BV ESTIMATES IN OPTIMAL TRANSPORTATION AND APPLICATIONS

GUIDO DE PHILIPPIS, ALPÁR RICHÁRD MÉSZÁROS, FILIPPO SANTAMBROGIO, AND BOZHIDAR VELICHKOV

Abstract. In this paper we study the BV regularity for solutions of variational problems in Optimal Transportation. As an application we recover BV estimates for solutions of some non-linear parabolic PDE by means of optimal transportation techniques. We also prove that the Wasserstein projection of a measure with BV density on the set of measures with density bounded by a given BV function f is of bounded variation as well. In particular, in the case $f = 1$ (projection onto a set of densities with an L^∞ bound) we precisely prove that the total variation of the projection does not exceed the total variation of the projected measure. This is an estimate which can be iterated, and is therefore very useful in some evolutionary PDEs (crowd motion,. . .). We also establish some properties of the Wasserstein projection which are interesting in their own, and allow for instance to prove uniqueness of such a projection in a very general framework.

1. Introduction

Among variational problems involving optimal transportation and Wasserstein distances, a very recurrent one is the following

$$\min_{\varrho\in\mathcal{P}_2(\Omega)} \frac{1}{2}W_2^2(\varrho, g) + \tau F(\varrho)\,, \tag{1.1}$$

where F is a given functional on probability measures, $\tau > 0$ a parameter which can possibly be small, and g is a given probability in $\mathcal{P}_2(\Omega)$ (the space of probability measures on $\Omega \subseteq \mathbb{R}^d$ with finite second moment $\int |x|^2 \,\mathrm{d}\varrho(x) < +\infty$). This very instance of the problem is exactly the one we face in the time-discretization of the gradient flow of F in $\mathcal{P}_2(\Omega)$, where $g = \varrho_k^\tau$ is the measure at step k, and the optimal ϱ will be the next measure ϱ_{k+1}^τ. Under suitable assumptions, at the limit when $\tau \to 0$, this sequence converges to a curve of measures which is the gradient flow of F (see [2, 1] for a general description of this theory).

But the same problem also appears in other frameworks as well, for fixed τ. For instance in image processing, if F is a smoothing functional, this is a model to find a better (smoother) image ϱ which is not so far from the original g (and the choice of the distance W_2 is justified by robustness arguments), see [15]. In urban planning (see [5, 23]) g can represent the distribution of some resources and ϱ that of population, who wants to be close to g but also to guarantee enough space to each individual. In this case the functional F favors diffuse measures, for instance $F(\varrho) = \int h(\varrho(x))\,\mathrm{d}x$ for a convex and superlinear function h, which gives a higher cost to high densities of ϱ (note that in this case, by Jensen's inequality, F is minimized by the uniform measure on Ω). Reciprocally, g could instead represent the distribution of population, and ϱ that of services, to be chosen so that they are close enough to g but more concentrated. In this case F will favor concentrated measures, instead.

When F takes only value 0 and $+\infty$, the above problem becomes a projection problem. Recently, the projection onto the set K_1 of densities bounded above by the constant 1[1] has received lot of attention because of its applications in the time-discretization of evolution problems with density constraints, in particular for crowd motion (see [22, 16], where a crowd is described as a population of particles which cannot overlap, and cannot go beyond a certain threshold density).

In this paper we concentrate on the case where $F(\varrho) = \int h(\varrho)$ for a convex integrand $h : \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$. The case of the projection on K_1 is obtained by taking the following function:

$$h(\varrho) = \begin{cases} 0, & \text{if } 0 \le \varrho \le 1 \\ +\infty, & \text{if } \varrho > 1\,, \end{cases}$$

We are interested in the estimates that one can give on the minimizer $\bar\varrho$, which can be roughly divided into two categories: those which are independent of g but depend on τ, and those which are uniform in τ but refer to similar bounds on g. For instance, by writing down the optimality conditions for (1.1) in the case $F(\varrho) = \int h(\varrho)$, we get $\varphi + \tau h'(\bar\varrho) = const$, where φ is the Kantorovich potential in the transport from $\bar\varrho$ to g (this equality would be true $\bar\varrho-$a.e., but we will forget this detail in this heuristic discussion). On a bounded domain, φ is Lipschitz continuous, and so is $\tau h'(\bar\varrho)$. If h is strictly convex and C^1, this allows to get continuity for $\bar\varrho$, but the bounds obviously degenerate as $\tau \to 0$. On the other hand, they do not really depend on g.

Another bound that one can prove is $\|\bar\varrho\|_{L^\infty} \le \|g\|_{L^\infty}$ (see [7, 23]), which, on the contrary, is independent of τ.

In this paper we are mainly concerned with BV estimates. As we also expect some uniform estimate, we get rid of the parameter τ that we only introduced for the sake of this presentation. We recall that for every function $\varrho \in L^1$ and every open set A the total variation of $\nabla\varrho$ in A is defined as

$$TV(u, A) = \int_A |\nabla\varrho| = \sup\left\{ \int \varrho \,\mathrm{div}\xi \,\mathrm{d}x \quad : \quad \xi \in C^1_c(A), \quad |\xi| \le 1 \right\}.$$

Our main theorem reads as follows:

Theorem 1.1. *Let $\Omega \subset \mathbb{R}^d$ be a (possibly unbounded) convex set, $h : \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$ be a convex function and $g \in \mathcal{P}_2(\Omega) \cap BV(\Omega)$. If $\bar\varrho$ is a minimizer of the following variational problem*

$$\min_{\varrho \in \mathcal{P}_2(\Omega)} \frac{1}{2} W_2^2(\varrho, g) + \int_\Omega h(\varrho(x))\,\mathrm{d}x\,,$$

then

$$\int_\Omega |\nabla\bar\varrho|\,\mathrm{d}x \le \int_\Omega |\nabla g|\,\mathrm{d}x\,.$$

As we said, this covers the case of the Wassertstein projection of g on the subset K_1 of $P_2(\Omega)$ given by the measures with density less than or equal to 1. When dealing with Wasserstein projections we are actually able to establish BV bounds in the more general case in which we project on the set of measures with density less then or equal to a prescribed BV function f. More precisely we have the following theorem

[1]Here and in the sequel we denote by K_f the set f absolutely continuous measure with density bounded by f, i.e.

$$K_f := \{\varrho \in \mathcal{P}(\Omega) \,:\, \varrho \le f\mathrm{d}x\}$$

Theorem 1.2. *Let $\Omega \subset \mathbb{R}^d$ be a (possibly unbounded) convex set, $g \in \mathcal{P}_2(\Omega) \cap BV(\Omega)$ and let $f \in BV_{\text{loc}}(\Omega)$ be a function with*

$$\int_\Omega f \,\mathrm{d}x \geq 1.$$

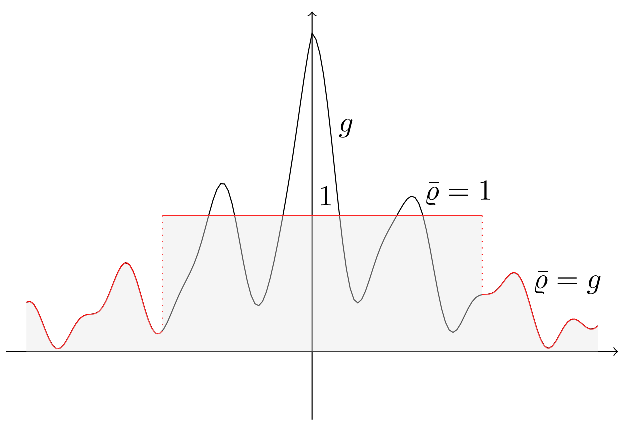
If

$$\bar{\varrho} = \operatorname{argmin}\Big\{ W_2^2(\varrho, g) : \varrho \in \mathcal{P}_2(\Omega), \quad \varrho \leq f \ \ a.e. \Big\}, \tag{1.2}$$

then

$$\int_\Omega |\nabla \bar{\varrho}| \,\mathrm{d}x \leq \int_\Omega |\nabla g| \,\mathrm{d}x + 2 \int_\Omega |\nabla f| \,\mathrm{d}x. \tag{1.3}$$

We would like to spend some words on this BV estimate for the projection, at least in the case of the projection onto K_1, which is the original motivation for this paper. This BV estimate is indeed natural, at least in the 1D case (all the oscillation beyond the level 1 are replaced by a possible jump between a value smaller than 1 and 1, as in the picture below).



For the higher-dimensional case, the situation is trickier, as, for instance, the projection of a measure $g = (1+\varepsilon)\mathbb{1}_{B(0,R)}$ (with $(1+\varepsilon)|B(0,R)| = 1$) is the indicator function of a bigger ball. The total variation involves two opposite effects: the perimeter of the ball increases, but the height of the jump passes from $1+\varepsilon$ to 1. It is not difficult to see that the combination of the two effects is such that the total variation decreases. It is also possible to adapt this argument to the case of a radially symmetric density g, but the general case is not evident.

These kind of BV estimates are useful when the projection is treated as one time-step of a discretized evolution process. For instance, a BV bound allows to transform weak convergence in the sense of measures into strong L^1 convergence (see Section 6.3). Also, if we consider a PDE mixing a smooth evolution, such as the Fokker-Planck evolution, and some projection steps (in order to impose a density constraint, as in crowd motion issues), one could wonder which bounds on the regularity of the solution are preserved in time. From the fact that the discontinuities in the projected measure destroy any kind of $W^{1,p}$ norm, it is natural to look for BV bounds. Notice by the way that, for these kind of applications, proving $\int_\Omega |\nabla \bar{\varrho}| \leq \int_\Omega |\nabla g|$ (with no multiplicative coefficient nor additional term) is crucial in order to iterate this estimate at every step.

The paper is structured as follows: In Section 2 we recall some preliminary results in optimal transportation, in Section 3 we establish our key *mother inequality*, in Section 4 we prove Theorem 1.1 while in Section 5 we collect some properties of solution of (1.2) which can be interesting in their own and we we prove Theorem 1.2. Eventually, in Section 6 we present some applications of the above results, connections with other variational and evolution problems and some open questions.

2. Notations and preliminaries

In this section we collect some facts about optimal transport that we will need in the sequel, referring the reader to [25] for more details. We will denote by $\mathcal{P}(\Omega)$ the set of probability measures in Ω and by $\mathcal{P}_2(\Omega)$ the subset of $\mathcal{P}(\Omega)$ given by those with finite second moment (i.e. $\mu \in \mathcal{P}_2(\Omega)$ if and only if $\int |x|^2 \mathrm{d}\mu < \infty$). We will also use the spaces $\mathcal{M}(\Omega)$ of finite measures on Ω and $L^1_+(\Omega)$ of non-negative functions in L^1. Notice $\{f \in L^1_+(\Omega) : \int f(x)\,\mathrm{d}x = 1\} = L^1_+(\Omega) \cap \mathcal{P}(\Omega)$. In the sequel we will *always* identify an absolutely continuous measure with its density (for instance writing $T_\# f$ for $T_\#(f\mathrm{d}x)$ and so on..).

Theorem 2.1. *Let $\Omega \subset \mathbb{R}^d$ be a given convex set and let $\varrho, g \in L^1_+(\Omega)$ be two probability densities on Ω. Then the following hold:*

(i) *The problem*

$$\frac{1}{2}W_2^2(\varrho, g) := \min\left\{\int_{\Omega\times\Omega} \frac{1}{2}|x-y|^2\,\mathrm{d}\gamma \ : \ \gamma \in \Pi(\varrho, g)\right\}, \tag{2.1}$$

where $\Pi(\varrho, g)$ is the set of the so-called transport plans, i.e. $\Pi(\varrho\mathrm{d}x, g\mathrm{d}x) := \{\gamma \in \mathcal{P}(\Omega \times \Omega) : (\pi^x)_\#\gamma = \varrho, (\pi^y)_\#\gamma = g\}$, has a unique solution, which is of the form $\gamma_{\hat{T}} := (id, \hat{T})_\#\varrho$, and $\hat{T} : \Omega \to \Omega$ is a solution of the problem

$$\min_{T_\#\varrho = g} \int_\Omega \frac{1}{2}|x - T(x)|^2\,\varrho(x)\,\mathrm{d}x\,. \tag{2.2}$$

(ii) *The map $\hat{T} : \{\varrho > 0\} \to \{g > 0\}$ is a.e. invertible and its inverse $\hat{S} := \hat{T}^{-1}$ is a solution of the problem*

$$\min_{S_\# g = \varrho} \int_\Omega \frac{1}{2}|x - S(x)|^2\,g(x)\,\mathrm{d}x. \tag{2.3}$$

(iii) *$W_2(\cdot,\cdot)$ is a distance on the space $\mathcal{P}_2(\Omega)$ of probabilities over Ω with finite second moment.*

(iv) *We have*

$$\frac{1}{2}W_2^2(\varrho, g) = \max\left\{\int_\Omega \varphi(x)\varrho(x)\,\mathrm{d}x + \int_\Omega \psi(y)g(y)\,\mathrm{d}y \ : \ \varphi(x) + \psi(y) \le \frac{1}{2}|x-y|^2, \ \forall x, y \in \Omega\right\}. \tag{2.4}$$

(v) *The optimal functions $\hat{\varphi}, \hat{\psi}$ in (2.4) are continuous, differentiable almost everywhere, Lipschitz if Ω is bounded, and such that:*
- $\hat{T}(x) = x - \nabla\hat{\varphi}(x)$ *and* $\hat{S}(x) = x - \nabla\hat{\psi}(x)$ *for a.e.* $x \in \Omega$; *in particular, the gradients of the optimal functions are uniquely determined (even in case of non-uniqueness of $\hat{\varphi}$ and $\hat{\psi}$) a.e. on $\{\varrho > 0\}$ and $\{g > 0\}$, respectively;*
- *the functions*

$$x \mapsto \frac{|x|^2}{2} - \hat{\varphi}(x) \quad \text{and} \quad x \mapsto \frac{|x|^2}{2} - \hat{\psi}(x),$$

are convex in Ω and hence $\hat{\varphi}$ and $\hat{\psi}$ are semi-concave;
- $\hat{\varphi}(x) = \max_{y\in\Omega}\left\{\frac{1}{2}|x-y|^2 - \hat{\psi}(y)\right\}$ *and* $\hat{\psi}(y) = \max_{x\in\Omega}\left\{\frac{1}{2}|x-y|^2 - \hat{\varphi}(x)\right\}$;

- *if we denote by χ^c the so-called c−transform of a function $\chi : \Omega \to \mathbb{R}$ defined through $\chi^c(y) = \inf_{x\in\Omega} \frac{1}{2}|x-y|^2 - \chi(x)$, then the maximal value in (2.4) is also equal to*

$$\max\left\{\int_\Omega \varphi(x)\varrho(x)\,\mathrm{d}x + \int_\Omega \varphi^c(y)g(y)\,\mathrm{d}y,\ \varphi\in C^0(\Omega)\right\} \tag{2.5}$$

and the optimal φ is the same $\hat\varphi$ as above, and is such that $\hat\varphi = (\hat\varphi^c)^c$ a.e. on $\{\varrho > 0\}$.

(vi) *The functional $W : \mathcal{M}(\Omega) \to \mathbb{R}\cup\{+\infty\}$ defined through*

$$W(\varrho) = \max\left\{\int_\Omega \varphi(x)\varrho(x)\,\mathrm{d}x + \int_\Omega \varphi^c(y)g(y)\,\mathrm{d}y,\ \varphi\in C^0(\Omega)\right\} = \begin{cases}\frac{1}{2}W_2^2(\varrho,g), & \text{if } \varrho\in\mathcal{P}_2(\Omega)\\ +\infty, & \text{otherwise,}\end{cases}$$

is convex and its subdifferential is given by

$$\partial W(\varrho) = \left\{\varphi\in C^0(\Omega)\,\text{which are optimal in (2.5)}\right\}.$$

The only non-standard point is the last one (the computation of the sub-differential of W): it is sketched in [5], and a more detailed presentation will be part of [24].

We also need some regularity results on optimal transport maps, see [8, 9].

Theorem 2.2. *Let $\Omega\subset\mathbb{R}^d$ be a bounded strictly convex set with smooth boundary and let $\varrho, g\in L^1_+(\Omega)$ be two probability densities on Ω away from zero and infinity*[2]*. Then, using the notations from Theorem 2.1, we have:*

(i) $\hat T\in C^{0,\alpha}(\Omega)$ *and* $\hat S\in C^{0,\alpha}(\Omega)$.

(ii) *If $\varrho\in C^{k,\beta}(\Omega)$ and $g\in C^{k,\beta}(\Omega)$, then $\hat T\in C^{k+1,\beta}(\Omega)$ and $\hat S\in C^{k+1,\beta}(\Omega)$.*

Most of our proofs will be done by approximation. To do this, we need a stability result

Theorem 2.3. *Let $\Omega\subset\mathbb{R}^d$ be a bounded convex set and let $\varrho_n\in L^1_+(\Omega)$ and $g_n\in L^1_+(\Omega)$ be two sequences of probability densities in Ω. Then, using the notations from Theorem 2.1, if $\varrho_n \rightharpoonup \varrho$ and $g_n\rightharpoonup g$ weakly as measures, then we have:*

(i) $W_2(\varrho,g) = \lim_{n\to\infty} W_2(\varrho_n, g_n)$.

(ii) *there exist two semi-concave functions φ,ψ such that $\nabla\hat\varphi_n\to\nabla\varphi$ and $\nabla\hat\psi_n\to\nabla\psi$ a.e. and $\nabla\varphi = \nabla\hat\varphi$ a.e. on $\{\varrho>0\}$ and $\nabla\psi=\nabla\hat\psi$ a.e. on $\{g>0\}$.*

If Ω is unbounded (for instance $\Omega=\mathbb{R}^d$), then the convergence $\varrho_n\rightharpoonup\varrho$ and $g_n\rightharpoonup g$ weakly as measures is not enough to guarantee (i) *but only implies $W_2(\varrho,g)\le\liminf_{n\to\infty} W_2(\varrho_n,g_n)$. Yet,* (i) *is satisfied if $W_2(\varrho_n,\varrho), W_2(g_n,g)\to 0$, which is a stronger condition.*

Proof. The proof of *(i)* can be found in [25]. We prove *(ii)*. (Actually this is a consequence of the Theorem 3.3.3. from [10], but for the sake of completeness we sketch its simple proof).

We first note that due to Theorem 2.1 *(v)* the sequences $\hat\varphi_n$ and $\hat\psi_n$ are equi-continuous. Moreover, since the Kantorovich potentials are uniquely determined up to a constant we may suppose that there is $x_0\in\Omega$ such that $\hat\varphi_n(x_0) = \hat\psi_n(x_0) = 0$ for every $n\in\mathbb{N}$. Thus, $\hat\varphi_n$ and $\hat\psi_n$ are locally uniformly bounded in Ω and, by the Ascoli-Arzelà Theorem, they converge uniformly up to a subsequence

$$\hat\varphi_n \xrightarrow[n\to\infty]{} \varphi_\infty \qquad\text{and}\qquad \hat\psi_n \xrightarrow[n\to\infty]{} \psi_\infty,$$

[2] We say that ϱ and g are away from zero and infinity if there is some $\varepsilon>0$ such that $\varepsilon\le\varrho\le 1/\varepsilon$ and $\varepsilon\le g\le 1/\varepsilon$ a.e. in Ω.

to some continuous functions $\varphi_\infty, \psi_\infty \in C(\Omega)$, satisfying

$$\varphi_\infty(x) + \psi_\infty(y) \leq \frac{1}{2}|x-y|^2, \quad \text{for every} \quad x, y \in \Omega.$$

In order to show that φ_∞ and ψ_∞ are precisely Kantorovich potentials, we use the characterization of the potentials as solutions to the problem (2.4). Indeed, let φ and ψ be such that $\varphi(x) + \psi(y) \leq \frac{1}{2}|x-y|^2$ for every $x, y \in \Omega$. Then, for every $n \in \mathbb{N}$ we have

$$\int_\Omega \hat{\varphi}_n(x)\varrho_n(x)\,\mathrm{d}x + \int_\Omega \hat{\psi}_n(y)g_n(y)\,\mathrm{d}y \geq \int_\Omega \varphi(x)\varrho_n(x)\,\mathrm{d}x + \int_\Omega \psi(y)g_n(y)\,\mathrm{d}y,$$

and passing to the limit we obtain

$$\int_\Omega \varphi_\infty(x)\varrho(x)\,\mathrm{d}x + \int_\Omega \psi_\infty(y)g(y)\,\mathrm{d}y \geq \int_\Omega \varphi(x)\varrho(x)\,\mathrm{d}x + \int_\Omega \psi(y)g(y)\,\mathrm{d}y,$$

which proves that φ_∞ and ψ_∞ are optimal. In particular, the gradient of these functions coincide with those of $\hat{\varphi}$ and $\hat{\psi}$ on the sets where the densities are strictly positive.

We now prove that $\nabla\hat{\varphi}_n \to \nabla\varphi_\infty$ a.e. in Ω. We denote with $\mathcal{N} \subset \Omega$ the set of points $x \in \Omega$, such that there is a function among $\hat{\varphi}$ and $\hat{\varphi}_n$, for $n \in \mathbb{N}$, which is not differentiable at x. We note that by Theorem 2.1 *(v)* the set $\mathcal{N}$ has Lebesgue measure zero. Let now $x_0 \in \Omega \setminus \mathcal{N}$ and suppose, without loss of generality, $x_0 = 0$. Setting

$$\alpha_n(x) := \frac{|x|^2}{2} - \hat{\varphi}_n(x) + \hat{\varphi}_n(0) + x \cdot \nabla\varphi_\infty(0) \quad \text{and} \quad \alpha(x) := \frac{|x|^2}{2} - \varphi_\infty(x) + \varphi_\infty(0) + x \cdot \nabla\varphi_\infty(0),$$

we have that α_n are all convex and such that $\alpha_n(0) = 0$, and hence $\alpha_n(x) \geq \nabla\alpha_n(0)\cdot x$. Moreover, $\alpha_n \to \alpha$ locally uniformly and $\nabla\alpha(0) = 0$. Suppose by contradiction that $\lim_{n\to\infty} \nabla\alpha_n(0) \neq 0$. Then, there is a unit vector $p \in \mathbb{R}^d$ and a constant $\delta > 0$ such that, up to a subsequence, $p \cdot \nabla\alpha_n \geq \delta$ for every $n > 0$. Then, for every $t > 0$ we have

$$\frac{\alpha(pt)}{t} = \lim_{n\to\infty} \frac{\alpha_n(pt)}{t} \geq \liminf_{n\to\infty} \left\{p \cdot \nabla\alpha_n(0)\right\} \geq \delta,$$

which is a contradiction with the fact that $\nabla\alpha(0) = 0$. □

In order to handle our approximation procedures, we also need to spend some words on the notion of $\Gamma-$*convergence* (see [11]).

Definition 2.1. *On a metric space X let $F_n : X \to \mathbb{R} \cup \{+\infty\}$ be a sequence of functions. We define the two lower-semicontinuous functions F^- and F^+ (called $\Gamma-\liminf$ and $\Gamma-\limsup$ of this sequence, respectively) by*

$$F^-(x) := \inf\{\liminf_{n\to\infty} F_n(x_n) \ : \ x_n \to x\},$$
$$F^+(x) := \inf\{\limsup_{n\to\infty} F_n(x_n) \ : \ x_n \to x\}.$$

Should F^- and F^+ coincide, then we say that F_n actually $\Gamma-$converges to the common value $F = F^- = F^+$.

This means that, when one wants to prove $\Gamma-$convergence of F_n towards a given functional F, one has actually to prove two distinct facts: first we need $F^- \geq F$ (i.e. we need to prove $\liminf_n F_n(x_n) \geq F(x)$ for any approximating sequence $x_n \to x$; not only, it is sufficient to prove it when $F_n(x_n)$ is bounded) and then $F^+ \leq F$ (i.e. we need to find a *recovery sequence* $x_n \to x$

such that $\limsup_n F_n(x_n) \le F(x)$). The definition of $\Gamma-$convergence for a continuous parameter $\varepsilon \to 0$ obviously passes through the convergence to the same limit for any subsequence $\varepsilon_n \to 0$.

Among the properties of $\Gamma-$convergence we have the following:

- if there exists a compact set $K \subset X$ such that $\inf_X F_n = \inf_K F_n$ for any n, then F attains its minimum and $\inf F_n \to \min F$,
- if $(x_n)_n$ is a sequence of minimizers for F_n admitting a subsequence converging to x, then x minimizes F (in particular, if F has a unique minimizer x and the sequence of minimizers x_n is compact, then $x_n \to x$),
- if F_n is a sequence $\Gamma-$converging to F, then $F_n + G$ will $\Gamma-$converge to $F + G$ for any continuous function $G : X \to \mathbb{R} \cup \{+\infty\}$.

In the sequel we will need the following two easy criteria to guarantee $\Gamma-$convergence.

Proposition 2.4. *If each F_n is l.s.c. and $F_n \to F$ uniformly, then F_n $\Gamma-$converges to F.*

If each F_n is l.s.c., $F_n \le F_{n+1}$ and $F(x) = \lim_n F_n(x)$ for all x, then F_n $\Gamma-$converges to F.

We will essentially apply the notion of $\Gamma-$convergence in the space $X = \mathcal{P}(\Omega)$ endowed with the weak convergence[3] (which is indeed metrizable on this bounded subset of the Banach space of measures) since the space $\mathcal{P}_2(\Omega)$ endowed with the W_2 convergence lacks compactness whenever Ω is not compact.

We conclude this section with the following simple lemma concerning properties of the functional

$$\mathcal{M}(\Omega) \ni \varrho \mapsto H(\varrho) = \begin{cases} \int_\Omega h(\varrho(x))\,\mathrm{d}x, & \text{if } \varrho \ll \mathrm{d}x, \\ +\infty, & \text{otherwise.} \end{cases}$$

Lemma 2.5. *Let Ω be an open set and $h : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be convex and superlinear at $+\infty$, then the functional $H : \mathcal{M}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ is convex and lower semicontinuous with respect to the weak convergence of measures. Moreover if $h \in C^1$ then we have*

$$\lim_{\varepsilon \to 0} \frac{H(\varrho + \varepsilon\chi) - H(\varrho)}{\varepsilon} = \int h'(\varrho)\,\mathrm{d}\chi$$

whenever $\rho, \chi \ll \mathrm{d}x$, $H(\varrho) < +\infty$ and $H(\varrho + \varepsilon\chi) < +\infty$ at least for small ε. As a consequence, $h'(\varrho)$ is the first variation of H and the we have

$$\partial H(\varrho) = \{h'(\varrho)\}.$$

For this classical fact, and in particular for the semicontinuity, we refer to [4] and [3].

3. The "mother" inequality

In this section we establish the key inequality needed in the proof of Theorems 1.1 and 1.2.

---

[3]We say that a family of probability measure μ_n weakly converges to a probability measure μ in Ω if

$$\int \varphi, \mathrm{d}\mu_n \to \int \varphi\,\mathrm{d}\mu \qquad \forall \varphi \in C_b(\Omega)\,,$$

where $C_b(\Omega)$ is the space of continuous and bounded functions on Ω

Lemma 3.1. *Suppose that $\varrho, g \in L^1_+$ are smooth probability densities, which are away from 0 and infinity, and let $H \in C^2(\Omega)$ be a convex function. Then we have the following inequality*

$$\int_\Omega \Big(\varrho\, \nabla \cdot \big[\nabla H(\nabla\varphi)\big] - g\, \nabla \cdot \big[\nabla H(-\nabla\psi)\big]\Big)\, \mathrm{d}x \le 0, \tag{3.1}$$

where φ and ψ are the corresponding Kantorovich potentials.

Proof. We first note that since ϱ and g are smooth and away from zero and infinity in Ω, Theorem 2.2 implies that φ, ψ are smooth as well. Now using the identity $S(T(x)) \equiv x$ and that $S_{\#} g = \varrho$ we get

$$\begin{aligned}
\int_\Omega \varrho(x)\, \nabla \cdot \big[\nabla H(\nabla\varphi(x))\big]\, \mathrm{d}x &= \int_\Omega g(x) \Big[\nabla \cdot \big[\nabla H(\nabla\varphi)\big]\Big](S(x))\, \mathrm{d}x \\
&= \int_\Omega g(x)\, \nabla \cdot \Big[\nabla H\big(\nabla\varphi \circ S\big)\Big](x)\, \mathrm{d}x \\
&\quad + \int_\Omega g(x) \Big(\Big[\nabla \cdot \big[\nabla H(\nabla\varphi)\big]\Big](S(x)) - \nabla \cdot \big[\nabla H\big(\nabla\varphi \circ S\big)\big](x)\Big)\, \mathrm{d}x,
\end{aligned}$$

and, by the equality

$$-\nabla\psi(x) = S(x) - x = S(x) - T(S(x)) = \nabla\varphi(S(x)),$$

we obtain

$$\begin{aligned}
&\int_\Omega \Big(\varrho\, \nabla \cdot \big[\nabla H(\nabla\varphi)\big] - g\, \nabla \cdot \big[\nabla H(-\nabla\psi)\big]\Big)\, \mathrm{d}x = \\
&\qquad = \int_\Omega g(x) \Big(\Big[\nabla \cdot \big[\nabla H(\nabla\varphi)\big]\Big](S(x)) - \nabla \cdot \big[\nabla H\big(\nabla\varphi \circ S\big)\big](x)\Big)\, \mathrm{d}x \qquad (3.2)\\
&\qquad = \int_\Omega \varrho(x) \Big(\nabla \cdot \big[\nabla H(\nabla\varphi)\big] - \Big[\nabla \cdot \big[\nabla H\big(\nabla\varphi\big) \circ S\big]\Big] \circ T\Big)\, \mathrm{d}x,
\end{aligned}$$

For simplicity we set

$$\begin{aligned}
E &= \nabla \cdot (\nabla H(\nabla\varphi)) - \big[\nabla \cdot (\nabla H(\nabla\varphi) \circ S)\,\big] \circ T \\
&= \nabla \cdot \xi - \big[\nabla \cdot (\xi \circ S)\big] \circ T,
\end{aligned} \tag{3.3}$$

where by ξ we denote the continuously differentiable function

$$\xi(x) = (\xi^1, \dots, \xi^d) := \nabla H(\nabla\varphi(x)),$$

whose derivative is given by

$$D\xi = D\big(\nabla H(\nabla\varphi)\big) = D^2 H(\nabla\varphi) \cdot D^2\varphi.$$

We now calculate

$$\begin{aligned}
\big[\nabla \cdot (\xi \circ S)\big] \circ T &= \sum_{i=1}^d \frac{\partial(\xi^i \circ S)}{\partial x^i} \circ T = \sum_{i=1}^d \sum_{j=1}^d \frac{\partial \xi^i}{\partial x^j}(S(T)) \frac{\partial S^j}{\partial x^i} \circ T \\
&= \mathrm{tr}\left(D\xi \cdot (DT)^{-1}\right) = \mathrm{tr}\left(D^2 H(\nabla\varphi) \cdot D^2\varphi \cdot (I_d - D^2\varphi)^{-1}\right),
\end{aligned} \tag{3.4}$$

where the last two equality follow by $DS \circ T = (DT)^{-1}$ and we also used that $(DT)^{-1} = (I_d - D^2\varphi)^{-1}$, where I_d is the d-dimensional identity matrix.

By (3.3) and (3.4) we have that

$$E = \mathrm{tr}\big[D^2 H(\nabla\varphi) \cdot D^2\varphi \cdot \big(I_d - (I_d - D^2\varphi)^{-1}\big)\,\big]$$
$$= -\mathrm{tr}\big[D^2 H(\nabla\varphi) \cdot \big[D^2\varphi\big]^2 \cdot (I_d - D^2\varphi)^{-1}\big].$$

Since we have that

$$I_d - D^2\varphi \geq 0,$$

and that the trace of the product of two positive matrices is positive, we obtain $E \leq 0$, which together with (3.2) concludes the proof. □

Lemma 3.2. *Let $\Omega \subset \mathbb{R}^d$ be bounded and convex, $\varrho, g \in W^{1,1}(\Omega)$ be two probability densities and $H \in C^2(\mathbb{R}^d)$ be a radially symmetric convex function. Then the following inequality holds*

$$\int_\Omega \Big(\nabla\varrho \cdot \nabla H(\nabla\varphi) + \nabla g \cdot \nabla H(\nabla\psi)\Big)\,\mathrm{d}x \geq 0, \tag{3.5}$$

where φ and ψ are the corresponding Kantorovich potentials.

Proof. Let us start noticing that due to the radial symmetry of H

$$\nabla H(\nabla\psi) = -\nabla H(-\nabla\psi). \tag{3.6}$$

Step 1. Proof in the smooth case. Suppose that the probability densities ϱ and g are smooth and bounded away from zero and infinity. As in Lemma 3.1, we note that under these assumption on ϱ and g the Kantorovich potentials are smooth, hence after integration by part the left hand side of (3.5) becomes

$$\begin{aligned}\int_\Omega \Big(\nabla\varrho \cdot \nabla H(\nabla\varphi) + \nabla g \cdot \nabla H(\nabla\psi)\Big)\,\mathrm{d}x \;&= \int_{\partial\Omega} \Big(\varrho\,\nabla H(\nabla\varphi)\cdot n + g\,\nabla H(\nabla\psi)\cdot n\Big)\,\mathrm{d}\mathcal{H}^{d-1}\\ &\quad - \int_\Omega \Big(\varrho\,\nabla\cdot\big[\nabla H(\nabla\varphi)\big] + g\,\nabla\cdot\big[\nabla H(\nabla\psi)\big]\Big)\,\mathrm{d}x\\ &\geq \int_{\partial\Omega} \Big(\varrho\,\nabla H(\nabla\varphi) + g\,\nabla H(\nabla\psi)\Big)\cdot n\,\mathrm{d}\mathcal{H}^{d-1},\end{aligned}$$

where we used Lemma 3.1 and (3.6). Moreover, by the radial symmetry of H gives that $\nabla H(z) = c(z)z$, for some $c(z) > 0$. Since the gradients of the Kantorovich potentials $\nabla\varphi$ and $\nabla\psi$ calculated in boundary points are pointing outward Ω (since $T(x) = x - \nabla\varphi(x) \in \Omega$, and $S(x) = x - \nabla\psi(x) \in \Omega$) we have that

$$\nabla H(\nabla\varphi(x))\cdot n(x) \geq 0 \qquad \text{and} \qquad \nabla H(\nabla\psi(x))\cdot n(x) \geq 0, \qquad \forall x \in \partial\Omega,$$

which concludes the proof of (3.5) if ϱ and g are smooth.

Step 2. Proof for generic $\varrho, g \in W^{1,1}(\Omega)$. We first note that for every $\varepsilon > 0$ there are smooth nonnegative functions $\varrho_\varepsilon \in C^1(\overline{\Omega})$ and $g_\varepsilon \in C^1(\overline{\Omega})$ such that

$$\varrho_\varepsilon \xrightarrow[\varepsilon\to 0]{W^{1,1}(\Omega)} \varrho \qquad \text{and} \qquad g_\varepsilon \xrightarrow[\varepsilon\to 0]{W^{1,1}(\Omega)} g.$$

Moreover, by adding a positive constant and then multiplying by another one, we may assume that ϱ_ε and g_ε are probability densities away from zero:

$$\varrho_\varepsilon \geq \varepsilon^2, \qquad g_\varepsilon \geq \varepsilon^2 \qquad \text{and} \qquad \int_\Omega \varrho_\varepsilon \,\mathrm{d}x = \int_\Omega g_\varepsilon \,\mathrm{d}x = 1.$$

Let $\varphi_\varepsilon \in C^{2,\beta}(\overline{\Omega})$ and $\psi_\varepsilon \in C^{2,\beta}(\overline{\Omega})$ be the Kantorovich potentials corresponding to the optimal transport maps between ϱ_ε and g_ε. By *Step 1* we have

$$\int_\Omega \Big(\nabla\varrho_\varepsilon \cdot \nabla H(\nabla\varphi_\varepsilon) + \nabla g_\varepsilon \cdot \nabla H(\nabla\psi_\varepsilon)\Big)\,\mathrm{d}x \geq 0. \tag{3.7}$$

On the other hand by Theorem 2.3 and by the fact that Ω is bounded we have that

$$|\nabla\varphi_\varepsilon|, |\nabla\psi_\varepsilon| \leq C, \qquad \nabla\varphi_\varepsilon \xrightarrow[\varepsilon\to 0]{a.e.} \nabla\varphi \qquad \text{and} \qquad \nabla\psi_\varepsilon \xrightarrow[\varepsilon\to 0]{a.e.} \nabla\psi,$$

and so passing to the limit as $\varepsilon \to 0$ in (3.7) (by dominated convergence, since ∇H is locally bounded and we can suppose that the convergence $\varrho_\varepsilon \to \varrho$ and $g_\varepsilon \to g$ holds a.e. and is dominated) we obtain (3.5), which concludes the proof. □

Remark 3.1. In Lemma 3.2 we can drop the convexity assumption on Ω if ϱ, g have compact support: indeed, it is enough to choose a ball $\Omega' \supset \Omega$ containing the supports of ϱ and g.

Remark 3.2. In Lemma 3.2 also remains true in the case of compactly supported densities g and ϱ, even if we drop the assumption on H, that $H(z) = H(|z|)$. In this case the inequality becomes

$$\int_{\mathbb{R}^d} \Big(\nabla\varrho \cdot \nabla H(\nabla\varphi) - \nabla g \cdot \nabla H(-\nabla\psi)\Big)\,\mathrm{d}x \geq 0.$$

Proof. The proof follows the same scheme of that of Lemma 3.2, first in the smooth case and then for approximation. We select a convex domain Ω large enough to contain the supports of ϱ and g in its interior: all the integrations and integration by parts are performed on Ω. The only difficulty is that we cannot guarantee the boundary term to be positive. Yet, we first take ϱ, g to be smooth and we approximate them by taking $\varrho_\varepsilon := \varepsilon\frac{1}{|\Omega|} + (1-\varepsilon)\varrho$ and $g_\varepsilon := \varepsilon\frac{1}{|\Omega|} + (1-\varepsilon)g$. For these densities and their corresponding potentials $\varphi_\varepsilon, \psi_\varepsilon$, we obtain the inequality

$$\int_\Omega \Big(\nabla\varrho_\varepsilon \cdot \nabla H(\nabla\varphi_\varepsilon) + \nabla g_\varepsilon \cdot \nabla H(\nabla\psi_\varepsilon)\Big)\,\mathrm{d}x \geq \int_{\partial\Omega} \Big(\varrho_\varepsilon\,\nabla H(\nabla\varphi_\varepsilon) + g_\varepsilon\,\nabla H(\nabla\psi_\varepsilon)\Big)\cdot n\,\mathrm{d}\mathcal{H}^{d-1}.$$

We can pass to the limit (by dominated convergence as before) in this inequality, and notice that the r.h.s. tends to 0, since $|\nabla H(\nabla\varphi_\varepsilon)|, |\nabla H(\nabla\psi_\varepsilon)| \leq C$ and $\varrho_\varepsilon = g_\varepsilon = \varepsilon/|\Omega|$ on $\partial\Omega$. Once the inequality is proven for smooth ϱ, g, a new approximation gives the desired result. □

By approximating $H(z) = |z|$ with $H(z) = \sqrt{\varepsilon^2 + |z|^2}$, Lemma 3.2 has the following important corollary, where we use the convention $\dfrac{z}{|z|} = 0$ for $z = 0$.

Corollary 3.3. *Let $\Omega \subset \mathbb{R}^d$ be a given bounded convex set and $\varrho, g \in W^{1,1}(\Omega)$ be two probability densities. Then the following inequality holds*

$$\int_\Omega \Big(\nabla\varrho \cdot \frac{\nabla\varphi}{|\nabla\varphi|} + \nabla g \cdot \frac{\nabla\psi}{|\nabla\psi|}\Big)\,\mathrm{d}x \geq 0, \tag{3.8}$$

where φ and ψ are the corresponding Kantorovich potentials.

4. BV estimates for minimizers

In this section we prove Theorem 1.1. Since we will need to perform several approximation arguments, and we want to use $\Gamma-$convergence, we need to provide uniqueness of the minimizers.

Lemma 4.1. *Let $g \in \mathcal{P}(\Omega) \cap L^1_+(\Omega)$, then the functional $\mu \mapsto W_2^2(\mu, g)$ is strictly convex on $\mathcal{P}_2(\Omega)$.*

Proof. Suppose by contradiction that $\mu_0 \neq \mu_1$ and $t \in]0,1[$ are such that

$$W_2^2(\mu_t, g) = (1-t)W_2^2(\mu_0, g) + tW_2^2(\mu_1, g),$$

where $\mu_t = (1-t)\mu_0 + t\mu_1$. Let γ_0 be the optimal transport plan in the transport from μ_0 to g (pay attention to the direction: it is a transport map if we see it backward: *from g to μ_0*). As the starting measure is absolutely continuous, by Brenier's Theorem, γ_0 is of the form $(T_0, id)_{\#}g$. Analogously, take $\gamma_1 = (T_1, id)_{\#}g$ optimal from μ_1 to g. Set $\gamma_t := (1-t)\gamma_0 + t\gamma_1 \in \Pi(\mu_t, g)$. We have

$$(1-t)W_2^2(\mu_0, g) + tW_2^2(\mu_1, g) = W_2^2(\mu_t, g) \leq \int |x-y|^2 \,\mathrm{d}\gamma_t = (1-t)\int |x-y|^2 \,\mathrm{d}\gamma_0 + t\int |x-y|^2 \,\mathrm{d}\gamma_1$$
$$= (1-t)W_2^2(\varrho_0, g) + tW_2^2(\varrho_1, g),$$

which implies that γ_t is actually optimal in the transport from g to μ_t. Yet γ_t is not induced from a transport map, unless $T_0 = T_1$ a.e. on $\{g > 0\}$. This is a contradiction with $\mu_0 \neq \mu_1$ and proves strict convexity. □

Let us denote by $\mathcal{C}$ the class of convex l.s.c. function $h : \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$, finite in a neighborhood of 0 and with finite right derivative $h'(0)$ at 0, and superlinear at $+\infty$.

Lemma 4.2. *If $h \in \mathcal{C}$ there exists a sequence of C^2 convex functions h_n, superlinear at ∞, with $h_n'' > 0$, $h_n \leq h_{n+1}$ and $h(x) = \lim_n h_n(x)$ for every $x \in \mathbb{R}_+$.*

Moreover, if $h : \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$ is a convex l.s.c. superlinear function, there exists a sequence of functions $h_n \in \mathcal{C}$ with $h_n \leq h_{n+1}$ and $h(x) = \lim_n h_n(x)$ for every $x \in \mathbb{R}_+$.

Proof. Let us start from the case $h \in \mathcal{C}$. Set $\ell^+ := \sup\{x : h(x) < +\infty\} \in \mathbb{R}_+ \cup \{+\infty\}$. Let us define an increasing function $\xi_n : \mathbb{R} \to \mathbb{R}$ in the following way:

$$\xi_n(x) := \begin{cases} h'(0) & \text{for } x \in]-\infty, 0] \\ h'(x) & \text{for } x \in [0, \ell^+ - \frac{1}{n}] \\ h'(\ell^+ - \frac{1}{n}) & \text{for } \ell^+ - \frac{1}{n} \leq x < \ell^+, \\ h'(\ell^+ - \frac{1}{n}) + n(x - \ell^+) & \text{for } x \geq \ell^+, \end{cases}$$

where, if the derivative if h does not exist somewhere, we just replace it with the right derivative. (Notice that when $\ell^+ = +\infty$, the last two cases do not apply).

Let $q \geq 0$ be a C^1 function with $\mathrm{spt}(q) \subset [-1, 0]$, $\int q(t)\,\mathrm{d}t = 1$ and let us set $q_n(t) = nq(nt)$. We define h_n as the primitive of the C^1 function

$$h_n'(x) := \int \left(\xi_n(t) - \frac{1}{n}e^{-t}\right) q_n(t-x)\,\mathrm{d}t,$$

with $h_n(0) = h(0)$. It is easy to check that all the required properties are satisfied: we have $h_n''(x) \geq \frac{1}{n}e^{-x}$, h_n is superlinear because $\lim_{x\to\infty} \xi_n(x) = +\infty$, and we have increasing convergence $h_n \to h$.

For the case of a generic function h, it is possible to approximate it with functions in $\mathcal{C}$ if we define $\ell^- := \inf\{x : h(x) < +\infty\} \in \mathbb{R}_+$ and take

$$h_n(x) = \begin{cases} h(\ell^- + \frac{1}{n}) + h'(\ell^- + \frac{1}{n})(x - \ell^- - \frac{1}{n}) + n|x - \ell^-| & \text{for } x \le \ell^- \\ h(\ell^- + \frac{1}{n}) + h'(\ell^- + \frac{1}{n})(x - \ell^- - \frac{1}{n}) & \text{for } x \in]\ell^-, \ell^- + \frac{1}{n}] \\ h(x) & \text{for } x \ge \ell^- + \frac{1}{n}. \end{cases}$$

In this case as well, it is easy to check that all the required properties are satisfied. □

Proof of Theorem 1.1.

Proof. Let us start from the case where g is $W^{1,1}$ and h is C^2, superlinear, with $h'' > 0$, and Ω is a bounded convex set. A minimizer $\bar\varrho$ exists (by semicontinuity of the criterion and compactness of $\mathcal{P}_2(\Omega)$). Thanks to Theorem 2.1 (vi) and Lemma 2.5, the optimality conditions are of the following there exists a Kantorovich potential φ for the transport from $\bar\varrho$ to g such that $0 = \varphi + h'(\bar\varrho)$. This shows that $h'(\bar\varrho)$ is Lipschitz continuous. Hence, $\bar\varrho$ is bounded. On bounded sets h' is a diffeomorphism with Lipschitz inverse, thanks to $h'' > 0$, which proves that $\bar\varrho$ itself is Lipschitz. Then we can apply Corollary 3.3 and get

$$\int_\Omega \Big(\nabla\bar\varrho \cdot \frac{\nabla\varphi}{|\nabla\varphi|} + \nabla g \cdot \frac{\nabla\psi}{|\nabla\psi|}\Big)\,\mathrm{d}x \ge 0.$$

Yet, from $\varphi = -h'(\bar\varrho)$ and $h'' > 0$, we get that $\nabla\varphi$ and $\nabla\bar\varrho$ are vectors with opposite directions. Hence we have

$$\int_\Omega |\nabla\bar\varrho|\,\mathrm{d}x \le \int_\Omega \nabla g \cdot \frac{\nabla\psi}{|\nabla\psi|}\,\mathrm{d}x \le \int_\Omega |\nabla g|\,\mathrm{d}x,$$

which is the desired estimate.

We can generalize to $h \in \mathcal{C}$ by using the previous lemma and approximating it with a sequence h_n. Thanks to monotone convergence we have $\Gamma-$convergence for the minimization problem that we consider. We also have compactness since $\mathcal{P}_2(\Omega)$ is compact, and uniqueness of the minimizer. Hence, the minimizers $\bar\varrho_n$ corresponding to h_n satisfy $\int_\Omega |\nabla\bar\varrho_n| \le \int_\Omega |\nabla g|$ and converge to the minimizer $\bar\varrho$ corresponding to h. By the semicontinuity of the total variation we conclude the proof in this case.

Similarly, we can generalize to other convex functions h, approximating them with functions in $\mathcal{C}$ (notice that this is only interesting if the function h allows the existence of at least a probability density with finite cost, i.e. if $h(1/|\Omega|) < +\infty$). Also, we can take $g \in BV$ and approximate it with $W^{1,1}$ functions. If the approximation is done for instance by convolution, then we have a sequence with $W_2(g_n, g) \to 0$, which guarantees uniform convergence of the functionals, and hence $\Gamma-$convergence.

We can also handle the case $\Omega = \mathbb{R}^d$, by first taking g to be compactly supported and $h \in \mathcal{C}$. In this case the same arguments as above hold, since the optimality condition $0 = \varphi + h'(\bar\varrho)$ imposes that $\bar\varrho$ is compactly supported. Indeed, on $\{\bar\varrho > 0\}$ we have $\varphi = \psi^c$, where ψ is the Kantorovich potential defined on $\mathrm{spt}(g)$, which is bounded. Hence φ grows at infinity quadratically, from $\varphi(x) = \inf_{y\in\mathrm{spt}(g)} \frac{1}{2}|x-y|^2 - \psi(y)$ and h' is bounded from below. As a consequence, it is not possible to have points with $\bar\varrho > 0$ too far. Once we know that the densities are compactly supported, the same arguments as above apply. Then one passes to the limit obtaining the result for any generic convex function h, and then we can also approximate g (as above, we select a sequence g_n of compactly supported densities converging to g in W_2). Notice that in this case the convergence is no more uniform on $\mathcal{P}_2(\Omega)$, but it is uniform on a bounded set $W_2(\varrho, g) \leq C$ which is the only one interesting in the minimization. □

5. Projected measures under density constraints

5.1. **Existence, uniqueness, characterization, stability of the projected measure.** In this section we will take $\Omega \subset \mathbb{R}^d$ be a given closed set with negligible boundary, $f : \Omega \to [0, +\infty[$ a measurable function in $L^1_{\text{loc}}(\Omega)$ with $\int_\Omega f \, \mathrm{d}x > 1$ and $\mu \in \mathcal{P}(\Omega)$ a given probability density on Ω. We will consider the following projection problem

$$\min_{\varrho \in K_f} W_2^2(\varrho, \mu), \tag{5.1}$$

where we set $K_f = \{\varrho \in L^1_+(\Omega) \ : \ \int_\Omega \varrho \, \mathrm{d}x = 1, \ \varrho \leq f\}$.

This section is devoted to the study of the above projection problem. We first want to summarize the main known results. Most of these results are only available in the case $f = 1$.

Existence. The existence of a solution to Problem (5.1) is a consequence of the direct method of calculus of variations. Indeed, take a minimizing sequence ϱ^n; it is tight thanks to the bound $W_2(\varrho^n, \mu) \leq C$; it admits a weakly converging subsequence and the limit minimizes the functional $W_2(\cdot, \mu)$ because of its semicontinuity and of the fact that the inequality $\varrho \leq f$ is preserved. We note that from the existence point of view, the case $f \equiv 1$ and the general case do not show any significant difference.

Characterization. The optimality conditions, derived in [22] exploiting the strategy developed in [16] (in the case $f = 1$, but they are easy to adapt to the general case) state the following: if ϱ is a solution to the above problem and φ is a Kantorovich potential in the transport from ϱ to μ, then there exists a threshold $\ell \in \mathbb{R}$ such that

$$\varrho(x) = \begin{cases} f(x), & \text{if } \varphi(x) < \ell, \\ 0, & \text{if } \varphi(x) > \ell, \\ \in [0, f(x)], & \text{if } \varphi(x) = \ell. \end{cases}$$

In particular, this shows that $\nabla\varphi = 0$ ϱ−a.e. on $\{\varrho < f\}$ and, since $T(x) = x - \nabla\varphi(x)$, that the optimal transport T from ϱ to μ is the identity on such set. If $\mu = g\mathrm{d}x$ is absolutely continuous, then one can write the Monge-Ampère equation

$$\det(DT(x)) = \varrho(x)/g(T(x))$$

and deduce $\varrho(x) = g(T(x)) = g(x)$ a.e. on $\{\varrho < f\}$. This suggests a sort of saturation result for the optimal ϱ, i.e. $\varrho(x)$ is either equal to $g(x)$ or to $f(x)$ (but one has to pay attention to the case $\varrho = 0$ and also to assume that g is absolutely continuous).

Uniqueness. For absolutely continuous measures $\mu = g \, \mathrm{d}x$ and generic f the uniqueness of the projection follows by Lemma 4.1. In the specific case $f = 1$ and Ω convex the uniqueness was proved in [16, 22] by a completely different method. In this case, as observed by A. Figalli, one can use displacement convexity along generalized geodesics. This means that if ϱ^0 and ϱ^1 are two solutions, one can take for every $t \in [0, 1]$ the convex combination $T^t = (1-t)T^0 + tT^1$ of the optimal transport maps T^i from g to ϱ^i and the curve $t \mapsto \varrho^t := ((1-t)T^0 + tT^1)_\# \mu$ in $\mathcal{P}_2$, interpolating from ϱ^0 to ϱ^1. It can be proven that ϱ^t still satisfies $\varrho^t \leq 1$ (but this can not be adapted to f, unless f is concave) and that $t \mapsto W_2^2(\varrho^t, g) < (1-t)W_2^2(\varrho^0, g) + tW_2^2(\varrho^1, g)$, which is a contradiction to the minimality. The assumption on μ can be relaxed but we need to ensure the existence of optimal transport maps: what we need to assume, is that μ gives no mass to "small" sets (i.e. $(d-1)-$dimensional); see [14] for the sharp assumptions and notions about this issue. Thanks to this uniqueness result, we can define a projection operator $P_{K_1} : \mathcal{P}_2(\Omega) \cap L^1(\Omega) \to \mathcal{P}_2(\Omega) \cap L^1(\Omega)$ through

$$P_{K_1}[g] := \operatorname{argmin}\{W_2^2(\varrho, g) \ : \ \varrho \in K_1\}.$$

Stability. From the same displacement interpolation idea, A. Roudneff-Chupin also proved ([22]) that the projection is Hölder continuous with exponent $1/2$ for the W_2 distance whenever Ω is a compact convex set. We do not develop the proof here, we just refer to Proposition 2.3.4 of [22]. Notice that the constant in the Hölder continuity depends a priori on the diameter of Ω. However, to be more precise, the following estimate is obtained (for g^0 and g^1 absolutely continuous)

$$W_2^2(P_{K_1}[g^0], P_{K_1}[g^1]) \le W_2^2(g^0, g^1) + W_2(g^0, g^1)(\operatorname{dist}(g^0, K_1) + \operatorname{dist}(g^1, K_1)), \tag{5.2}$$

which shows that, even on unbounded domains, we have a local Hölder behavior.

In the rest of the section, we want to recover similar results in the largest possible generality, i.e. for general f, and without the assumptions on μ and Ω.

We will first get a saturation characterization for the projections, which will allow for a general uniqueness result. Continuity will be an easy corollary.

In order to proceed, we first need the following lemma.

Lemma 5.1. *Let ϱ be a solution of the Problem 5.1. Let moreover $\gamma \in \Pi(\varrho, \mu)$ the optimal plan from ϱ to μ. If $(x_0, y_0) \in \operatorname{spt}(\gamma)$ then $\varrho = f$ a.e. in $B(y_0, R)$, where $R = |y_0 - x_0|$.*

Proof. Let us suppose that this is not true and there exists a compact set $K \subset B(y_0, R)$ with positive Lebesgue measure such that $\varrho < f$ a.e. in K. Let $\varepsilon := \operatorname{dist}(\partial B(y_0, R), K) > 0$.

By the definition of the support, for all $r > 0$ we have that

$$0 < \gamma(B(x_0, r) \times B(y_0, r)) \le \int_{B(x_0,r)} \varrho \, \mathrm{d}x \le \int_{B(x_0,r)} f \, \mathrm{d}x.$$

By the absolute continuity of the integral, for $r > 0$ small enough there exists $0 < \alpha \le 1$ such that

$$\gamma(B(x_0, r) \times B(y_0, r)) = \alpha \int_K (f - \varrho) \, \mathrm{d}x =: \alpha m.$$

Now we construct the following measures $\tilde{\gamma}, \eta \in \mathcal{P}(\Omega \times \Omega)$ as

$$\tilde{\gamma} := \gamma - \gamma \llcorner (B(x_0, r) \times B(y_0, r)) + \eta \quad \text{and} \quad \eta := \alpha(f - \varrho)\mathrm{d}x \llcorner K \otimes (\pi^y)_\# \gamma \llcorner (B(x_0, r) \times B(y_0, r)).$$

It is immediate to check that $(\pi^y)_\# \tilde{\gamma} = \mu$. On the other hand

$$\tilde{\varrho} := (\pi^x)_\# \tilde{\gamma} = \varrho - \varrho \llcorner B(x_0, r) + \alpha(f - \varrho) \llcorner K \le f$$

is an admissible competitor in Problem (5.1) and we have the following

$$\begin{aligned} W_2^2(\tilde{\varrho}, \mu) &\le \int_{\Omega \times \Omega} |x - y|^2 \, \mathrm{d}\tilde{\gamma}(x, y) \\ &\le W_2^2(\varrho, g) - \int_{B(x_0,r) \times B(y_0,r)} |x - y|^2 \, \mathrm{d}\gamma(x, y) + \int_{K \times B(y_0,r)} |x - y|^2 \, \mathrm{d}\eta(x, y) \\ &\le W_2^2(\varrho, g) - (R - 2r)^2 \alpha m + (R - \varepsilon + r)^2 \alpha m. \end{aligned}$$

Now if we chose $r > 0$ small enough to have $R - 2r > R - \varepsilon + r$, i.e. $r < \varepsilon/3$ we get that

$$W_2^2(\tilde{\varrho}, g) < W_2^2(\varrho, g),$$

which is clearly a contradiction, hence the result follows. □

The following proposition establishes uniqueness of the projection on K_f as well as a very precise description of it. For a given measure μ we are going to denote by μ^{ac} the density of its absolutely continuous part with respect to the Lebesgue measure, i.e.

$$\mu = \mu^{\mathrm{ac}}\mathrm{d}x + \mu^s,$$

with $\mu^s \perp \mathrm{d}x$.

Proposition 5.2. *Let $\Omega \subset \mathbb{R}^d$ be a convex set and let $f \in L^1_{\mathrm{loc}}(\Omega)$, $f \geq 0$ be such that $\int_\Omega f \geq 1$. Then, for every probability measure $\mu \in \mathcal{P}(\Omega)$, there is a unique solution ϱ of the problem (5.1). Moreover, ϱ is of the form*

$$\varrho = \mu^{\mathrm{ac}}\mathbb{1}_B + f\mathbb{1}_{B^c}, \tag{5.3}$$

for a measurable set $B \subset \Omega$.

Proof. We first note that by setting $f = 0$ on Ω^c we can assume that $\Omega = \mathbb{R}^d$. Existence of a solution in Problem 5.1 follows by the direct methods in the calculus of variations by noticing that the set K_f is closed with respect to the weak convergence of measures.

Let us prove now the saturation result (5.3). Let us first premise the following fact: if $\mu, \nu \in \mathcal{P}(\Omega)$, $\gamma \in \Pi(\mu, \nu)$ and we define the set

$$A(\gamma) := \{x \in \Omega : \text{the only point } (x, y) \in \mathrm{spt}(\gamma) \text{ is } (x, x)\},$$

then

$$\mu \llcorner A(\gamma) \leq \nu \llcorner A(\gamma). \tag{5.4}$$

In particular $\mu^{\mathrm{ac}} \leq \nu^{\mathrm{ac}}$ for a.e. $x \in A(\gamma)$. To prove (5.4), let $\phi \geq 0$ and write

$$\begin{aligned}\int_{A(\gamma)} \phi \,\mathrm{d}\mu = \int \phi(x)\mathbb{1}_{A(\gamma)}(x)\,\mathrm{d}\gamma(x, y) &= \int \phi(x)\mathbb{1}^2_{A(\gamma)}(x)\,\mathrm{d}\gamma(x, y)\\ &= \int \phi(y)\mathbb{1}_{A(\gamma)}(y)\mathbb{1}_{A(\gamma)}(x)\,\mathrm{d}\gamma(x, y)\\ &\leq \int \phi(y)\mathbb{1}_{A(\gamma)}(y)\,\mathrm{d}\gamma(x, y) = \int_{A(\gamma)} \phi\,\mathrm{d}\nu,\end{aligned}$$

where we used the fact that γ−a.e. $\mathbb{1}_{A(\gamma)}(x) > 0$ implies $x = y$.

Now, for an optimal transport plan $\gamma \in \Pi(\varrho, \mu)$, let us define

$$B := \mathrm{Leb}(f) \cap \mathrm{Leb}(\mu^{\mathrm{ac}}) \cap \mathrm{Leb}(\varrho) \cap \{\varrho < f\}^{(1)} \cap \{\varrho \neq \mu^{\mathrm{ac}}\}^{(1)} \cap A(\gamma)^{(1)} \cap A(\tilde{\gamma})^{(1)}.$$

Here $\tilde{\gamma} \in \Pi(g, \varrho)$ is the transport plan obtained by seeing γ "the other way around", i.e. $\tilde{\gamma}$ is the image of γ through the maps $(x, y) \mapsto (y, x)$ while $\mathrm{Leb}(h)$ is the set of Lebesgue points of h and for a set A we denote by $A^{(1)} := \mathrm{Leb}(\mathbb{1}_A)$ the set of its density one points.

Let now $x_0 \in B$ and let us consider the following two cases:

Case 1. $\varrho(x_0) < \mu^{\mathrm{ac}}(x_0)$. Since, in particular, $\mu^{\mathrm{ac}}(x_0) > 0$ and $x_0 \in \mathrm{Leb}(\mu^{\mathrm{ac}})$ we have that $x_0 \in \mathrm{spt}(\mu)$. From Lemma 5.1 wee see that $(y_0, x_0) \in \mathrm{spt}(\gamma)$ implies $y_0 = x_0$. Indeed if this were not the case there would exist a ball where $\varrho = f$ a.e. and x_0 would be in the middle of this ball; from $x_0 \in \mathrm{Leb}(f) \cap \mathrm{Leb}(\varrho)$ we would get $\varrho(x_0) = f(x_0)$ a contradiction with $x_0 \in B$. Hence, if we use the set $A(\tilde{\gamma})$ defined above with $\nu = \varrho$, we have $x_0 \in A(\tilde{\gamma})$. From $x_0 \in \mathrm{Leb}(\mu^{\mathrm{ac}}) \cap \mathrm{Leb}(\varrho)$ we get $\mu^{\mathrm{ac}}(x_0) \leq \varrho(x_0)$, which is a contradiction.

Case 2. $\mu^{\rm ac}(x_0) < \varrho(x_0)$. Exactly as in the previous case we have that $x_0 \in {\rm spt}(\varrho)$ and, by the Lemma 5.1, we have again that $(x_0, y_0) \in {\rm spt}(\gamma)$ implies $y_0 = x_0$. Indeed, otherwise x_0 would be on the boundary of a ball where $\varrho = f$ a contradiction with $x_0 \in \{\varrho < f\}^{(1)}$. Hence, we get $x_0 \in A(\gamma)$ and $\varrho(x_0) \le \mu^{\rm ac}(x_0)$, again a contradiction.

Hence we get that $\mu^{\rm ac} = \varrho$ for $x \in B$. By the definition of B,

$$B^c \subset_{\rm a.e.} \{\varrho = f\} \cup A(\gamma)^c \cup A(\tilde\gamma)^c\,,$$

where a.e. refers to the Lebesgue measure. By applying Lemma 5.1, this implies that $\varrho = f$ a.e. on B^c, and concludes the proof of (5.3).

Uniqueness of the projection it is now an immediate consequence of the saturation property (5.3). Indeed, suppose that ϱ_0 and ϱ_1 were two different projections of a same measure g. Define $\varrho_{1/2} = \frac{1}{2}\varrho_0 + \frac{1}{2}\varrho_1$. Then, by convexity of $W_2^2(\cdot, \mu)$, we get that $\varrho_{1/2}$ is also optimal. But its density is not saturated on the set where the densities of ϱ_0 and ϱ_1 differ, in contradiction with (5.3). □

Corollary 5.3. *For fixed f, the map $P_{K_f} : \mathcal{P}_2(\Omega) \to \mathcal{P}_2(\Omega)$ defined through*

$$P_{K_f}[\mu] := {\rm argmin}\{W_2^2(\varrho, \mu) \ : \ \varrho \in K_f\}$$

is continuous in the following sense: if $\mu_n \to \mu$ for the W_2 distance, then $P_{K_f}[\mu_n] \rightharpoonup P_{K_f}[\mu]$ in the weak convergence.

Moreover, in the case where $f = 1$ and Ω is a convex set, the projection is also locally $\frac{1}{2}-$Hölder continuous for W_2 on the whole $\mathcal{P}(\Omega)$ and satisfies (5.2).

Proof. This is just a matter of compactness and uniqueness. Indeed, take a sequence $\mu_n \to \mu$ and look at $P_{K_f}[\mu_n]$. It is a tight sequence of measures since

$$W_2(P_{K_f}[\mu_n], \mu) \le W_2(P_{K_f}[\mu_n], \mu_n) + W_2(\mu_n, \mu) \le W_2(\varrho, \mu) + 2W_2(\mu_n, \mu)\,, \tag{5.5}$$

where $\varrho \in K_f$ is any admissible measure. Hence we can extract a weakly converging subsequence to some measure $\tilde\varrho \in K_f$ (recall that K_f is weakly closed). Moreover, by the lower semicontinuity of W_2 with respect to the weak convergence and since $W_2(\mu_n, \mu) \to 0$, passing to the limit in (5.5) we get

$$W_2(\tilde\varrho, \mu) \le W_2(\varrho, \mu) \qquad \forall\, \varrho \in K_f.$$

Uniqueness of the projection implies $\tilde\varrho = P_{K_f}(\mu)$ and thus that the limit is independent on the extracted subsequence, this proves the desired continuity.

Concerning the second part of the statement, we take arbitrary μ^1 and μ^2 (not necessarily absolutely continuous) and we approximate them in the W_2 distance with absolutely continuous measures g_n^i ($i = 1, 2$; for instance by convolution), then we have, from (5.2)

$$W_2^2(P_{K_1}[g_n^0], P_{K_1}[g_n^1]) \le W_2^2(g_n^0, g_n^1) + W_2(g_n^0, g_n^1)({\rm dist}(g_n^0, K_1) + {\rm dist}(g_n^1, K_1)),$$

and we can pass to the limit as $n \to \infty$. □

The following technical lemma will be used in the next section and establishes the continuity of the projection with respect to f. To state it let us consider for given $f \in L^1_{\rm loc}$ and $\mu \in \mathcal{P}_2(\Omega)$ let us consider following functional

$$\mathcal{F}_f(\varrho) := \begin{cases} \frac{1}{2}W_2^2(\mu, \varrho), & \text{if } \varrho \in K_f \\ +\infty, & \text{otherwise.} \end{cases}$$

Proposition 5.2 can be restated by saying that the functional $\mathcal{F}_f$ has a unique minimizer in $\mathcal{P}_2(\Omega)$.

Lemma 5.4. *Let $f_n, f \in L^1_{\mathrm{loc}}(\Omega)$ with $\int_\Omega f_n \,\mathrm{d}x \geq 1$, $\int_\Omega f \,\mathrm{d}x \geq 1$ and let us assume that $f_n \to f$ in $L^1_{\mathrm{loc}}(\Omega)$ and almost everywhere. Also assume $f_n \in \mathcal{P}_2(\Omega)$ if $\int_\Omega f_n \,\mathrm{d}x = 1$ and $f \in \mathcal{P}_2(\Omega)$ if $\int_\Omega f \,\mathrm{d}x = 1$. Then, for every $\mu \in \mathcal{P}_2(\Omega)$,*

(i) *The sequence $(P_{K_{f_n}}(\mu))_n$ is tight.*
(ii) *We have $P_{K_{f_n}}(\mu) \rightharpoonup P_{K_f}(\mu)$.*
(iii) *If $\int_\Omega f > 1$, then $\mathcal{F}_{f_n}$ $\Gamma-$converges to $\mathcal{F}_f$ with respect to the weak convergence of measures.*

Proof. Let us denote by $\bar{\varrho}_n$ the projection $P_{K_{f_n}}(\mu)$ and let us start from proving its tightness, i.e. *(i)*. We fix $\varepsilon > 0$: there exists a radius R_0 such that $\mu(B(0, R_0)) > 1 - \frac{\varepsilon}{2}$ and $\int_{B(0,R_0)} f > 1 - \frac{\varepsilon}{2}$. By L^1_{loc} convergence, there exists n_0 such that $\int_{B(0,R_0)} f_n > 1 - \varepsilon$ pour $n > n_0$. Now, take $R > 3R_0$ and suppose $\bar{\varrho}_n(B(0, R)^c) > \varepsilon$ for $n \geq n_0$. Then, the optimal transport T from $\bar{\varrho}_n$ to μ should move some mass from $B(0, R)^c$ to $B(0, R_0)$. Let us take a point $x_0 \in B(0, R)^c$ such that $T(x_0) \in B(0, R_0)$. From Lemma 5.1, this means that $\bar{\varrho}_n = f_n$ on the ball $B(T(x_0), |x_0 - T(x_0)|) \supset B(T(x_0), 2R_0) \supset B(0, R_0)$. But this means $\int_{B(0,R_0)} \bar{\varrho}_n = \int_{B(0,R_0)} f_n > 1 - \varepsilon$, and hence $\bar{\varrho}_n(B(0, R)^c) \leq \varepsilon$, which is a contradiction. This shows that $\bar{\varrho}_n$ is tight.

Now, if $\int_\Omega f = 1$, then the weak limit of $\bar{\varrho}_n$ up to subsequences can only be f itself, since it must be a probability density bounded above by f. And $f = P_{K_f}(\mu)$. This proves *(ii)* in the case $\int_\Omega f = 1$. In the case $\int_\Omega f > 1$, this will be a consequence of *(iii)*. Notice that in this case we necessarily have $\int_\Omega f_n > 1$ for n large enough.

Let us prove *(iii)*. Since $\varrho_n \leq f_n$, $\varrho_n \rightharpoonup \varrho$ and $f_n \to f$ in L^1_{loc} immediately implies that $\varrho \leq f$, the $\Gamma-$liminf inequality simply follows by the lower semicontinuity of W_2.

Concerning the $\Gamma-$limsup, we need to prove that every density $\varrho \in \mathcal{P}_2(\Omega)$ with $\varrho \leq f$ a.e. can be approximated by a sequence $\varrho_n \leq f_n$ a.e. with $W_2(\varrho_n, g) \to W_2(\varrho, g)$. In order to do this let us define $\tilde{\varrho}_n := \min\{\varrho, f_n\}$. Note that $\tilde{\varrho}_n$ is not admissible since it is not a probability, because in general $\int \tilde{\varrho}_n < 1$. Yet, we have $\int \tilde{\varrho}_n \to 1$ since $\tilde{\varrho}_n \to \min\{\varrho, f\} = \varrho$ and this convergence is dominated by ϱ. We want to "complete" $\tilde{\varrho}_n$ so as to get a probability, stay admissible, and converge to ϱ in W_2, since this will imply that $W_2(\varrho_n, g) \to W_2(\varrho, g)$.

Let us select a ball B such that $\int_{B\cap\Omega} f > 1$ and note that we can find $\varepsilon > 0$ such that the set $\{f > \varrho + \varepsilon\} \cap B$ is of positive measure, i.e. $m := |\{f > \varrho + \varepsilon\} \cap B| > 0$. Since $f_n \to f$ a.e., the set $B_n := \{f_n > \varrho + \frac{\varepsilon}{2}\} \cap B$ has measure larger than $m/2$ for large n. Now take $B'_n \subset B_n$ with $|B'_n| = \frac{2}{\varepsilon}(1 - \int \tilde{\varrho}_n) \to 0$, and define

$$\varrho_n := \tilde{\varrho}_n + \frac{\varepsilon}{2} \mathbb{1}_{B'_n}.$$

By construction, $\int \varrho_n = 1$ and $\varrho_n \leq f_n$ a.e. since on B'_n we have $\tilde{\varrho}_n = \varrho$ and $\varrho + \frac{\varepsilon}{2} < f_n$ while on the complement of B'_n, $\tilde{\varrho}_n \leq f_n$ a.e. by definition. To conclude the proof we only need to check $W_2(\varrho_n, \varrho) \to 0$. This is equivalent (see, for instance, [2] or [25]) to

$$\int \phi \varrho_n \to \int \phi \varrho \tag{5.6}$$

for all continuous functions ϕ with such that $\phi \leq C(1 + |x|^2)$. Since $\varrho \in \mathcal{P}_2(\Omega)$ and $\tilde{\varrho}_n \leq \varrho$, thank to the dominated convergence theorem it is enough to show that $\int \phi(\varrho_n - \tilde{\varrho}_n) \to 0$. But

$\varrho_n - \tilde{\varrho}_n$ converges to 0 in L^1 and it is supported in $B'_n \subset B$. Since ϕ is bounded on B we obtain the desired conclusion. □

Remark 5.1. Let us conclude this section with the following *open question*: in a Hilbert space, the only fact that the projection onto a set K is uniquely determined for every starting point implies that K is convex and thus that the projection is 1-Lipschitz continuous. Here we are in a metric space which has a sort of Hilbertian manifold structure (see [2, 14]), and we could wonder if the same is true stays true. The set K_f is always convex in the usual sense, but it is also geodesically convex (which seems to be more pertinent in this setting) when $f = 1$, and also convex w.r.t. generalized geodesics.

For $f = 1$ the projection is continuous and we can even provide Hölder bounds on P_{K_1}. The question whether P_{K_1} is 1-Lipschitz, as far as we know, is open. Let us underline that some sort of 1-Lipschitz results have been proven in [6] for solutions of similar variational problems, but seem impossible to adapt in this framework.

For the case $f \neq 1$ even the continuity of the projection with respect to the Wasserstein distance seems delicate.

5.2. **BV estimates for P_{K_f}.** In this section, we prove Theorem 1.2. Notice that the case $f = 1$ has already been proven as a particular case of Theorem 1.1. To handle the general case, we develop a slightly different strategy, based on the standard idea to approximate L^∞ bounds with L^p penalizations.

Let $m \in \mathbb{N}$ and let us assume that $\inf f > 0$, for $\mu \in \mathcal{P}_2(\Omega)$, we define the approximating functionals $\mathcal{F}_m : L^1_+(\Omega) \to \mathbb{R} \cup \{+\infty\}$ by

$$\mathcal{F}_m(\varrho) := \frac{1}{2} W_2^2(\mu, \varrho) + \frac{1}{m+1} \int_\Omega \left(\frac{\varrho}{f}\right)^{m+1} \mathrm{d}x + \frac{\varepsilon_m}{2} \int_\Omega \left(\frac{\varrho}{f}\right)^2 \mathrm{d}x$$

and the limit functional $\mathcal{F}$ as

$$\mathcal{F}(\varrho) := \begin{cases} \frac{1}{2} W_2^2(\mu, \varrho), & \text{if } \varrho \in K_f \\ +\infty, & \text{otherwise} \end{cases}$$

Here $\varepsilon_m \downarrow 0$ is a small parameter to be chosen later.

Lemma 5.5. *Let $\Omega \subset \mathbb{R}^d$ and $f : \Omega \to (0, +\infty)$ be a measurable function, bounded from below and from above by positive constants and let $\mu \in \mathcal{P}_2(\Omega)$. Then:*

(i) *There are unique minimizers ϱ, ϱ_m in $L^1(\Omega)$ for each of the functionals $\mathcal{F}$ and $\mathcal{F}_m$, respectively.*
(ii) *The family of functionals $\mathcal{F}_m$ Γ-converges for the weak convergence of probability measures to $\mathcal{F}$, and the minimizers ϱ_m weakly converge to ϱ, as $m \to \infty$.*
(iii) *The minimizers ϱ_m of $\mathcal{F}_m$ satisfy*

$$\varphi_m + \left(\frac{\varrho_m}{f}\right)^m \frac{1}{f} + \varepsilon_m \left(\frac{\varrho_m}{f}\right) \frac{1}{f} = 0, \tag{5.7}$$

for a suitable Kantorovich potential φ_m in the transport from ϱ_m to μ.

Proof. Existence and uniqueness of minimizers of $\mathcal{F}$ has been established in Proposition 5.2. Existence of minimizers of $\mathcal{F}_m$ is again a simple application of the direct methods in the calculus of variations and uniqueness follows from strict convexity.

Let us prove the $\Gamma-$convergence in *(ii)*. In order to prove the $\Gamma-$liminf inequality, let $\varrho_m \rightharpoonup \varrho$. If $\mathcal{F}_m(\varrho_m) \leq C$, then for every $m_0 \leq m$ and every finite measure set $A \subset \Omega$, we have

$$\|\varrho_m/f\|_{L^{m_0}(A)} \leq |A|^{\frac{1}{m_0}-\frac{1}{m+1}}(C(m+1))^{\frac{1}{m+1}}.$$

If we pass to the limit $m \to \infty$, from $\frac{\varrho_m}{f} \rightharpoonup \frac{\varrho}{f}$, we get $||\varrho/f||_{L^{m_0}(A)} \leq |A|^{\frac{1}{m_0}}$. Letting m_0 go to infinity we obtain $||\varrho/f||_{L^\infty} \leq 1$, i.e. $\varrho \in K_f$. Since

$$\mathcal{F}_m(\varrho_m) \geq \frac{1}{2}W_2^2(\mu, \varrho_m),$$

the lower semicontinuity of W_2^2 with respect to weak converges proves the $\Gamma-$liminf inequality.

In order to prove $\Gamma-$limsup, we use the constant sequence $\varrho_m = \varrho$ as a recovery sequence. Since we can assume $\varrho \leq f$ (otherwise there is nothing to prove, since $\mathcal{F}(\varrho) = +\infty$), it is clear that the second and third parts of the functional tend to 0, thus proving the desired inequality.

The last part of the statement finally follows form Theorem 2.1 (vi) and Lemma 2.5 □

Proof of Theorem 1.2

Proof. Clearly we can assume that $TV(g, \Omega)$ and $TV(f, \Omega)$ are finite and that $\int_\Omega f > 1$ since otherwise the conclusion is trivial.

Step 1. Assume that the support of g is compact, that $f \in C^\infty(\Omega)$ is bounded from above and below by positive constants, and let ϱ_m be the minimizer of $\mathcal{F}_m$. As in the proof of Theorem 1.1, we can use the optimality condition (5.7) to prove that ϱ is compactly supported. Also, the same condition imply that ϱ is Lipschitz continuous. Indeed, we can write (5.7) as

$$\varphi f + H_m'\left(\frac{\varrho}{f}\right) = 0,$$

where $H_m(t) = \frac{1}{m+1}t^{m+1} + \frac{\varepsilon_m}{2}t^2$. Since H_m is smooth and convex and H_m'' is bounded from below by a positive constant H_m' is invertible and

$$\varrho = f \cdot (H_m')^{-1}(\varphi f),$$

where $(H_m')^{-1}$ is Lipschitz continuous. Since φ and f are locally Lipschitz, this gives Lipschitz continuity for ϱ on a neighborhood of its support.

Taking the derivative of the optimality condition (5.7) we obtain

$$\nabla\varphi_m + \left(m\left(\frac{\varrho_m}{f}\right)^{m-1} + \varepsilon_m\right)\frac{f\nabla\varrho_m - \varrho_m\nabla f}{f^3} - \left(\left(\frac{\varrho_m}{f}\right)^m + \varepsilon_m\frac{\varrho_m}{f}\right)\frac{\nabla f}{f^2} = 0.$$

Rearranging the terms we have

$$\nabla\varphi_m + A\nabla\varrho_m - B\nabla f = 0,$$

where by A and B we denote the (positive!) functions

$$A := \left(m\left(\frac{\varrho_m}{f}\right)^{m-1} + \varepsilon_m\right)\frac{1}{f^2} \quad \text{and} \quad B := \left(m\left(\frac{\varrho_m}{f}\right)^{m-1} + \varepsilon_m\right)\frac{\varrho_m}{f^3} + \left(\left(\frac{\varrho_m}{f}\right)^m + \varepsilon_m\frac{\varrho_m}{f}\right)\frac{1}{f^2}.$$

Now we will use the inequality from Corollary 3.3 for ϱ_m and g in the form

$$\int_\Omega |\nabla\varrho_m|\,\mathrm{d}x \leq \int_\Omega |\nabla g|\,\mathrm{d}x + \int_\Omega \nabla\varrho_m \cdot \left(\frac{\nabla\varrho_m}{|\nabla\varrho_m|} + \frac{\nabla\varphi_m}{|\nabla\varphi_m|}\right)\mathrm{d}x.$$

In order to estimate the second integral on the right-hand side we use the inequality

$$\left|\frac{a}{|a|}-\frac{b}{|b|}\right| \le \left|\frac{a}{|a|}-\frac{b}{|a|}\right| + \left|\frac{b}{|a|}-\frac{b}{|b|}\right| = \frac{|a-b|}{|a|} + \frac{|b|-|a|}{|a|} \le \frac{2}{|a|}|a-b|, \tag{5.8}$$

for all non-zero $a, b \in \mathbb{R}^d$ (that we apply to $a = A\nabla\varrho_m$ and $b = -\nabla\varphi_m$), and we obtain

$$\begin{aligned}\int_\Omega |\nabla\varrho_m|\,\mathrm{d}x &\le \int_\Omega |\nabla g|\,\mathrm{d}x + \int_\Omega |\nabla\varrho_m|\cdot\left|\frac{A\nabla\varrho_m}{A|\nabla\varrho_m|}+\frac{\nabla\varphi_m}{|\nabla\varphi_m|}\right|\,\mathrm{d}x\\ &\le \int_\Omega |\nabla g|\,\mathrm{d}x + 2\int_\Omega \frac{1}{A}\left|A\nabla\varrho_m+\nabla\varphi_m\right|\mathrm{d}x\\ &\le \int_\Omega |\nabla g|\,\mathrm{d}x + 2\int_\Omega \frac{B}{A}|\nabla f|\,\mathrm{d}x.\end{aligned}$$

We must now estimate the ration B/A. If we denote by λ the ratio ϱ_m/f we may write

$$\frac{B}{A} = \lambda + \lambda\frac{\varepsilon_m+\lambda^{m-1}}{\varepsilon_m+m\lambda^{m-1}} \le \lambda\left(1+\frac{1}{m}\right) + \frac{\varepsilon_m\lambda}{\varepsilon_m+m\lambda^{m-1}}.$$

Now, consider that

$$\max_{\lambda\in\mathbb{R}_+}\frac{\varepsilon_m\lambda}{\varepsilon_m+m\lambda^{m-1}} = \frac{m-2}{m-1}\left(\frac{\varepsilon_m}{m(m-2)}\right)^{1/(m-1)} =: \delta_m$$

is a quantity depending on m and tending to 0 if ε_m is chosen small enough (for instance $\varepsilon_m = 2^{-m^2}$). This allows to write

$$\int_\Omega |\nabla\varrho_m|\,\mathrm{d}x \le \int_\Omega |\nabla g|\,\mathrm{d}x + 2\left(1+\frac{1}{m}\right)\int_\Omega \frac{\varrho_m}{f}|\nabla f|\,\mathrm{d}x + 2\delta_m\int_\Omega |\nabla f|\,\mathrm{d}x.$$

In the limit, as $m \to +\infty$, we obtain

$$\int_\Omega |\nabla\varrho|\,\mathrm{d}x \le \int_\Omega |\nabla g|\,\mathrm{d}x + 2\int_\Omega \frac{\varrho}{f}|\nabla f|\,\mathrm{d}x.$$

Using the fact that $\varrho \le f$, we get

$$\int_\Omega |\nabla\varrho|\,\mathrm{d}x \le \int_\Omega |\nabla g|\,\mathrm{d}x + 2\int_\Omega |\nabla f|\,\mathrm{d}x.$$

Step 2. To treat the case $g, f \in BV_{\mathrm{loc}}(\Omega)$ we proceed by approximation as in the proof of Theorem 1.1. To do this we just note that Corollary 5.3 and Lemma 5.4 give the desired continuity property of the projection with respect both to g and f, lower semicontinuity of the total variation with respect to the weak convergence then implies the conclusion. □

Remark 5.2. We conclude this section by noticing that the constant 2 in Equation (1.3) can not be replaced by any smaller constant. Indeed if $\Omega = \mathbb{R}$, $f = \mathbb{1}_{\mathbb{R}_+}$, $g = \frac{1}{n}\mathbb{1}_{[-n,0]}$ then $\varrho = P_{K_f}(g) = \mathbb{1}_{[0,1]}$ and $\int|\nabla\varrho| = 2$, $\int|\nabla f| = 1$, $\int|\nabla g| = \frac{2}{n}$.

6. Applications

In this section we discuss some applications of Theorems 1.1 and 1.2 and we present some open problem.

6.1. **Partial transport.** The projection problem on K_f is a particular case of the so called *partial transport problem*, see [12, 13]. Indeed, the problem is to transport μ to a part of the measure f, which is a measure with mass larger than 1. As typical in the partial transport problem, the solution has an active region, which is given by f restricted to a certain set. This set satisfies a sort of interior ball condition, with a radius depending on the distance between each point and its image. In the partial transport case some regularity ($C^{1,\alpha}$) is known for the optimal map away from the intersection of the supports of the two measures.

A natural question is how to apply the technique that we developed here in the framework of more general partial transport problems (in general, both measures could have mass larger than 1 and could be transported only partially), and/or whether results or ideas from partial transport could be translated into the regularity of the free boundary in the projection.

6.2. **Shape optimization.** If we take a set $A \subset \mathbb{R}^d$ with $|A| < 1$ and finite second moment $\int_A |x|^2 \,\mathrm{d}x < +\infty$, a natural question is which is the set B with volume 1 such that the uniform probability density on B is closest to that on A. This means solving a shape optimization problem of the form

$$\min\{W_2^2(\mathbb{1}_B, \frac{1}{|A|}\mathbb{1}_A) \;:\; |B| = 1\}.$$

The considerations in Section 5.1 show that solving such a problem is equivalent to solving

$$\min\{W_2^2(\varrho, \frac{1}{|A|}\mathbb{1}_A) \;:\; \varrho \in \mathcal{P}_2(\mathbb{R}^d)\}$$

and that the optimal ϱ is of the form $\varrho = \mathbb{1}_B$, $B \supset A$. Also, from our Theorem 1.2 (with $f = 1$), we deduce that if A is of finite perimeter, then the same is true for B, and $\mathrm{Per}(B) \le \frac{1}{|A|}\mathrm{Per}(A)$ (i.e. the perimeter is bounded by the Cheeger ratio of A).

It is interesting to compare this problem with this perimeter bound with the problem studied in [19], which has the same words but in different order: more precisely: here we minimize the Wasserstein distance and we try to get an information on the perimeter, in [19] the functional to be minimized is a combination of W_2 and the perimeter. Hence, the techniques to prove any kind of results are different, because here W_2 cannot be considered as a lower order perturbation of the perimeter.

As a consequence, many natural questions arise: if A is a nice closed set, can we say that B contains A in its interior? if A is convex is B convex? what about the regularity of ∂B?

6.3. **Set evolution problems.** Consider the following problem. For a given set $A \subset \mathbb{R}^d$ we define $\varrho_0 = \mathbb{1}_A$. For a time interval $[0,T]$ and a time step $\tau > 0$ (and $N + 1 := \left[\frac{T}{\tau}\right]$) we consider the following scheme $\varrho_0^\tau := \varrho_0$ and

$$\varrho_{k+1}^\tau := P_{K_1}\left[(1+\tau)\varrho_k^\tau\right],\; k \in \{0, \dots, N-1\}, \tag{6.1}$$

(here we extend the notion of Wasserstein distance and projection to measures with the same mass, even if different from 1: in particular, the mass of ϱ_k^τ will be $|A|(1+\tau)^k$ and at every step we project ϱ_k^τ on the set of finite positive measure, with the same mass of ϱ_k^τ, and with density bounded by 1, and we still denote this set by K_1 and the projection operator in the sense of the quadratic Wasserstein distance onto this set by P_{K_1}). We want to study the convergence of this algorithm as $\tau \to 0$. This is a very simplified model for the growth of a biological population, which increases exponentially in size (supposing that there is enough food: see [17] for a more sophisticated model) but is subject to a density constraint because each individual needs a certain amount of space. Notice that this scheme formally follows the same evolution as in the Hele-Shaw flow (this can be justified by the fact that, close to uniform density the W_2 distance and the H^{-1} distance are asymptotically the same).

Independently of the compactness arguments that we need to prove the convergence of the scheme, we notice that, for fixed $\tau > 0$, all the densities ϱ_k^τ are indeed indicator functions (this comes from the consideration in Section 5.1). Thus we have an evolution of sets. A natural question is whether this stays true when we pass to the limit as $\tau \to 0$. Indeed, we generally prove convergence of the scheme in the weak sense of measures, and it is well-known that, in general, a weak limit of indicator functions is not necessarily an indicator itself. However Theorem 1.2 provides an a priori bound the perimeter of these sets. This BV bound allows to transform weak convergence as measures into strong L^1 convergence, and to preserve the fact that these densities are indicator functions.

Notice on the other hand that the same result could not be applied in the case where the projection was performed onto K_f, for a non-constant f. The reason lies in the term $2\int|\nabla f|$ in the estimate we provided. This means that, a priori, instead of being decreasing, the total variation could increase at each step of a fixed amount $2\int|\nabla f|$. When $\tau \to 0$, the number of iterations diverges and this does not allow to prove any BV estimate on the solution. Yet, a natural question would be to prove that the set evolution is well-defined as well, using maybe the fact that these sets are increasing in time.

6.4. **Crowd movement with diffusion.** In [16, 22] crowd movement models where a density ϱ evolves according to a given vector field v, but subject to a density constraint $\varrho \le 1$ are studied. This means that, without the density constraint, the equation would be $\partial_t\varrho + \nabla\cdot(\varrho v) = 0$, and a natural way to discretize the constrained equation would be to set $\tilde\varrho_{k+1}^\tau = (id + \tau v)_\#\varrho_k^\tau$ and then $\varrho_{k+1}^\tau = P_{K_1}[\tilde\varrho_{k+1}^\tau]$.

What happens if we want to add some diffusion, i.e. if the continuity equation is replaced by a Fokker-Planck equation $\partial_t\varrho - \Delta\varrho + \nabla\cdot(\varrho v) = 0$? among other possible methods, one discretization idea is the following: define $\tilde\varrho_{k+1}^\tau$ by following the unconstrained Fokker-Planck equation for time τ starting from ϱ_k^τ, and then project. In order to get some compactness of the discrete curves we need to estimate the distance between ϱ_k^τ and $\tilde\varrho_{k+1}^\tau$. It is not difficult to see that the speed of the solution of the Heat Equation (and also of the Fokker-Planck equation) for the distance W_p is related to $\|\nabla\varrho\|_{L^p}$. It is well known that these parabolic equations regularize and so the L^p norm of the gradient will not blow up in time, but we have to keep into account the projections that we perform every time step τ. From the discontinuities that appear in the projected measures, one cannot expected that $W^{1,p}$ bounds on ϱ are preserved. The only reasonable bound is for $p = 1$, i.e. a BV bound, which is exactly what is provided in this paper.

The application to crowd motion with diffusion are a matter of current study by the second and third author [20].

6.5. **BV estimates for some degenerate diffusion equation.** In this subsection we apply our main Theorem 1.1 to establish BV estimates for for some degenerate diffusion equation. BV estimates for these equations are usually known and they can be derived by looking at the evolution in time of the BV norm of the solution. Theorem Theorem 1.1 allows to give an optimal transport proof of these estimates. Let $h : \mathbb{R}^+ \to \mathbb{R}$ be a given super-linear convex function and let us consider the problem

$$\begin{cases} \partial_t \varrho_t = \nabla \cdot (h''(\varrho_t)\rho_t \nabla \rho_t)\,, & \text{in } (0,T] \times \mathbb{R}^d, \\ \varrho(0,\cdot) = \varrho_0, & \text{in } \mathbb{R}^d, \end{cases} \tag{6.2}$$

where ϱ_0 is a non-negative BV probability density. We remark that by the evolution for any $t \in (0,T]$ ϱ_t will remain a non-negative probability density. In the case $h(\rho) = \rho^m/(m-1)$ in equation (6.2) we get precisely the *porous medium equation* $\partial_t \rho = \Delta(\rho^m)$ (see [26]).

Since the seminal work of F. Otto ([21]) we know that the problem (6.2) can be seen as a gradient flow of the functional

$$\mathcal{F}(\varrho) := \int_{\mathbb{R}^d} h(\varrho)$$

in the space $(\mathcal{P}(\mathbb{R}^d), W_2)$. As a gradient flow, this equation can be discretized in time through an implicit Euler scheme. More precisely let us take a time step $\tau > 0$ and let us consider the following scheme: $\varrho_0^\tau := \varrho_0$ and

$$\varrho_{k+1}^\tau := \mathrm{argmin}_\varrho \left\{ \frac{1}{2\tau} W_2^2(\varrho, \varrho_k^\tau) + \int h(\varrho) \right\},\ k \in \{0, \dots, N-1\}. \tag{6.3}$$

where $N := \left[\frac{T}{\tau}\right]$. Defining piecewise constant and geodesic interpolations between the ϱ_k^τ's with the corresponding velocities and momentums, it is possible to show that as $\tau \to 0$ we will get a curve $\varrho_t,\ t \in [0,T]$ in $(\mathcal{P}(\mathbb{R}^d), W_2)$ which solves

$$\begin{cases} \partial_t \rho_t + \nabla \cdot (\varrho_t v_t) = 0 \\ v_t = -h''(\varrho_t) \nabla \varrho_t, \end{cases}$$

hence

$$\partial_t \varrho_t - \nabla \cdot (h''(\varrho_t)\varrho_t \nabla \varrho_t) = 0,$$

that is ϱ_t is a solution to (6.2), see [2] for a rigorous presentation of these facts.

We now note that Theorem 1.1 implies that

$$\int_{\mathbb{R}^d} |\nabla \varrho_{k+1}^\tau| \,\mathrm{d}x \le \int_{\mathbb{R}^d} |\nabla \varrho_k^\tau| \,\mathrm{d}x,$$

hence the total variation decreases for the sequence $\varrho_0^\tau, \dots, \varrho_N^\tau$. As the estimations do not depend on $\tau > 0$ this will remain true also in the limit $\tau \to 0$. Hence (assuming uniqueness for the limiting equation) we get that for any $t, s \in [0,T],\ t > s$

$$TV(\varrho_t, \mathbb{R}^d) \le TV(\varrho_s, \mathbb{R}^d),$$

and in particular for any $t \in [0,T]$

$$TV(\varrho_t, \mathbb{R}^d) \le TV(\varrho_0, \mathbb{R}^d).$$

References

[1] L. Ambrosio, Movimenti minimizzanti, *Rend. Accad. Naz. Sci. XL Mem. Mat. Sci. Fis. Natur.*, 113 (1995), 191-246.

[2] L. Ambrosio, N. Gigli, G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Math., ETH Zürich, (2005).

[3] G. Bouchitté, G. Buttazzo, New lower semicontinuity results for nonconvex functionals defined on measures, *Nonlinear Anal.*, 15 (1990), No. 7, 679-692.

[4] G. Buttazzo, *Semicontinuity, relaxation, and integral representation in the calculus of variations*, Longman Scientific & Technical, 1989.

[5] G. Buttazzo, F. Santambrogio, A model for the optimal planning of an urban area, *SIAM J. Math. Anal.* 37 (2005), No. 2, 514-530.
[6] E. Carlen, K. Craig, Contraction of the proximal map and generalized convexity of the Moreau-Yosida regularization in the 2-Wasserstein metric, *Math. and Mech. Compl. Syst.*, Vol. 1 (2013), No. 1, 33–65.
[7] G. Carlier, F. Santambrogio, A variational model for urban planning with traffic congestion, *ESAIM Contr. Opt. Calc. Var.* Vol. 11, No. 4, (2005), 595-613.
[8] L. A. Caffarelli, The regularity of mappings with a convex potential, *J. Amer. Math. Soc.*, 5 (1992), No. 1, 99-104.
[9] L. A. Caffarelli, Boundary regularity of maps with convex potentials, *Comm. Pure Appl. Math.*, 45 (1992), No. 9, 1141-1151.
[10] P. Cannarsa, C. Sinestrari, *Semiconcave functions, Hamilton-Jacobi euqations, and optimal control*, Birkhäuser, (2004).
[11] G. Dal Maso: *An Introduction to* $\Gamma-$*convergence*. Birkhäuser, Basel, (1992).
[12] A. Figalli, A note on the regularity of the free boundaries in the optimal partial transport problem, *Rendiconti del Circolo Matematico di Palermo*, 58 (2009), 283-286.
[13] A. Figalli, The optimal partial transport problem, *Arch. Rat. Mech. Anal.*, 195 (2010), 533-560.
[14] N. Gigli, On the inverse implication of Brenier-McCann theorems and the structure of $(\mathcal{P}_2(M), W_2)$, *Meth. Appl. of Anal.*, Vol 18 (2011), no 2, 127-158.
[15] J. Lellmann, D.A. Lorenz, C. Schönlieb, T. Valkonen, Imaging with Kantorovich-Rubinstein discrepancy, preprint (available at `http://arxiv.org/pdf/1407.0221v1.pdf`
[16] B. Maury, A. Roudneff-Chupin, F. Santambrogio A macroscopic crowd motion model of gradient flow type, *Math. Models and Methods in Appl. Sciences*, Vol. 20 (2010), No. 10, 1787-1821.
[17] B. Maury, A. Roudneff-Chupin, F. Santambrogio Congestion-driven dendritic growth, *Discr. Cont. Dyn. Syst.*, Vol. 34 (2014), No. 4, 1575-1604.
[18] R. J. McCann, A convexity principle for interacting gases. *Adv. Math.* 128 (1997), No. 1, 153-159.
[19] E. Milakis, On the regularity of optimal sets in mass transfer problems, *Comm. Partial Differential Equations*, 31 (2006), no. 4-6, 817-826.
[20] A. R. Mészáros, F. Santambrogio, A diffusive model for macroscopic crowd movements with congestion, *in preparation*.
[21] F. Otto, The geometry of dissipative evolution equations: the porous medium equation, *Commun. in PDE*, 26 (2001), No. 1-2, 101-174.
[22] A. Roudneff-Chupin, Modélisation macroscopique de mouvements de foule, PhD Thesis, Université Paris-Sud, (2011), available at `http://www.math.u-psud.fr/∼roudneff/Images/these_roudneff.pdf`
[23] F. Santambrogio, Transport and concentration problems with interaction effects. *J. Global Optim.*, 38 (2007), no. 1, 129–141.
[24] F. Santambrogio, Optimal Transport for Applied Mathematicians. To be published by Birkäuser. An incomplete version is available at `http://www.math.u-psud.fr/ santambr/OTAM.pdf`
[25] Cédric Villani, *Optimal Transport. Old and New.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 338. Springer-Verlag, Berlin, 2009.
[26] J. L Vázquez, *The porous medium equation. Mathematical theory*, The Clarendon Press, Oxford University Press, (2007).

Institut für Mathematik, Universität Zürich, Zürich, Switzerland
E-mail address: `guido.dephilippis@math.uzh.ch`

Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, 91405 Orsay Cedex, France
E-mail address: `alpar.meszaros@math.u-psud.fr`

Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, 91405 Orsay Cedex, France
E-mail address: `filippo.santambrogio@math.u-psud.fr`

Dipartimento di Matematica, Università di Pisa, Pisa, Italy
E-mail address: `b.velichkov@sns.it`