = P^T B^{1/2} d$. Denote by $a_1, \dots, a_n > 0$ the diagonal coefficients of D . Then, for all $k \in \{1, \dots, n\}$, one gets

$$(22) \quad \bar{w}_{i+1,k} = (2 + a_k \epsilon^2) \bar{w}_{i,k} - \bar{w}_{i-1,k}, \quad \forall i \geq 1.$$

The characteristic equation of this second order one dimensional scheme is $r^2 - (2 + a_k \epsilon^2) r + 1 = 0$, whose roots are the following:

$$r_k(\epsilon) = \frac{1}{2} \left(2 + a_k \epsilon^2 - \sqrt{(2 + a_k \epsilon^2)^2 - 4} \right) \quad \text{and} \quad s_k(\epsilon) = \frac{1}{2} \left(2 + a_k \epsilon^2 + \sqrt{(2 + a_k \epsilon^2)^2 - 4} \right).$$

Therefore, we conclude that $(\bar{w}_{i,k})_{i \geq 0}$ is a linear combination of $(r_k^i)_{i \geq 0}$ and $(s_k^i)_{i \geq 0}$. Since $s_k > 1$, the sequence $(s_k^i)_{i \geq 0}$ does not belong to $\ell^2(\mathbb{R})$. Therefore, the only possibility is that $\bar{w}_{i,k} = \bar{w}_{0,k} r_k^i$, $i \geq 0$. Note that

$$S(\bar{u}_i) = \frac{1}{2} \bar{u}_i^T A^{-1} \bar{u}_i = \frac{1}{2} \bar{v}_i^T M \bar{v}_i = \frac{1}{2} w_i^T D w_i = \frac{1}{2} \sum_{k=1}^n a_k \bar{w}_{i,k}^2 = \frac{1}{2} \sum_{k=1}^n a_k r_k^{2i} \bar{w}_{0,k}^2$$

and, similarly,

$$R^*(\bar{u}_i - \bar{u}_{i+1}) = \frac{1}{2} (\bar{w}_i - \bar{w}_{i+1})^T I (\bar{w}_i - \bar{w}_{i+1}) = \frac{1}{2} \sum_{k=1}^n (1 - r_k)^2 r_k^{2i} \bar{w}_{0,k}^2.$$

Thus,

$$\begin{aligned} S_\epsilon(d) &= \frac{\epsilon}{2} \sum_{i=1}^{+\infty} \sum_{k=1}^n a_k r_k^{2i} \bar{w}_{0,k}^2 + \frac{1}{2\epsilon} \sum_{i=0}^{+\infty} \sum_{k=1}^n (1 - r_k)^2 r_k^{2i} \bar{w}_{0,k}^2 \\ &= \frac{\epsilon}{2} \sum_{k=1}^n a_k \frac{r_k^2}{1 - r_k^2} \bar{w}_{0,k}^2 + \frac{1}{2\epsilon} \sum_{k=1}^n (1 - r_k)^2 \frac{1}{1 - r_k^2} \bar{w}_{0,k}^2 \end{aligned}$$

A simple calculation shows that

$$r_k(\epsilon) = 1 - \sqrt{a_k} \epsilon + o(\epsilon)$$

as $\epsilon \rightarrow 0$. This easily implies that

$$S_\epsilon(d) \rightarrow \frac{1}{2} \sum_{k=1}^n \sqrt{a_k} \bar{w}_{0,k}^2 = \frac{1}{2} \bar{w}_0^T D^{1/2} \bar{w}_0 = \frac{1}{2} \bar{v}_0^T M^{1/2} \bar{v}_0 = \frac{1}{2} d^T B^{1/2} M^{1/2} B^{1/2} d,$$

as $\epsilon \rightarrow 0$. With the same limiting argument as in Corollary 4.3, we conclude that the Kantorovich potential ϕ from μ to ν satisfies

$$S_\phi(d) \leq \frac{1}{2} \langle d, B^{1/2} M^{1/2} B^{1/2} d \rangle, \quad \forall d \in \mathbb{R}^n,$$

which completes the proof. \square

It is also possible to look at degenerate anisotropies, in particular when μ, ν are supported on affine subspaces. Before stating the result, let us recall some elementary facts about subgradient of convex functions defined on strict affine subspaces of \mathbb{R}^n . Let $\phi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be such that $\text{dom} \phi = K$ where K is some affine subspace of \mathbb{R}^n . Then

$$(23) \quad \partial \phi(x) = \partial \phi_K(x) + \vec{K}^\perp,$$

where \vec{K}^\perp is the orthogonal complement of \vec{K} and $\partial \phi_K(x) \subset \vec{K}$ is the subgradient at x of $\phi_K : K \rightarrow \mathbb{R} : y \mapsto \phi(y)$ (the restriction of ϕ to K). In other words, $v \in \vec{K}^\perp$ belongs to $\partial \phi_K(x)$ if

$$\phi(y) \geq \phi(x) + \langle v, y - x \rangle, \quad \forall y \in K.$$

Theorem 5.5. *Let $\mu(dx) = e^{-V(x)} \mathcal{H}_K(dx)$ and $\nu(dy) = e^{-W(y)} \mathcal{H}_L(dy)$ be two probability measures where K, L are affine subspaces of \mathbb{R}^n and V, W are twice continuously differentiable on K and L respectively and satisfy $\nabla^2 V \leq \alpha_V Id$ and $\nabla^2 W \geq \beta_W Id$, with $\alpha_V, \beta_W > 0$, and let ϕ be the Kantorovich potential ϕ from μ to ν . Then the restriction ϕ_K of ϕ to the affine subspace K is continuously differentiable and $\nabla\phi_K : K \rightarrow \vec{K}$ satisfies: for all $x_0, x_1 \in K$*

$$\|\nabla\phi_K(x_1) - \nabla\phi_K(x_0)\| \leq \sqrt{\frac{\alpha_V}{\beta_W}} C_{K,L} \|x_1 - x_0\|,$$

with $C_{K,L} := \sup_{e \in \vec{K}, \|e\|=1} \|\text{Proj}_{\vec{L}}(e)\|$ and where $\text{Proj}_{\vec{L}}$ denotes the orthogonal projection on L .

In the above statement, the hessian of V and W are understood as operators defined on the tangent spaces \vec{K} and \vec{L} respectively.

Remark 5.6. *Let μ, ν and ϕ be as in Theorem 5.5.*

- *If $K = \mathbb{R}^n$, then $C_{K,L} = 1$ and one obviously recovers the conclusions of Theorem 1.1.*
- *If π^* is an optimal coupling between μ and ν , then it is well known that*

$$\pi^*(\cup_{x \in K} \{x\} \times \partial\varphi(x)) = 1.$$

Equivalently, writing $\pi^(dxdy) = \mu(dx)\pi_x^*(dy)$ the disintegration of π^* with respect to its first marginal, one gets*

$$\pi_x^*(\partial\varphi(x)) = 1$$

for μ almost every $x \in K$. According to Theorem 5.5 and (23), one gets

$$\pi_x^*(\nabla\phi_K(x) + \vec{K}^\perp) = 1$$

for μ almost every $x \in K$ and $\nabla\phi_K$ is Lipschitz continuous on K . In other words, the mass coming from x is moved “horizontally” through a Lipschitz map $\nabla\phi_K$ along \vec{K} and then distributed “vertically” along \vec{K}^\perp . Only the horizontal component of mass transports is controlled.

- *In the case where K and L are two orthogonal vector subspaces, then $C_{K,L} = 0$ and so $\nabla\phi_K$ is constant (in fact equal to 0). The transport of mass is purely vertical in this case. In this degenerate situation, any coupling (X, Y) with $X \sim \mu$ and $Y \sim \nu$ is optimal. Indeed, by orthogonality,*

$$\|X - Y\|^2 = \|X\|^2 + \|Y\|^2.$$

- *If K, L are close from being orthogonal, then the constant $C_{K,L}$ can be small, and the Lipschitz constant of $\nabla\phi_K$ can be strictly smaller than $\sqrt{\alpha_V/\beta_W}$.*

Proof. The moduli satisfy the following $S_V(d) \leq \frac{\alpha_V}{2} \|d\|^2$ for $d \in \vec{K}$, $R_W(d) \geq \frac{\beta_W}{2} \|d\|^2$ for $d \in \vec{L}$ and $+\infty$ otherwise. Thus Corollary 4.5 ensures $S_\phi(d) \leq \bar{S}(d)$, with

$$\bar{S}(d) = \frac{1}{2} \sqrt{\frac{\alpha_V}{\beta_W}} \|d\| \|\text{Proj}_{\vec{L}}(d)\|$$

for $d \in \vec{K}$ and $+\infty$ otherwise. A simple calculation shows that the Legendre transform of \bar{S} is given by

$$\bar{S}^*(s) = \frac{1}{2} \sqrt{\frac{\beta_W}{\alpha_V}} \sup_{e \in \vec{K}, \|e\|=1} \frac{\langle s, e \rangle^2}{\|\text{Proj}_{\vec{L}}(e)\|}, \quad s \in \mathbb{R}^n,$$

with the convention $0/0 = 0$. Note in particular that $\bar{S}^*(s) = +\infty$ if $s \in \vec{L}^\perp \setminus \{0\}$. Thus, by Proposition 2.11, we have for $x_0, x_1 \in K$ and $y_0 \in \partial\phi(x_0), y_1 \in \partial\phi(x_1)$

$$(24) \quad \bar{S}^*(y_1 - y_0) \leq \bar{S}(x_1 - x_0).$$

In particular, $\bar{S}^*(y_1 - y_0) < +\infty$. Thus, if $e \in \vec{K} \cap \left(\vec{L}^\perp\right)$, then $\langle y_1 - y_0, e \rangle = 0$. With this remark, we then see that (24) is equivalent to

$$\langle y_1 - y_0, e \rangle^2 \leq \frac{\alpha_V}{\beta_W} \|x_1 - x_0\| \|\text{Proj}_{\vec{L}}(x_1 - x_0)\| \|\text{Proj}_{\vec{L}}(e)\|,$$

for all $e \in \vec{K}$ with $\|e\| = 1$. By definition of $C_{K,L}$ we deduce that

$$\langle y_1 - y_0, e \rangle^2 \leq \frac{\alpha_V}{\beta_W} C_{K,L}^2 \|x_1 - x_0\|^2,$$

for all $e \in \vec{K}$ with $\|e\| = 1$. Optimizing over e yields

$$\|\text{Proj}_{\vec{K}}(y_1) - \text{Proj}_{\vec{K}}(y_0)\| \leq \sqrt{\frac{\alpha_V}{\beta_W}} C_{K,L} \|x_1 - x_0\|.$$

In particular, taking $x_0 = x_1 = x \in K$ one concludes, in view of (23), that $\partial\phi_K(x)$ is reduced to a single point. This proves that ϕ_K is continuously differentiable on K and $\partial\phi_K(x) = \text{Proj}_{\vec{K}}(\partial\phi(x)) = \{\nabla\phi_K(x)\}$, $x \in K$, and completes the proof. \square

5.3. Growth estimates. It is important to note that the estimation of the smoothness modulus of the optimal transport map is still valid when the target measure is not log-concave. The estimates still hold for ρ_W which is not non-negative and are meaningful as soon as $\lim_{r \rightarrow +\infty} \rho_W(r) = +\infty$. In particular it is possible to obtain growth estimates for the optimal transport map (deduced from the modulus of continuity between far apart points).

Corollary 5.7. *Let $\mu(dx) = e^{-V(x)} dx$ and $\nu(dy) = e^{-W(y)} dy$ be two probability measures of \mathbb{R}^n . Assume that V is σ -smooth with $\sigma : \mathbb{R} \rightarrow \mathbb{R}_+$ non-decreasing and that W is ρ -convex with $\rho : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$. Then the optimal transport T from μ to ν satisfies*

$$\|T(x)\| \leq \|T(0)\| + 2(\rho^{**})^{-1}(\sigma(\|x\|)).$$

Proof. By Corollary 4.3 we have

$$\|T(x) - T(0)\| \leq \frac{2}{\|x\|} \int_0^{\|x\|} (\rho^{**})^{-1}(\sigma(s)) ds.$$

Since $\sigma \geq 0$ is non-decreasing and $(\rho^{**})^{-1}$ is non-decreasing over \mathbb{R}_+ we have the result. \square

This result is to be compared with the growth estimates obtained via concentration of measure in [18].

Theorem 5.8 ([18]). *Let α be a continuous decreasing function such that $\nu(\|\cdot\| \geq r) \leq \alpha(r)$ for $r \geq r_0$. Then for any $\lambda > 0$ and $x \in \mathbb{R}^n$, we have*

$$\|T(x)\| \leq \max(3r_0, 3\alpha^{-1}(\mu(B(x + 2\lambda u, \lambda))))$$

where $u(x) = \frac{T(x)}{\|T(x)\|}$ and T is the optimal transport map from μ to ν .

In order to compare the two results above let us first recall some results on the concentration of measure phenomenon.

Proposition 5.9 ([22]). *Let $p \in \mathcal{P}(\mathbb{R}^n)$ such that*

$$\alpha(W_1(p, q)) \leq H(q | p)$$

for any $q \in \mathcal{P}(\mathbb{R}^n)$ where $\alpha : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a non-decreasing function such that $\alpha(0) = 0$. We say that p satisfies a transport-entropy inequality of modulus α . Then p satisfies the following concentration property: for any 1-Lipschitz function f and $r \geq 0$ we have

$$p\left(f > \int f dp + r_0 + r\right) \leq e^{-\alpha(r)}$$

where $r_0 = \alpha^{-1}(\ln(2))$.

In order to deduce a concentration result for ν it is sufficient to prove that under the conditions of Corollary 5.7 ν satisfies a transport-entropy inequality. This is the goal of the following lemma.

Lemma 5.10. *Let $\nu(dx) = e^{-W(x)}dx$ be a probability measure over \mathbb{R}^n with finite second moment. Assume that W is ρ -convex. Then ν satisfies*

$$\rho^{**}(W_1(\nu, p))^+ \leq H(p | \nu)$$

for any $p \in \mathcal{P}(\mathbb{R}^n)$.

Proof. Let $x, y \in \mathbb{R}^n$. Following the proof of Proposition 2.7 we know that

$$V(y) - V(x) - \langle \nabla V(x), y - x \rangle \geq \rho(\|x - y\|) \geq \rho^{**}(\|x - y\|)$$

Thus by combining Proposition 1.1 of [13] which ensures that ν satisfies a transport-entropy inequality of modulus ρ and the convexity of ρ^{**} we have that $\rho^{**}(W_1(\nu, \cdot)) \leq H(\cdot | \nu)$. Finally the positivity of entropy grants the result. \square

Combining the two results above ensure that in the context of Corollary 5.7 we have the following concentration estimate

$$\nu(\|\cdot\| \geq r) \leq e^{-\rho^{**}(r-r_0)^+}$$

which holds for any $r \geq r_0 = \int \|\cdot\| d\nu + ((\rho^{**})^+)^{-1}(\ln(2))$. It now remains to lower bound the ball probability of μ in the context of Corollary 5.7. This is the goal of the following Lemma.

Lemma 5.11. *Let $\mu(dx) = e^{-V(x)}dx$ be a probability measure over \mathbb{R}^n with finite second moment. Assume that V is σ -smooth with σ non decreasing and that V admits a minimizer x^* at which it is differentiable. Then for any $x, u \in \mathbb{R}^n$ such that $\|u\| = 1$ we have that*

$$\ln(\mu(B(x + 2\|x\|u, \|x\|))) \geq -\min V - 2\sigma(\|x^*\| + 3\|x\|) + \ln(\omega_n \|x\|^n)$$

where ω_n is the volume of the unit ball.

Proof. Let $x \in \mathbb{R}^n$ we have by Jensen's inequality

$$\begin{aligned} \ln(\mu(B(x + 2\|x\|u, \|x\|))) &= \ln\left(\int_{B(x+2\|x\|u, \|x\|)} e^{-V(y)} dy\right) \\ &\geq -\frac{1}{\omega_n \|x\|^n} \int_{B(x+2\|x\|u, \|x\|)} V(y) dy + \ln(\omega_n \|x\|^n). \end{aligned}$$

Since V is σ -smooth Lemma 2.4 ensures

$$\begin{aligned} \frac{1}{\omega_n \|x\|^n} \int_{B(x+2\|x\|u, \|x\|)} V(y) dy &\leq \\ \frac{1}{\omega_n \|x\|^n} \int_{B(x+2\|x\|u, \|x\|)} V(x + 2\|x\|u) + \langle \nabla V(x + 2\|x\|u), y - (x + 2\|x\|u) \rangle + \sigma(\|y - (x + 2\|x\|u)\|) dy & \\ \leq V(x + 2\|x\|u) + \sigma(\|x\|). & \end{aligned}$$

Finally using the same Lemma 2.4 we have $V(x + 2\|x\|u) \leq V(x^*) + \sigma(\|x + 2\|x\|u - x^*\|)$. We conclude using the non-decreasingness of σ . \square

Finally, under the assumptions of Corollary 5.7, combining the estimates for μ and ν gives the following growth estimate on the optimal transport map according to Theorem 5.8:

$$\|T(x)\| \leq 9r_0 + 3((\rho^{**})^+)^{-1}((\min V) - n \ln(\|x\|) - \ln(\omega_n) + 2\sigma(\|x^*\| + 3\|x\|)).$$

So, assuming that V achieves its minimum at 0 and that $\sigma(\|x\|) - n \ln(\omega_n \|x\|) \geq 0$ as $x \rightarrow +\infty$, one gets

$$\|T(x)\| \leq 6((\rho^{**})^+)^{-1}(\sigma(3\|x\|)),$$

as soon as $\|x\|$ is large enough, which is of the same order than the bound given by Corollary 5.7. This shows that Fathi's bound gives back the conclusion of Corollary 5.7 at a large scale.

To conclude, let us observe now that if μ is the standard Gaussian measure and ν is a *rotationnally* invariant probability measure, then the bound of Corollary 5.7 is always better than the bound

given by Theorem 5.8. Indeed, in this case $\sigma(u) = u^2/2$ and $T(x) = \tau(\|x\|)\frac{x}{\|x\|}$, for all $x \neq 0$, with $\tau : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, and so we get

$$\begin{aligned} \mu(B(x + 2\lambda u, \lambda)) &\leq \mu(\{y \in \mathbb{R}^n : \langle y, u \rangle \geq \langle x, u \rangle + \lambda\}) \\ &\leq \exp(-\frac{1}{2}(\langle x, u \rangle + \lambda)^2) \\ &= \exp(-\frac{1}{2}(\|x\| + \lambda)^2) \\ &\leq \exp(-\frac{1}{2}\|x\|^2), \end{aligned}$$

where $u = T(x)/\|T(x)\| = x/\|x\|$. Denoting $\alpha(r) = e^{-\rho^{**}(r-r_0)}$, $r \geq r_0 := \int \|\cdot\| d\nu + ((\rho^{**})^+)^{-1}(\ln(2))$, one gets

$$\begin{aligned} \inf_{\lambda > 0} \max(3r_0, 3\alpha^{-1}(\mu(B(x + 2\lambda u, \lambda)))) &\geq 3\alpha^{-1}(\exp(-\frac{1}{2}\|x\|^2)) \\ &= 3r_0 + 3((\rho^{**})^+)^{-1}(\sigma(\|x\|)) \\ &\geq \|T(0)\| + 2((\rho^{**})^+)^{-1}(\sigma(\|x\|)), \end{aligned}$$

since $T(0) = 0$.

APPENDIX A. PROOF OF PROPOSITION 2.12

Proof of Proposition 2.12. Suppose for a moment that, for all $x, y \in \mathbb{R}^n$, we have

$$(25) \quad f_\alpha\left(\frac{x+y}{2}\right) + \frac{1}{2}A\left(\frac{\|x-y\|}{2}\right) \leq \frac{1}{2}f_\alpha(x) + \frac{1}{2}f_\alpha(y).$$

Then, by convexity of f_α , for all $t \in [0, 1/2]$, we get

$$\begin{aligned} f_\alpha((1-t)x + ty) &= f_\alpha\left((1-2t)x + 2t\frac{x+y}{2}\right) \leq 2tf_\alpha\left(\frac{x+y}{2}\right) + (1-2t)f_\alpha(x) \\ &\leq (1-t)f_\alpha(x) + tf_\alpha(y) - tA\left(\frac{\|x-y\|}{2}\right) \\ &\leq (1-t)f_\alpha(x) + tf_\alpha(y) - t(1-t)A\left(\frac{\|x-y\|}{2}\right). \end{aligned}$$

By exchanging x and y , with see that the same inequality holds when $t \in [1/2, 1]$. It is thus enough to prove (25).

Suppose that (25) holds for all $x, y \in \mathbb{R}^n$ such that $x \cdot y \geq 0$. Then, if x, y are such that $x \cdot y \leq 0$, applying (25) to $x, -y$ yields

$$f_\alpha\left(\frac{x-y}{2}\right) + \frac{1}{2}f_\alpha\left(\frac{x+y}{2}\right) \leq \frac{1}{2}f_\alpha(x) + \frac{1}{2}f_\alpha(y).$$

Since, in this case, $\|x+y\| \leq \|x-y\|$, we get that

$$\begin{aligned} f_\alpha\left(\frac{x-y}{2}\right) &\geq \frac{1}{2}f_\alpha\left(\frac{x+y}{2}\right) + \frac{1}{2}f_\alpha\left(\frac{x-y}{2}\right) \\ &= \frac{1}{2}f_\alpha\left(\frac{x+y}{2}\right) + \frac{1}{2}A\left(\frac{\|x-y\|}{2}\right) \end{aligned}$$

and so (25) also holds in this case.

To conclude the proof, let us prove (25) for all x, y such that $x \cdot y \geq 0$. Let $u \in [0, 1]$ be such that $x \cdot y = \|x\|\|y\|u$. Then (25) is equivalent to

$$L(u) := A(N(u)) + \frac{1}{2}A(N(-u)) \leq \frac{1}{2}f_\alpha(x) + \frac{1}{2}f_\alpha(y),$$

with $N(u) = \frac{1}{2}\sqrt{\|x\|^2 + \|y\|^2 + 2\|x\|\|y\|u}$. The derivative of the function L is given by

$$L'(u) = \frac{\|x\|\|y\|}{4} \left[\frac{\alpha(N(u))}{N(u)} - \frac{1}{2} \frac{\alpha(N(-u))}{N(-u)} \right]$$

For $u \in [0, 1]$, $N(u) \geq N(-u)$, and so since the function $\alpha(t)/t$ is non-decreasing, we get that $L'(u) \geq 0$. So, it is enough to prove the inequality for $u = 1$, that is when $y = \lambda x$ for some $\lambda \geq 0$. Set $a = \|x\|$, $b = \|y\|$ and assume without loss of generality that $a \leq b$. We are reduced to prove that

$$\int_0^{\frac{a+b}{2}} \alpha(t) dt + \frac{1}{2} \int_0^{\frac{b-a}{2}} \alpha(t) dt \leq \frac{1}{2} \int_0^a \alpha(t) dt + \frac{1}{2} \int_0^b \alpha(t) dt.$$

Since the right hand side is equal to

$$\int_0^a \alpha(t) dt + \frac{1}{2} \int_a^{\frac{a+b}{2}} \alpha(t) dt + \frac{1}{2} \int_{\frac{a+b}{2}}^b \alpha(t) dt$$

this inequality amounts to

$$\int_a^{\frac{a+b}{2}} \alpha(t) dt + \int_0^{\frac{b-a}{2}} \alpha(t) dt \leq \int_{\frac{a+b}{2}}^b \alpha(t) dt,$$

that is

$$(26) \quad \int_0^{\frac{b-a}{2}} \alpha(a+t) dt + \int_0^{\frac{b-a}{2}} \alpha(t) dt \leq \int_0^{\frac{b-a}{2}} \alpha\left(\frac{a+b}{2} + t\right) dt.$$

Since $t \mapsto \frac{\alpha(t)}{t}$ is non-decreasing, the function α is super-additive:

$$\alpha(u+v) \geq \alpha(u) + \alpha(v), \quad \forall u, v \geq 0.$$

In particular, if $t \in [0, \frac{b-a}{2}]$,

$$\alpha\left(\frac{a+b}{2} + t\right) = \alpha\left(a+t + \frac{b-a}{2}\right) \geq \alpha(a+t) + \alpha\left(\frac{b-a}{2}\right) \geq \alpha(a+t) + \alpha(t).$$

Integrating this inequality gives (26) and completes the proof. \square

REFERENCES

- [1] Dominique Azé and Jean-Paul Penot. Uniformly convex and uniformly smooth convex functions. *Ann. Fac. Sci. Toulouse, Math. (6)*, 4(4):705–730, 1995.
- [2] Espen Bernton, Promit Ghosal, and Marcel Nutz. Entropic optimal transport: geometry and large deviations. *Duke Math. J.*, 171(16):3363–3400, 2022.
- [3] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.
- [4] Luis A. Caffarelli. The regularity of mappings with a convex potential. *J. Am. Math. Soc.*, 5(1):99–104, 1992.
- [5] Luis A. Caffarelli. Monotonicity properties of optimal transportation and the fkg and related inequalities. *Communications in Mathematical Physics*, 214(3):547–563, November 2000.
- [6] Luis A. Caffarelli. 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(2):449–450, 2002.
- [7] Piermarco Cannarsa and Carlo Sinestrari. *Semiconcave Functions, Hamilton—Jacobi Equations, and Optimal Control*. Birkhäuser Boston, 2004.
- [8] Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM Journal on Mathematical Analysis*, 49(2):1385–1418, January 2017.
- [9] Guillaume Carlier, Alessio Figalli, and Filippo Santambrogio. On Optimal Transport Maps Between $1/d$ -Concave Densities. Preprint, arXiv:2404.05456 [math.AP] (2024), 2024.
- [10] Sinho Chewi and Aram-Alexandre Pooladian. An entropic generalization of caffarelli’s contraction theorem via covariance inequalities. *Comptes Rendus. Mathématique*, 361(G9):1471–1482, November 2023.
- [11] 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(4):1491–1519, 2017.
- [12] Dario Cordero-Erausquin. Some applications of mass transport to Gaussian-type inequalities. *Arch. Ration. Mech. Anal.*, 161(3):257–269, 2002.

- [13] Dario Cordero-Erausquin. Transport inequalities for log-concave measures, quantitative forms, and applications. *Canadian Journal of Mathematics*, 69(3):481–501, June 2017.
- [14] Thomas A. Courtade, Max Fathi, and Ashwin Pananjady. Quantitative stability of the entropy power inequality. *IEEE Trans. Inform. Theory*, 64(8):5691–5703, 2018.
- [15] Guido De Philippis and Alessio Figalli. Rigidity and stability of Caffarelli’s log-concave perturbation theorem. *Nonlinear Anal.*, 154:59–70, 2017.
- [16] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications.*, volume 38 of *Stoch. Model. Appl. Probab.* Berlin: Springer, 2nd ed., corrected 2nd printing edition, 2010.
- [17] D. C. Dowson and B. V. Landau. The Fréchet distance between multivariate normal distributions. *J. Multivariate Anal.*, 12(3):450–455, 1982.
- [18] Max Fathi. Growth estimates on optimal transport maps via concentration inequalities, 2024.
- [19] Max Fathi, Nathael Gozlan, and Maxime Prod’homme. A proof of the Caffarelli contraction theorem via entropic regularization. *Calc. Var. Partial Differ. Equ.*, 59(3):18, 2020. Id/No 96.
- [20] Max Fathi, Dan Mikulincer, and Yair Shenfeld. Transportation onto log-Lipschitz perturbations. *Calc. Var. Partial Differential Equations*, 63(3):Paper No. 61, 25, 2024.
- [21] Matthias Gelbrich. On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Math. Nachr.*, 147:185–203, 1990.
- [22] Nathael Gozlan and Christian Léonard. Transport inequalities. a survey. 2010.
- [23] Gilles Hargé. A particular case of correlation inequality for the Gaussian measure. *Ann. Probab.*, 27(4):1939–1951, 1999.
- [24] Gilles Hargé. Inequalities for the Gaussian measure and an application to Wiener space. *C. R. Acad. Sci. Paris Sér. I Math.*, 333(8):791–794, 2001.
- [25] Young-Heon Kim and Emanuel Milman. A generalization of Caffarelli’s contraction theorem via (reverse) heat flow. *Math. Ann.*, 354(3):827–862, 2012.
- [26] A. V. Kolesnikov. Global Hölder estimates for optimal transportation. *Mat. Zametki*, 88(5):708–728, 2010.
- [27] A. V. Kolesnikov. On Sobolev regularity of mass transport and transportation inequalities. *Theory Probab. Appl.*, 57(2):243–264, 2013.
- [28] L. Leindler. On a certain converse of Hölder’s inequality. II. *Acta Sci. Math. (Szeged)*, 33(3-4):217–223, 1972.
- [29] Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4):1533–1574, 2014.
- [30] Dan Mikulincer and Yair Shenfeld. On the Lipschitz properties of transportation along heat flows. In *Geometric aspects of functional analysis*, volume 2327 of *Lecture Notes in Math.*, pages 269–290. Springer, Cham, [2023] ©2023.
- [31] Dan Mikulincer and Yair Shenfeld. The Brownian transport map. *Probab. Theory Related Fields*, 190(1-2):379–444, 2024.
- [32] Emanuel Milman. Spectral estimates, contractions and hypercontractivity. *J. Spectr. Theory*, 8(2):669–714, 2018.
- [33] Marcel Nutz. Introduction to entropic optimal transport, lecture notes, columbia university, 2021.
- [34] Marcel Nutz and Johannes Wiesel. Entropic optimal transport: convergence of potentials. *Probability Theory and Related Fields*, 184(1–2):401–424, November 2021.
- [35] András Prékopa. Logarithmic concave measures with application to stochastic programming. *Acta Sci. Math. (Szeged)*, 32:301–316, 1971.
- [36] András Prékopa. On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)*, 34:335–343, 1973.
- [37] Stefán Ingi Valdimarsson. On the Hessian of the optimal transport potential. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 6(3):441–456, 2007.
- [38] A. A. Vladimirov, Yu. E. Nesterov, and Yu. N. Chekanov. On uniformly convex functionals. *Mosc. Univ. Comput. Math. Cybern.*, 1978(3):10–21, 1978.
- [39] C. Zălinescu. On uniformly convex functions. *J. Math. Anal. Appl.*, 95:344–374, 1983.
- [40] C. Zălinescu. *Convex analysis in general vector spaces.* Singapore: World Scientific, 2002.

N. G. : UNIVERSITÉ PARIS CITÉ, CNRS, MAP5, F-75006 PARIS
 Email address: nathael.gozlan@u-paris.fr

M. S. : CEREMADE, UNIVERSITÉ PARIS-DAUPHINE, UNIVERSITÉ PSL, CNRS, MOKAPLAN, INRIA PARIS, 75016 PARIS, FRANCE
 Email address: maxime.sylvestre@dauphine.psl.eu