

Doubly nonlinear diffusive PDEs: new existence results via generalized Wasserstein gradient flows

Thibault Caillet¹ and Filippo Santambrogio¹

¹Universite Claude Bernard Lyon 1, ICJ UMR5208, CNRS, Ecole Centrale de Lyon, INSA Lyon, Université Jean Monnet, 69622 Villeurbanne, France. `caillet,santambrogio@math.univ-lyon1.fr`

February 3, 2024

Abstract

We prove an existence result for a large class of PDEs with a nonlinear Wasserstein gradient flow structure. We use the classical theory of Wasserstein gradient flow to derive an EDI formulation of our PDE and prove that under some integrability assumptions on the initial condition the PDE is satisfied in the sense of distributions.

1 Introduction

The most classical theory of gradient flows is concerned with evolution equations of the form

$$x'(t) = -\nabla \mathcal{F}(x(t)),$$

for instance when x is a curve valued in a certain Hilbert space H and $\mathcal{F} : H \rightarrow \mathbb{R}$ is a given function on such a space. It is well-known that a natural time-discretization of this equation takes the form of a sequence of optimization problems, where we define, for a given time step $\tau > 0$, a sequence $(x_k)_k$ of points in H via

$$x_{k+1} = \operatorname{argmin} \left\{ \mathcal{F}(x) + \frac{\|x - x_k\|^2}{2\tau} \right\}.$$

This corresponds to the well-known implicit Euler scheme for the above equation, but has the advantage that it admits a formulation in metric spaces, where the norm $\|x - x_k\|$ is replaced by a generic distance. Such an iterated minimization procedure is usually called *minimizing movement scheme* (see [8]).

If one replaces the power 2 in these optimization problems with another power, the natural scaling is the following

$$x_{k+1} = \operatorname{argmin} \left\{ \mathcal{F}(x) + \frac{\tau}{p} \left(\frac{\|x - x_k\|}{\tau} \right)^p \right\}$$

and leads to a solution of

$$x'(t) = -(\nabla \mathcal{F}(x(t)))^{q-1},$$

where for a vector v and an exponent $\alpha > 0$ we denote by v^α the vector whose direction is the same as that of v and whose norm is $\|v\|^\alpha$, i.e. $\|v\|^{\alpha-1}v$, and the exponent q is the conjugate exponent of p . As a consequence, we do not face a true gradient flow, but a different variational evolution that we call *generalized gradient flow*. A similar construction could be made in metric spaces, of course, where the notion of (generalized) gradient flow should be suitably defined (and we refer to [3] or [17] for the details).

A particular case of interest is that of the Wasserstein spaces of probability measures endowed with the distances W_p induced by optimal transport. These distances are now used in a variety of contexts, from PDEs to machine learning, and we will not provide here extra details on them. The main facts that we will use will be presented in Section 2.2, and the interested reader is encouraged to refer to [20] or [18]. Each distance W_p being defined as the p -root of a transport cost, it is natural to use each W_p only in the framework of a generalized gradient flow with exponent p .

The procedure consisting in studying the limits (as $\tau \rightarrow 0$) of the following iterated optimization problems

$$\rho_{k+1} = \operatorname{argmin} \left\{ \mathcal{F}(\rho) + \frac{W_2^2(\rho, \rho_k)}{2\tau} \right\}$$

has been introduced in [13], is now known as *Jordan-Kinderlehrer-Otto* scheme, and is widely used to study PDEs of the form

$$\partial_t \rho = \nabla \cdot \left(\rho \nabla \left(\frac{\delta \mathcal{F}}{\delta \rho} \right) \right),$$

where $\frac{\delta \mathcal{F}}{\delta \rho}$ denotes the first variation of the functional \mathcal{F} . When $\mathcal{F}(\rho) = \int f(\rho(x)) dx$ this gives rise to a parabolic PDE of the form

$$\partial_t \rho = \nabla \cdot (\rho \nabla (f'(\rho))).$$

In this work, we investigate instead the general case $p \neq 2$. This consists in considering the iterated optimization problems

$$\rho_{k+1} = \operatorname{argmin} \left\{ \mathcal{F}(\rho) + \frac{W_p^p(\rho, \rho_k)}{p\tau^{p-1}} \right\},$$

that we will call p -JKO scheme. The PDE has now the form

$$\partial_t \rho = \nabla \cdot \left(\rho \left(\nabla \left(\frac{\delta \mathcal{F}}{\delta \rho} \right) \right)^{q-1} \right).$$

When \mathcal{F} has the form $\int f(\rho(x)) dx$ for a convex integrand f , this equation becomes

$$\partial_t \rho = \Delta_q (h(\rho)),$$

where h is a function such that $h'(s) = s^{p-1} f''(s)$ and, again, p and q are conjugate exponents. Equivalently, one can also write this as

$$\partial_t b(u) = \Delta_q (u)$$

with $b = h^{-1}$. This is why these equations are called *doubly non-linear*. They are anyway quasilinear, since the right-hand side is linear w.r.t. the second derivatives in space. However, the nonlinearity in terms of the gradient is source of difficulty.

Among the first studies on these equations we mention [2] where strong structural assumptions were required to prove existence and uniqueness, but we are more interested in the works which already attacked these equations in terms of generalized gradient flows and p -JKO schemes. The founding papers on this variational approach to these equations are [1] and [16], where it is proven that the discrete solutions obtained by the p -JKO scheme converge, under some conditions, to a solution of the PDE, in weak sense.

Both [1] and [16] use estimates obtained from the scheme in order to prove the necessary compactness of both the density ρ and its gradient $\nabla\rho$ and pass to the limit the discrete optimality conditions at each step. They require some assumptions on the function f : [16] only considers the case of powers, i.e. $f(z) = \frac{z^m}{m-1}$, and requires $m > 0$; [1] requires f to satisfy Mc Cann's condition for geodesic convexity (see [15]). These restrictions on f are disappointing since the case of the parabolic q -Laplacian equation $\partial_t\rho = \Delta_q\rho$ requires to use a function f such that $f''(z) = cz^{-p}$, i.e. $f(z) \approx z^{2-p}$, which could be out of the set of assumptions of these papers.

Our goal in the present paper is to generalize the results to more generic convex functions f , and at the same time to change the strategy of the proof. Instead of relying on fine arguments for the strong convergence of the gradient, we will make use of the *Energy-Dissipation Principle* and prove an *Energy Dissipation Inequality on the limit*. The idea, well-known in gradient flows (see, for instance [3]) is the following: a curve $t \mapsto x(t)$ is a solution of $x' = -\nabla\mathcal{F}(x)$ on $[0, T]$ if and only if we have

$$\mathcal{F}(x(T)) + \int_0^T \left(\frac{1}{2} \|x'(t)\|^2 + \frac{1}{2} \|\nabla\mathcal{F}(x(t))\|^2 \right) dt \leq \mathcal{F}(x(0)). \quad (1.1)$$

This can be seen from the fact that the chain rule provides $\frac{d}{dt}(\mathcal{F}(x(t))) = \nabla\mathcal{F}(x(t)) \cdot x'(t)$ and, by Young's inequality, the opposite inequality of (1.1) is always true and the only equality case is given by $x' = -\nabla\mathcal{F}(x)$.

Analogously, we have $x'(t) = -(\nabla\mathcal{F}(x(t)))^{q-1}$ if and only if we have $\mathcal{F}(x(T)) + \int_0^T \left(\frac{1}{p} \|x'(t)\|^p + \frac{1}{q} \|\nabla\mathcal{F}(x(t))\|^q \right) dt \leq \mathcal{F}(x(0))$. The corresponding condition in the Wasserstein space when considering $\mathcal{F}(\rho) := \int f(\rho(x))dx$ would be: a curve $t \mapsto \rho_t$, solution of $\partial_t\rho + \nabla \cdot (\rho v) = 0$, is a solution of the PDE if and only if we have

$$\mathcal{F}(\rho(T)) + \frac{1}{p} \int_0^T \int_{\Omega} \rho |v|^p dx dt + \frac{1}{q} \int_0^T \int_{\Omega} \rho |\nabla(f'(\rho))|^q dx dt \leq \mathcal{F}(\rho(0)).$$

Proving such an inequality is possible, up to some technicalities, using suitable estimates on the discrete scheme and semicontinuity arguments. The important point is that it is in general hard to prove that this characterizes the solutions of the PDE, i.e. the condition $v = -\nabla f'(\rho)$, since it is hard to prove that we do have $\frac{d}{dt}\mathcal{F}(\rho_t) = \int \nabla(f'(\rho_t)) \cdot v_t d\rho_t$, i.e. proving a chain rule in the Wasserstein space.

The only other paper, to the best of our knowledge, where the EDI condition is used to characterize solutions of these doubly nonlinear PDEs is [7], whose goal is to prove the convergence of a numerical

scheme. The chain rule is there proven under the assumption that \mathcal{F} is geodesically convex, i.e. f satisfies Mc Cann's condition. One of the hard tasks of the present work is to remove this assumption, and this is done by decomposing, after some approximation, an arbitrary function f into the difference of two functions satisfying Mc Cann's condition. If we do not need to perform such a decomposition, i.e. if f itself satisfies Mc Cann's condition, then the only assumption that we will require on the initial datum ρ_0 will be $\int f(\rho_0) < +\infty$, which is natural. If not, then we need a certain L^α summability, for an exponent α depending on the dimension d , on the exponent p , and on the function f .

This condition on the integrability of the initial datum is an improvement compared to [1], where most of the paper is presented under the assumption that ρ_0 is bounded from above and below by strictly positive constants, and then completed by a discussion on how to remove the lower bounds by approximation and then the upper bound by replacing it with $\rho_0 \in L^p$ (the same p as the the exponent in the Wasserstein space). Our exponent α could be smaller than p but, in particular, we do not need it when f satisfies Mc Cann's condition, as it is the case in [1].

We stress on the other hand two points where our paper is less general than the previous ones, essentially for the sake of simplicity:

- After performing very long computations we decided to stick to the case $\mathcal{F}(\rho) = \int f(\rho(x))dx$ and ignore the more general case $\mathcal{F}(\rho) = \int f(\rho(x))dx + \int Vd\rho$ where a potential energy is added. Such a case is, instead, considered in [1, 7]. Yet, it makes the computations much harder in the case where \mathcal{F} is not geodesically convex, and we believe the interest for these equations is limited, because, differently from the case $p = 2$, the gradient of the potential does not act as a linear drift in the equation.
- We also decided to ignore the case where the Wasserstein distance W_p are replaced by transport costs which have not the homogeneity of a power. Indeed, [1] considers a JKO-like scheme obtained by looking at

$$\rho_{k+1} = \operatorname{argmin} \{ \mathcal{F}(\rho) + \tau \mathcal{T}_{c,\tau}(\rho, \rho_k) \},$$

where $T_{c,\tau}$ is the minimal transport cost associated with the cost $(x, y) \mapsto c((x - y)/\tau)$. In this case the equation we obtain is

$$\partial_t \rho = \nabla \cdot \left(\rho \nabla c^* \left(\frac{\delta F}{\delta \rho} \right) \right).$$

For an example of a very interesting PDE with this structure we cite [14] on the so-called *relativistic heat equation*, where $c(v) = 1 - \sqrt{1 - |v|^2}$.

The main result of our paper is the following theorem, where we use the notation $L_f(z) = z f'(z) - f(z)$, so that the function L_f satisfies $L'_f(z) = z f''(z)$ and, formally, $\nabla L_f(\rho) = \rho \nabla(f'(\rho))$.

Theorem 1.1. *Let $f : \mathbb{R}_+ \rightarrow \mathbb{R} \cup \{+\infty\}$ be a l.s.c. convex function such that $f \in C^2(0, +\infty)$, $f'' > 0$, and $f''(z) \geq Cz^{-\theta}$ when z is large enough for some $\theta \in \mathbb{R} \cup \{+\infty\}$. Let $\varrho_0 \in L^1(\Omega)$ be a probability density satisfying*

$$\int_{\Omega} f(\varrho_0) dx < \infty.$$

If f does not satisfy McCann's condition, set $\alpha = 2 - q(1 + \frac{1}{d}) + \theta(q - 1)$, and assume as well $\varrho_0 \in L^\alpha$ whenever $\alpha > 1$, or $\int_\Omega \varrho_0 \log \varrho_0 \, dx < +\infty$ if $\alpha = 1$.

Then there exists $\varrho \in C^0([0, T], \mathbb{W}_p(\Omega))$ and $v \in L^p(\varrho_t \, dx \, dt)$ such that

$$\begin{cases} \partial_t \varrho_t + \nabla \cdot (\varrho_t v_t) = 0 \text{ in the sense of distributions,} \\ v_t = - \left(\frac{\nabla L_f(\varrho_t)}{\varrho_t} \right)^{q-1} \varrho_t \text{ a.e. for a.e. } t, \\ \varrho(0) = \varrho_0. \end{cases} \quad (1.2)$$

We underline that the summability assumption on ρ_0 (which is void if $\alpha < 1$) only plays a role in case f does not satisfy McCann's condition.

To achieve the goal of proving the above theorem, the paper is organized as follows. Section 2 is devoted to some preliminaries: notation (2.1), bases of optimal transport, including some considerations on the McCann's condition that we mentioned many times so far (2.2), and the details of an approximation procedure that we will use for every step of the p -JKO scheme, i.e. replacing f with f_ε defined via $f_\varepsilon(z) = f(z) + \varepsilon z \log z$. In section 3 we prove that the solutions obtained in the iterated minimization converge, up to subsequences, to a curve ρ_t satisfying the EDI condition. This requires suitable interpolations of the discrete sequence (3.1). The proof of the compactness and of the limit inequality is contained in Section 3.2, and in Section 3.3 we prove additional integrability properties on the limit curve.

Section 4 is devoted to the hard task of proving the chain rule and is divided into several steps, where we prove the desired results on arbitrary curves satisfying some integrability assumptions. Since the curve obtained in Section 3 satisfies them, at the end of the section we can prove Theorem 1.1.

We also prove additional estimates on the solution constructed as a limit of the p -JKO scheme, in particular in terms of its BV norm. This is a consequence of the so-called *five gradients inequality* which has been recently generalized to the case of the transport cost $|x - y|^p$ in [6]. This is the object of Section 5.

A short appendix concludes the paper, where a technical point in the discretization of the integral in time in Section 4 is detailed.

Acknowledgments The authors were supported by the European Union via the ERC AdG 101054420 EYAWKAJKOS project.

2 Notations and Preliminaries

2.1 Notations

Below is a list of notations we will use in the sequel :

- $d \in \mathbb{N}^*$ is the space dimension,
- Ω is an open convex bounded subset of \mathbb{R}^d ,
- $T > 0$ is a set time horizon and $\Omega_T = \Omega \times [0, T]$,

- $p > 1$ and $q > 1$ are dual integrability exponents : $\frac{1}{p} + \frac{1}{q} = 1$,
- W_p is the p -Wasserstein distance.
- If $z \in \mathbb{R}^d$ is any vector and $\alpha > 1$, we write $z^\alpha = |z|^{\alpha-1}z$
- For a differentiable function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$, L_f is the function defined by $L_f(z) = zf'(z) - f(z)$
- For a function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$, $\mathcal{F} : \mathcal{P}(\Omega) \rightarrow \mathbb{R} \cup \{+\infty\}$ is its associated functional according to definition 2.4. We use similarly $\mathcal{H}, \tilde{\mathcal{F}}, \tilde{\mathcal{F}}_1, \tilde{\mathcal{F}}_2, \mathcal{L}$ for the functionals associated with the functions $h, \tilde{f}, \tilde{f}_1, \tilde{f}_2, \ell$ respectively.

For a function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$, we define the following assumptions which we will use throughout this paper :

- **(H1)** f is convex and lower semi-continuous
- **(H2)** $f \in C^2((0, +\infty))$
- **(H3)** $f'' > 0$
- **(H4)** $f''(z) \geq Cz^{-\theta}$ when z is large enough
- **(G1)** (Mc-Cann's condition) $f(0) = 0$ and $(0, +\infty) \ni s \mapsto s^d f(s^{-d})$ is convex and non increasing.

2.2 Preliminaries on Optimal transport, Gradient Flows and Geodesic Convexity

We refer to [3, 18, 19] for the general theory of optimal transport and its application to gradient flows, and will compile below a selection of helpful facts we will use in the sequel.

Theorem 2.1. *Let $\varrho, g \in \mathcal{P}(\Omega)$ be two probabilities on Ω . The following statements are classical :*

1. *The problem*

$$W_p^p(\mu, \nu) := \min \left\{ \int_{\Omega \times \Omega} |x - y|^p d\gamma \quad ; \quad \gamma \in \Pi(\mu, \nu) \right\},$$

where $\Pi(\mu, \nu)$ is the set of probabilities on $\Omega \times \Omega$ with first marginal μ and second marginal ν , admits a solution γ^* . If $\mu = \varrho dx$, with $\varrho \in L^1(\Omega)$, then the solution is unique, and given by $\gamma^* = (\text{id}, T)_{\#}\varrho$ for some $T : \Omega \rightarrow \Omega$ which is called the optimal transport map.

2. *We have*

$$\frac{1}{p} W_p^p(\mu, \nu) = \max \left\{ \int_{\Omega} \varphi d\mu + \int_{\Omega} \psi d\nu \quad ; \quad \varphi(x) + \psi(y) \leq \frac{1}{p} |x - y|^p \quad \forall x, y \in \Omega \right\}.$$

The optimal φ we call Kantorovich potentials, and they can be taken to be c -concave, meaning of the form

$$\varphi(x) = \inf_{y \in \Omega} \frac{1}{p} |x - y|^p - \psi(y), \quad \psi(y) = \varphi^c(y) := \inf_{x \in \Omega} \frac{1}{p} |x - y|^p - \varphi(x),$$

from which one can prove that φ and φ^c are Lipschitz.

3. The optimal transport map T is given by

$$T(x) = x - (\nabla\varphi(x))^{q-1},$$

where φ is any Kantorovich potential. It is well defined almost everywhere since a Lipschitz function is differentiable almost everywhere.

4. W_p is a distance on $\mathcal{P}(\Omega)$ which metrizes the weak convergence of measures. Endowed with this metric, we call $\mathcal{P}(\Omega)$ the Wasserstein Space, which we denote by $\mathbb{W}_p(\Omega)$

5. $\mathbb{W}_p(\Omega)$ is a geodesic space, and the geodesic between μ and ν is given by $\mu_t := (\pi_t)_\# \gamma$, where $\pi_t(x, y) = (1-t)x + ty$ and γ is an optimal transport plan between μ and ν . If γ is given by an optimal transport map T , then $\mu_t = ((1-t)\text{id} + tT)_\# \mu$.

We will now define the Benamou-Brenier functional which will be useful in the sequel. We take $p \in (1, +\infty)$, and set $K_p = \left\{ (a, b) \in \mathbb{R} \times \mathbb{R}^d; a + \frac{1}{q}|b|^q \leq 0 \right\}$, and for $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$,

$$B_p(t, x) = \begin{cases} \frac{1}{p} \frac{|x|^p}{t^{p-1}}, & \text{if } t > 0, \\ 0, & \text{if } t = 0, x = 0, \\ +\infty & \text{if } t = 0, x \neq 0, \text{ or } t < 0. \end{cases}$$

Definition 2.2 (Benamou-Brenier functional). If X is a compact measurable space and $(\varrho, E) \in \mathcal{M}_b(X) \times \mathcal{M}_b(X)^d$,

$$\mathcal{B}_p(\varrho, E) := \sup \left\{ \int a \, d\varrho + \int b \cdot dE; (a, b) \in C_b(X, K_p) \right\}.$$

Proposition 2.3. \mathcal{B}_p is convex and lower semi-continuous on the space $\mathcal{M}_b(X) \times \mathcal{M}_b(X)^d$ for the weak convergence, and moreover we have

1. $\mathcal{B}_p(\varrho, E) \geq 0$,
2. If ϱ and E are both absolutely continuous with respect to a positive measure λ , then $\mathcal{B}_p(\varrho, E) = \int B_p(\varrho, E) \, d\lambda$,
3. $\mathcal{B}_p(\varrho, E) < \infty$ only if $\varrho \geq 0$ and $E \ll \varrho$,
4. In that case, we can write $E = v \cdot \varrho$ and $\mathcal{B}_p(\varrho, E) = \int \frac{1}{p} |v|^p \, d\varrho$

Definition 2.4. Let $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ be a convex and l.s.c. function, and let $L := \lim_{t \rightarrow +\infty} \frac{f(t)}{t}$. We define

$$\mathcal{F}(\mu) := \int_{\Omega} f(\varrho(x)) \, dx + L\mu^S(\bar{\Omega}),$$

with $\mu = \varrho \cdot dx + \mu^S$, where ϱ and μ^S are respectively the absolutely continuous (with respect to the Lebesgue measure) and the singular part of μ . It is a convex and l.s.c. functional for the weak convergence of measures.

We will similarly use the notations $\mathcal{H}, \tilde{\mathcal{F}}, \tilde{\mathcal{F}}_1, \tilde{\mathcal{F}}_2, \mathcal{L}$ to denote the functionals associated with the convex functions $h, \tilde{f}, \tilde{f}_1, \tilde{f}_2, \ell$ respectively.

Proposition 2.5. If f verifies **(G1)**, then the functional \mathcal{F} is geodesically convex in \mathbb{W}_p , i.e. it is convex along geodesics in \mathbb{W}_p .

Lemma 2.6. Let $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a superlinear convex function which is 0 in a neighborhood of 0. There exists a smooth strictly convex superlinear function $\tilde{\Phi}$ satisfying **(G1)** and a constant $C > 0$ such that $\Phi \leq C(\varphi + 1)$

Proof. We look at the function

$$\mathcal{M}\phi(s) := s^d \phi(s^{-d}),$$

and define $\tilde{\Phi}$ to be its lower convex hull. Since ϕ is 0 near 0, we deduce that $\tilde{\Phi}$ is eventually constantly 0, which implies that $\tilde{\Phi}$ is decreasing, being a convex function. We take

$$\Phi(z) = \mathcal{M}^{-1}\tilde{\Phi}(z) = \tilde{\Phi}(z^{-\frac{1}{d}})z,$$

and find that Φ is superlinear if $\tilde{\Phi}(0^+) = +\infty$. Supposing that this is not the case, i.e. that there exists some $M > 0$ such that $\tilde{\Phi} \leq M$ near 0, we set

$$a = \sup_{s>0} \frac{M + 1 - \mathcal{M}\phi(s)}{s} < +\infty,$$

which is well defined because since ϕ is superlinear, $\mathcal{M}\phi(0^+) = +\infty$. We therefore have that

$$\mathcal{M}\phi(s) \geq M + 1 - as,$$

so that the lower convex hull of $\mathcal{M}\phi$, being the supremum of affine functions lower than itself, verifies $\tilde{\Phi}(0^+) \geq M + 1$ when we take $s \rightarrow 0$ which yields a contradiction. Up to finding a smooth approximation of $\tilde{\Phi}$, for example via convolution, we can also assume that $\tilde{\Phi}$ is smooth. It is known (see for example [19] chapter 17) that a superlinear convex function satisfying Mc-Cann's condition **(G1)** is strictly convex away from 0, therefore up to adding a convex function to $\tilde{\Phi}$, that is linear near 0 (which means adding a sublinear function to φ near $+\infty$) and that strictly decreases to a constant thereafter (which keeps Φ bounded near 0), we can guarantee that $\Phi = \mathcal{M}^{-1}\tilde{\Phi}$ is strictly convex everywhere, and that $\Phi \leq C(\varphi + 1)$ for some constant $C > 0$. \square

2.3 Approximated problem

In this first section we will look at the following variational problem defining the next step in the JKO scheme we will use later :

$$\varrho \in \operatorname{argmin} \frac{W_p^p(\varrho, g)}{p\tau^{p-1}} + \mathcal{F}(\varrho). \quad (2.1)$$

We will assume that f satisfies **(H1,H2,H3)**. To gain some regularity for the minimizers of the problem, we will approximate it by adding a small entropy term.

$$\mu_\varepsilon \in \operatorname{argmin}_\mu \frac{W_p^p(\mu, g)}{p\tau^{p-1}} + \mathcal{F}_\varepsilon(\mu) \quad (2.2)$$

where \mathcal{F}_ε is the functional associated with $f_\varepsilon(z) = f(z) + \varepsilon z \log(z)$, i.e.

$$\mathcal{F}_\varepsilon(\mu) = \mathcal{F}(\mu) + \begin{cases} \varepsilon \int_{\Omega} \varrho \log(\varrho) \, dx & \text{if } \mu = \varrho \, dx \\ +\infty & \text{otherwise.} \end{cases}$$

Lemma 2.7. *The solution of 2.2 is given by a density that is bounded from above and away from 0, and is actually Lipschitz continuous.*

Proof. An easy adaptation of the proof from [4], appendix B.4 shows that the conditions $f_\varepsilon(0^+) = -\infty$ and $f'_\varepsilon(+\infty) = +\infty$ implies that there exists $\delta > 0$ such that the solution ϱ_ε of 2.2 satisfies $\delta \leq \varrho_\varepsilon \leq \frac{1}{\delta}$. Since f''_ε is bounded from below on compact sets, using the optimality condition (see [18] section 7.2.3 for its derivation), we have

$$\frac{\varphi}{p\tau^{p-1}} + f'_\varepsilon(\varrho_\varepsilon) = C \text{ a.e.,}$$

from which we deduce that ϱ_ε is a Lipschitz function. Indeed, we can write

$$\varrho_\varepsilon = (f'_\varepsilon)^{-1} \left(C - \frac{\varphi}{p\tau^{p-1}} \right),$$

and since φ is Lipschitz we deduce the result. \square

Lemma 2.8 (Flow Interchange Technique). *Let $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ satisfy (H1-H2-H3-G1) and ϱ_ε be the solution of 2.2. Define \mathcal{H} as in definition 2.4. If $\mathcal{H}(g) < +\infty$ then we have*

$$\mathcal{H}(g) - \mathcal{H}(\varrho_\varepsilon) \geq \tau \int_{\Omega} \nabla(L_h(\varrho_\varepsilon)) \cdot (\nabla f'_\varepsilon(\varrho_\varepsilon))^{q-1} \, dx \geq 0.$$

Proof. Assumption (G1) implies that \mathcal{H} is convex along geodesics, therefore we will look at the geodesic $[0, 1] \ni t \mapsto \varrho_t$ from ϱ_ε to g . By convexity we have

$$\mathcal{H}(g) - \mathcal{H}(\varrho_\varepsilon) \geq \lim_{t \rightarrow 0} \frac{\mathcal{H}(\varrho_t) - \mathcal{H}(\varrho_\varepsilon)}{t}.$$

Following [3] section 10.4.3, we find that

$$\lim_{t \rightarrow 0} \frac{\mathcal{H}(\varrho_t) - \mathcal{H}(\varrho_\varepsilon)}{t} \geq - \int_{\Omega} L_h(\varrho_\varepsilon) \text{Tr} \left(\tilde{\nabla}(T - id) \right) \, dx,$$

where T is the optimal transport map from ϱ_ε to g and $\tilde{\nabla}T$ denotes its approximate gradient. Again adapting the proof of Theorem 10.4.5 of [3], up to approximating T with BV functions with nonnegative distributional divergence and using that $L_h(\varrho_\varepsilon) \geq 0$ is $W^{1,1}(\Omega)$, we have

$$- \int_{\Omega} L_h(\varrho_\varepsilon) \text{Tr} \left(\tilde{\nabla}(T - id) \right) \, dx \geq \int_{\Omega} \nabla(L_h(\varrho_\varepsilon)) \cdot (T - id) \, dx - \int_{\partial\Omega} L_h(\varrho_\varepsilon)(T - id) \cdot n \, d\sigma$$

Using the convexity of the domain and the fact that T points inwards ϱ_ε a.e. and $L_h(0) = 0$, the boundary integral is non positive. Finally, using the optimality conditions for ϱ_ε we have

$$T - id = -(\nabla\varphi)^{q-1} = \tau (\nabla f'_\varepsilon(\varrho_\varepsilon))^{q-1},$$

and we obtain

$$\begin{aligned} - \int_{\Omega} L_h(\varrho_\varepsilon) \text{Tr} \left(\tilde{\nabla}(T - id) \right) dx &\geq \tau \int_{\Omega} \nabla(L_h(\varrho_\varepsilon)) \cdot \nabla (f'_\varepsilon(\varrho_\varepsilon))^{q-1} dx \\ &= \tau \int_{\Omega} h''(\varrho_\varepsilon) f''_\varepsilon(\varrho_\varepsilon)^{q-1} |\nabla \varrho_\varepsilon|^q d\varrho_\varepsilon \geq 0 \end{aligned}$$

□

Lemma 2.9. *We have Γ -convergence of the functionals $\mathcal{F}_\varepsilon \xrightarrow[\varepsilon \rightarrow 0]{\Gamma} \mathcal{F}$ for the topology of weak convergence of measures.*

Proof. Since $z \mapsto z \log(z)$ is bounded from below, up to adding a constant we can assume that the sequence $(\mathcal{F}_\varepsilon)_\varepsilon$ is nonincreasing as $\varepsilon \rightarrow 0$. This monotonicity ensures that the sequence Gamma converges to the lower semi-continuous relaxation of its pointwise limit (see [5])

$$\tilde{\mathcal{F}} : \mu \mapsto \mathcal{F}(\mu) + \begin{cases} 0 & \text{if } \mu = \varrho dx \text{ and } \int_{\Omega} \varrho \log(\varrho) dx < \infty \\ +\infty & \text{otherwise.} \end{cases}$$

We now have to show that the l.s.c relaxation $\text{sc}^- \tilde{\mathcal{F}}$ of this functional is indeed \mathcal{F} . First we can notice that $\mathcal{F} \leq \tilde{\mathcal{F}}$ and since \mathcal{F} is l.s.c., we have $\mathcal{F} \leq \text{sc}^- \tilde{\mathcal{F}}$. To get the opposite inequality, it is enough to find a recovery sequence of probability measures that are given by bounded densities, since bounded densities have finite entropy. We take a probability measure $\mu \in \mathcal{P}(\bar{\Omega})$ whose Lebesgue decomposition is $\mu = \varrho dx + \mu^S$ and we can assume that $\mathcal{F}(\mu) < \infty$. We define the set $A_n = \{\varrho > n\}$ and

$$\varrho_n(x) = \begin{cases} \varrho(x) & \text{when } x \in A_n^c \\ \int_{A_n} \varrho(y) dy & \text{when } x \in A_n. \end{cases}$$

We have, using dominated convergence,

$$\int_{\Omega} |\varrho(x) - \varrho_n(x)| dx = \int_{A_n} \left| \varrho(x) - \int_{A_n} \varrho(y) dy \right| dx \leq 2 \int_{A_n} \varrho(x) dx \xrightarrow{n \rightarrow \infty} 0,$$

so that ϱ_n converges to ϱ in L^1 and $\varrho_n(\Omega) = \varrho(\Omega)$. For the singular part, we approximate μ^S with linear combinations of Dirac masses supported on Ω , and then take the convolution with a standard mollifier with small enough support so that no mass escapes outside of the domain. We obtain a sequence μ_n^S of L^∞ densities such that $\mu_n^S \rightarrow \mu^S$ as measures, and $\mu_n(\Omega) = \mu(\Omega)$. We have therefore constructed a sequence of probability measures $\mu_n = \varrho_n dx + \mu_n^S$ that converges to μ , and we have, using the convexity of f ,

$$\tilde{\mathcal{F}}(\mu_n) = \int_{\Omega} f(\varrho_n(x) + \mu_n^S(x)) dx \leq \int_{\Omega} f(\varrho_n) dx + L \int_{\Omega} \mu_n(x) dx.$$

For the first term, we use Jensen's inequality to get

$$\begin{aligned} \int_{\Omega} f(\varrho_n) dx &= \int_{A_n^c} f(\varrho) dx + \int_{A_n} f \left(\int_{A_n} \varrho(y) dy \right) dx \\ &\leq \int_{A_n^c} f(\varrho) dx + \int_{A_n} \left(\int_{A_n} f(\varrho(y)) dy \right) dx = \int_{\Omega} f(\varrho) dx. \end{aligned}$$

Remembering that we constructed μ_n^S so that the mass of the singular part was conserved, we therefore obtain

$$\limsup_n \tilde{\mathcal{F}}(\mu_n) \leq \int_{\Omega} f(\varrho) dx + L\mu^S(\bar{\Omega}) = \mathcal{F}(\mu),$$

so that $\text{sc}^- \tilde{\mathcal{F}} = \mathcal{F}$ □

Lemma 2.10. *The solution μ_ε of the approximated problem 2.2 weakly converges as measures to the solution μ of 2.1 when $\varepsilon \rightarrow 0$.*

Proof. Using Lemma 2.9, and the fact that the Wasserstein distance metrizes the weak convergence of measures, for a given probability measure $g \in \mathcal{P}(\bar{\Omega})$, we have

$$\mu \mapsto \frac{W_p^p(\mu, g)}{p\tau^{p-1}} + \mathcal{F}_\varepsilon(\mu) \xrightarrow{\varepsilon \rightarrow 0} \mu \mapsto \frac{W_p^p(\mu, g)}{p\tau^{p-1}} + \mathcal{F}(\mu).$$

Since $\bar{\Omega}$ is compact, $\mathcal{P}(\bar{\Omega})$ is compact for the weak convergence of probability measures so that we can assume that the sequence of minimums weakly converge. Because the limit problem has a unique solution μ , we deduce that the sequence of minima μ_ε converges to μ . □

Remark 2.11. In the sequel, we will always have $g \in L^1(\Omega)$. As a consequence of Lemmas 2.6, 2.8 and 2.10, one can show that the solution of problem 2.1 is an absolutely continuous measure $\mu = \varrho dx$ (see the proof of Lemma 3.6). We shall therefore use the usual abuse of notation of denoting by ϱ any measure $\mu = \varrho dx$ when the context is clear.

3 Energy Dissipation Condition and properties of the limit curve

3.1 Interpolations

We look at the problem

$$\varrho_{k+1}^\tau \in \operatorname{argmin} \frac{W_p^p(\varrho, \varrho_k^\tau)}{p\tau^{p-1}} + \mathcal{F}(\varrho) \tag{3.1}$$

Under our hypotheses, this problem has a unique solution and the sequence $(\varrho_k)_{k \in \mathbb{N}}$ can be defined recursively.

We can define multiple interpolations which are all useful for our analysis :

Definition 3.1 (Piecewise constant interpolation). We define the piecewise constant interpolation as a piecewise constant curve $\bar{\varrho}_t^\tau$ and an associated velocity \bar{v}_t^τ : for $t \in (k\tau, (k+1)\tau]$,

$$\begin{aligned}\bar{\varrho}_t^\tau &= \varrho_{k+1}^\tau, \\ \bar{v}_t^\tau &= v_k^\tau := \frac{\text{Id} - T_k}{\tau},\end{aligned}$$

where T_k is the optimal transport map from ϱ_k^τ to ϱ_{k+1}^τ . We also define the piecewise constant momentum variable $\bar{E}_t^\tau = \bar{\varrho}_t^\tau \bar{v}_t^\tau$

Definition 3.2 (Piecewise geodesic interpolation). We define the piecewise geodesic interpolation as a piecewise geodesic curve $\tilde{\varrho}_t^\tau$ and an associated velocity \tilde{v}_t^τ : for $t \in (k\tau, (k+1)\tau]$,

$$\begin{aligned}\tilde{\varrho}_t^\tau &= (\text{Id} - (k\tau - t)v_k^\tau)_\# \varrho_k^\tau, \\ \tilde{v}_t^\tau &= v_k^\tau \circ (\text{Id} - (k\tau - t)v_k^\tau)^{-1},\end{aligned}$$

which is constructed to satisfy the continuity equation

$$\partial_t \tilde{\varrho}_t^\tau + \nabla \cdot (\tilde{\varrho}_t^\tau \tilde{v}_t^\tau) = 0.$$

Also, we have for a.e. t ,

$$\|\tilde{v}_t^\tau\|_{L^p(\tilde{\varrho}_t^\tau)} = |(\tilde{\varrho}^\tau)'|(t) = \frac{W_p(\varrho_{k+1}^\tau, \varrho_k^\tau)}{\tau},$$

and therefore the following holds :

$$\frac{\tau}{p} \|\tilde{v}_t^\tau\|_{L^p(\tilde{\varrho}_t^\tau)} = \frac{W_p^p(\varrho_{k+1}^\tau, \varrho_k^\tau)}{p\tau^{p-1}}. \quad (3.2)$$

We also define the piecewise geodesic momentum variable $\tilde{E}_t^\tau = \tilde{\varrho}_t^\tau \tilde{v}_t^\tau$

Definition 3.3 (De Giorgi variational interpolation). We define the De Giorgi variational interpolation as a curve $\hat{\varrho}_t^\tau$: for $t \in (k\tau, (k+1)\tau]$, we define $s = \frac{t-k\tau}{\tau} \in (0, 1]$ and

$$\hat{\varrho}_t^\tau = \operatorname{argmin} \frac{W_p^p(\varrho, \varrho_k^\tau)}{p(s\tau)^{p-1}} + \mathcal{F}(\varrho). \quad (3.3)$$

In particular for $t = (k+1)\tau$ we have $s = 1$ and we do retrieve $\hat{\varrho}_{k+1}^\tau = \varrho_{k+1}^\tau$, and if $t = k\tau$, $s = 0$ and the minimizer has to be $\varrho = \varrho_k^\tau$.

This interpolation is useful in deriving the following precursor to the EDI interpretation of our gradient flow :

$$\mathcal{F}(\varrho_k^\tau) \geq \mathcal{F}(\varrho_{k+1}^\tau) + \frac{\tau}{q} \int_{k\tau}^{(k+1)\tau} \frac{W_p^p(\hat{\varrho}_t^\tau, \varrho_k^\tau)}{s(s\tau)^{p-1}} dt + \frac{W_p^p(\varrho_{k+1}^\tau, \varrho_k^\tau)}{p\tau^{p-1}}. \quad (3.4)$$

Indeed if we define the function

$$g : t \in (k\tau, (k+1)\tau] \mapsto \min_{\varrho} \frac{W_p^p(\varrho, \varrho_k^\tau)}{p(s\tau)^{p-1}} + \mathcal{F}(\varrho),$$

we find that g is an nonincreasing function of t , and therefore differentiable almost everywhere. Furthermore, we have a one sided fundamental theorem of analysis :

$$\begin{aligned} \int_{k\tau}^{(k+1)\tau} g'(t) dt &\geq g((k+1)\tau) - g(k\tau) \\ &= \frac{W_p^p(\varrho_{k+1}^\tau, \varrho_k^\tau)}{p\tau^{p-1}} + \mathcal{F}(\varrho_{k+1}^\tau) - \mathcal{F}(\varrho_k^\tau). \end{aligned}$$

By the envelope theorem, at differentiability points of g we necessarily have

$$g'(t) = \frac{1}{q} \frac{W_p^p(\hat{\varrho}_t^\tau, \varrho_k^\tau)}{(s\tau)^p},$$

which gives 3.4 when combined with the previous result. We can now use the optimality conditions for 3.3 : if we denote by T_s and φ_s respectively the optimal transport map and a Kantorovich potential between $\hat{\varrho}_t^\tau$ and ϱ_k^τ , we have

$$\frac{\nabla \varphi_s}{(s\tau)^{p-1}} = -\nabla f'(\hat{\varrho}_s^\tau)$$

which we can replace in the formula for W_p^p using the definition of the optimal transport map :

$$T_s - \text{Id} = -(\nabla \varphi_s)^{q-1},$$

and obtain

$$\frac{1}{q} \frac{W_p^p(\hat{\varrho}_t^\tau, \varrho_k^\tau)}{(s\tau)^p} = \frac{1}{q} \int_{\Omega} \frac{|x - T_s|^p}{(s\tau)^p} d\hat{\varrho}_s^\tau = \frac{1}{q} \int_{\Omega} |\nabla f'(\hat{\varrho}_t^\tau)|^q d\hat{\varrho}_t^\tau.$$

Finally, using 3.2 in 3.4 yields

$$\mathcal{F}(\varrho_k^\tau) \geq \mathcal{F}(\varrho_{k+1}^\tau) + \frac{1}{q} \int_{k\tau}^{(k+1)\tau} \int_{\Omega} |\nabla f'(\hat{\varrho}_t^\tau)|^q d\hat{\varrho}_t^\tau dt + \frac{1}{p} \int_{k\tau}^{(k+1)\tau} \int_{\Omega} |\tilde{v}_t^\tau|^p d\tilde{\varrho}_t^\tau dt.$$

To easily justify the previous computations and later make it easier to pass to the limit we will use the approximation introduced in 2.2 to obtain a different (but formally equivalent) form for the slope :

$$\varrho^\varepsilon = \operatorname{argmin} \frac{W_p^p(\varrho, \varrho_k^\tau)}{p(s\tau)^{p-1}} + \mathcal{F}(\varrho) + \varepsilon \int_{\Omega} \varrho \log \varrho dx.$$

We know ϱ^ε are Lipschitz and weakly converge to $\hat{\varrho}_t^\tau$, and with the same computations as above we can deduce, using that $\varepsilon \log(\varrho^\varepsilon)$ is a increasing function of $f'(\varrho^\varepsilon)$,

$$\frac{1}{q} \frac{W_p^p(\varrho^\varepsilon, \varrho_k^\tau)}{(s\tau)^p} = \frac{1}{q} \int_{\Omega} |\nabla (f'(\varrho^\varepsilon) + \varepsilon \log(\varrho^\varepsilon))|^q d\varrho^\varepsilon \geq \frac{1}{q} \int_{\Omega} \left| \frac{\nabla L_f(\varrho^\varepsilon)}{\varrho^\varepsilon} \right|^q d\varrho^\varepsilon.$$

Since weak convergence of measures implies the convergence in Wasserstein distance, we have, using the lower semi-continuity of the slope proved in the following lemma :

$$\frac{1}{q} \frac{W_p^p(\hat{\varrho}_t^\tau, \varrho_k^\tau)}{(s\tau)^p} \geq \liminf_{\varepsilon} \frac{1}{q} \int_{\Omega} \left| \frac{\nabla L_f(\varrho^\varepsilon)}{\varrho^\varepsilon} \right|^q d\varrho^\varepsilon \geq \frac{1}{q} \int_{\Omega} \left| \frac{\nabla L_f(\hat{\varrho}_t^\tau)}{\hat{\varrho}_t^\tau} \right|^q d\hat{\varrho}_t^\tau$$

Lemma 3.4. *The functional*

$$\varrho \mapsto \frac{1}{q} \int_{\Omega} \left| \frac{\nabla L_f(\varrho)}{\varrho} \right|^q d\varrho = \mathcal{B}_q(\varrho, \nabla L_f(\varrho))$$

is lower semi-continuous with respect to the weak L^1 convergence.

Proof. Let ϱ_n be a sequence of probability measures weakly converging in L^1 to some probability measure ϱ . We can assume that $\mathcal{B}_q(\varrho_n, \nabla L_f(\varrho_n))$ is bounded and from Hölder's inequality and $\frac{q}{p} = q-1$ we get the following basic estimate :

$$\int_{\Omega} |\nabla L_f(\varrho_n)| dx = \int_{\Omega} \frac{|\nabla L_f(\varrho_n)|}{\varrho_n^{1/p}} \varrho_n^{1/p} dx \leq \left(\int_{\Omega} \frac{|\nabla L_f(\varrho_n)|^q}{\varrho_n^{q/p}} dx \right)^{\frac{1}{q}} \left(\int_{\Omega} \varrho_n(x) dx \right)^{\frac{1}{p}} \leq \mathcal{B}_q(\varrho_n, \nabla L_f(\varrho_n))^{\frac{1}{q}}.$$

Using this bound we can therefore deduce that $L_f(\varrho_n) - C_n$ is bounded in L^1 for some constants C_n , and using the compactness embedding of BV in L^1 we can deduce that up to subsequences, we have, for some $u \in BV(\Omega)$

$$\begin{aligned} L_f(\varrho_n) - C_n &\rightarrow u \text{ almost everywhere and in } L^1, \\ \nabla L_f(\varrho_n) &\rightarrow du \text{ weakly as measures.} \end{aligned}$$

This strong convergence implies that C_n is actually bounded. Indeed, if for some further subsequence we had $C_n \rightarrow \infty$, this would imply that $L_f(\varrho_n) \rightarrow \infty$ a.e. which in turn implies that $\varrho_n \rightarrow \infty$ a.e. because L_f is strictly increasing. Similarly, if $C_n \rightarrow -\infty$ for some subsequence, we would have $L_f(\varrho_n) \rightarrow -\infty$ a.e. and $\varrho_n \rightarrow 0$ a.e. Both situations are impossible in light of Lemma 3.5 below, so that we can assume that $(C_n)_n$ is bounded. Up to changing u by a constant, we can therefore assume that $L_f(\varrho_n) \rightarrow u$ a.e. and in L^1 , which implies that $\varrho_n \rightarrow (L_f)^{-1}(u)$ a.e. Again using Lemma 3.5, the weak convergence $\varrho_n \rightarrow \varrho$ then gives $u = L_f(\varrho)$ a.e. Finally, using 2.3, we get that du has a density which is $\nabla L_f(\varrho)$ and

$$\frac{1}{q} \int_{\Omega} \left| \frac{\nabla L_f(\varrho)}{\varrho} \right|^q d\varrho = \mathcal{B}_q(\varrho, \nabla L_f(\varrho)) \leq \liminf_n \mathcal{B}_q(\varrho_n, \nabla L_f(\varrho_n)) = \liminf_n \frac{1}{q} \int_{\Omega} \left| \frac{\nabla L_f(\varrho_n)}{\varrho_n} \right|^q d\varrho_n$$

□

Lemma 3.5. *Let $(u_n)_n \in L^1(\Omega)$ be a sequence of functions that weakly converges to some u in $L^1(\Omega)$, such that u_n converges pointwise to some measurable function v almost everywhere on $A \subset \Omega$, then $u = v$ almost everywhere on A .*

Proof. First by Fatou's Lemma, we find

$$\int_A |v| \, dx \leq \|u\|_{L^1(A)},$$

so that v is finite almost everywhere. By Egorov's theorem, for $\varepsilon > 0$, there exists some $A_\varepsilon \subset A$ such that $|A \setminus A_\varepsilon| \leq \varepsilon$ and u_n converges to g uniformly on A_ε . By the weak convergence, we have

$$\int_{A_\varepsilon} u_n(u - v) \, dx \rightarrow \int_{A_\varepsilon} u(u - v) \, dx.$$

By the uniform convergence, we find

$$\int_{A_\varepsilon} u_n(u - v) \, dx \rightarrow \int_{A_\varepsilon} v(u - v) \, dx,$$

so that we have

$$\int_{A_\varepsilon} (u - v)^2 \, dx = 0,$$

and $u = v$ almost everywhere on A_ε . Taking $\varepsilon \rightarrow 0$, we find $f = g$ almost everywhere on A . □

We can therefore exchange the slope term in the previous partial EDI and have

$$\mathcal{F}(\varrho_k^\tau) \geq \mathcal{F}(\varrho_{k+1}^\tau) + \frac{1}{q} \int_{k\tau}^{(k+1)\tau} \int_\Omega \left| \frac{\nabla L_f(\hat{\varrho}_t^\tau)}{\hat{\varrho}_t^\tau} \right|^q \, d\hat{\varrho}_t^\tau \, dt + \frac{1}{p} \int_{k\tau}^{(k+1)\tau} \int_\Omega |\tilde{v}_t^\tau|^p \, d\tilde{\varrho}_t^\tau \, dt.$$

Finally, summing on k we get our EDI formulation :

$$\mathcal{F}(\varrho_0) \geq \mathcal{F}(\varrho_T^\tau) + \frac{1}{q} \int_0^T \int_\Omega \left| \frac{\nabla L_f(\hat{\varrho}_t^\tau)}{\hat{\varrho}_t^\tau} \right|^q \, d\hat{\varrho}_t^\tau \, dt + \frac{1}{p} \int_0^T \int_\Omega |\tilde{v}_t^\tau|^p \, d\tilde{\varrho}_t^\tau \, dt. \quad (3.5)$$

3.2 Compactness estimates and passing to the limit

From our initial JKO scheme, comparing the value of the optimal minimizer ϱ_{k+1}^τ with that of the current step ϱ_k^τ we have

$$\mathcal{F}(\varrho_{k+1}^\tau) + \frac{W_p^p(\varrho_{k+1}^\tau, \varrho_k^\tau)}{p\tau^{p-1}} \leq \mathcal{F}(\varrho_k^\tau),$$

and summing these inequalities yields

$$\sum_k \frac{W_p^p(\varrho_{k+1}^\tau, \varrho_k^\tau)}{p\tau^{p-1}} \leq \mathcal{F}(\varrho_0) - \inf \mathcal{F} < \infty.$$

This is useful in deriving the following uniform continuity estimate on the geodesic interpolations : if $s < t$ we have

$$\begin{aligned}
W_p(\tilde{\varrho}_s^\tau, \tilde{\varrho}_t^\tau) &\leq \int_s^t |(\tilde{\varrho}^\tau)'|(r) \, dr \\
&\leq (t-s)^{\frac{1}{q}} \left(\int_s^t |(\tilde{\varrho}^\tau)'|(r)^p \, dr \right)^{\frac{1}{p}} \\
&= (t-s)^{\frac{1}{q}} \left(\int_s^t \|\tilde{v}_t^\tau\|_{L^p(\tilde{\varrho}_t^\tau)}^p \, dr \right)^{\frac{1}{p}} \\
&\leq (t-s)^{\frac{1}{q}} \left(\sum_k \tau \frac{W_p^p(\varrho_{k+1}^\tau, \varrho_k^\tau)}{\tau^p} \right)^{\frac{1}{p}} \\
&\leq C(t-s)^{\frac{1}{q}}.
\end{aligned}$$

For the piecewise constant interpolation, we can first say that for $t \in (k\tau, (k+1)\tau]$,

$$W_p(\tilde{\varrho}_t^\tau, \tilde{\varrho}_t^\tau) = W_p(\varrho_{k+1}^\tau, \tilde{\varrho}_t^\tau) \leq C\tau^{\frac{1}{q}}$$

and therefore the convergence of $\tilde{\varrho}_t^\tau$ will imply the convergence of $\tilde{\varrho}_t^\tau$. Finally, for the De Giorgi interpolation, for $t \in (k\tau, (k+1)\tau]$,

$$\begin{aligned}
W_p(\hat{\varrho}_t^\tau, \tilde{\varrho}_t^\tau) &\leq W_p(\hat{\varrho}_t^\tau, \tilde{\varrho}_t^\tau) + W_p(\tilde{\varrho}_t^\tau, \tilde{\varrho}_t^\tau) \\
&\leq W_p(\hat{\varrho}_t^\tau, \varrho_k^\tau) + W_p(\varrho_k^\tau, \varrho_{k+1}^\tau) + W_p(\tilde{\varrho}_t^\tau, \tilde{\varrho}_t^\tau)
\end{aligned}$$

and using

$$\begin{aligned}
W_p^p(\hat{\varrho}_t^\tau, \varrho_k^\tau) &\leq (\mathcal{F}(\varrho_k) - \mathcal{F}(\hat{\varrho}_t^\tau)) p(t-k\tau)^{p-1} \\
&\leq (\mathcal{F}(\varrho_0) - \inf \mathcal{F}) p\tau^{p-1} \\
&\leq C\tau^{\frac{p}{q}},
\end{aligned}$$

we get

$$W_p(\hat{\varrho}_t^\tau, \tilde{\varrho}_t^\tau) \leq C\tau^{\frac{1}{q}}.$$

Using the Arzelà-Ascoli theorem we can deduce that up to a subsequence which we do not relabel, when $\tau \rightarrow 0$, all interpolated curves uniformly converge to some $\varrho \in C^{\frac{1}{q}}([0, T], \mathbb{W}_p(\Omega))$

We now wish to prove compactness for the momentum variable, to pass to the limit in our EDI formulation. To this aim, we can notice that using the Jensen inequality,

$$|\tilde{E}_t^\tau|(\Omega) = \int_\Omega |\tilde{v}_t^\tau| \, d\tilde{\varrho}_t^\tau \leq \left(\int_\Omega |\tilde{v}_t^\tau|^p \, d\tilde{\varrho}_t^\tau \right)^{\frac{1}{p}} = \|\tilde{v}_t^\tau\|_{L^p(\tilde{\varrho}_t^\tau)}$$

and, adding the time variable,

$$|\tilde{E}^\tau|([0, T] \times \Omega) \leq \int_0^T \|\tilde{v}_t^\tau\|_{L^p(\tilde{\varrho}_t^\tau)} dt \leq T^{\frac{1}{q}} \left(\int_0^T \|\tilde{v}_t^\tau\|_{L^p(\tilde{\varrho}_t^\tau)}^p dt \right)^{\frac{1}{p}} \leq CT^{\frac{1}{q}}$$

We can therefore assume that as $\tau \rightarrow 0$ and for a subsequence which we do not relabel, \tilde{E}^τ weakly converges as a measure to some measure E .

Using the Benamou-Brenier functional which is lower semi-continuous for the weak convergence of measures, we therefore have that there exists $v \in L^p([0, T] \times \Omega, d\varrho_t dt)$ such that $E = \varrho v$ and

$$\liminf_\tau \mathcal{B}_p(\tilde{\varrho}^\tau, \tilde{E}^\tau) \geq \mathcal{B}_p(\varrho, E) = \frac{1}{p} \int_0^T \int_\Omega |v_t|^p d\varrho_t dt$$

Moreover the continuity equation which is linear in (ϱ, E) readily passes to the limit and we have

$$\partial_t \varrho + \nabla \cdot (\varrho v) = 0$$

Up to changing v we can also assume it is the (unique) velocity vector field of minimal L^p norm. For the slope term, we only have to use Lemma 3.4 and Fatou's Lemma to get

$$\liminf_\tau \frac{1}{q} \int_0^T \int_\Omega \left| \frac{\nabla L_f(\hat{\varrho}_t^\tau)}{\hat{\varrho}_t^\tau} \right|^q d\hat{\varrho}_t^\tau dt \geq \int_0^T \liminf_\tau \mathcal{B}_p(\hat{\varrho}_t^\tau, \nabla L_f(\hat{\varrho}_t^\tau)) dt \geq \frac{1}{q} \int_0^T \int_\Omega \left| \frac{\nabla L_f(\varrho_t)}{\varrho_t} \right|^q d\varrho_t dt.$$

Therefore, passing to the limit in 3.4 gives

$$\mathcal{F}(\varrho_0) \geq \mathcal{F}(\varrho_T) + \frac{1}{q} \int_0^T \int_\Omega \left| \frac{\nabla L_f(\varrho_t)}{\varrho_t} \right|^q d\varrho_t dt + \frac{1}{p} \int_0^T \int_\Omega |v_t|^p d\varrho_t dt. \quad (3.6)$$

3.3 Estimates on the limit curve

Lemma 3.6. *There exists $\Phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, smooth, strictly convex and superlinear such that*

$$\sup_{t \in [0, T]} \int_\Omega \Phi(\varrho_t) dx < \infty$$

Proof. We know by Dunford-Pettis' theorem that since $\varrho_0 \in L^1(\Omega)$, there exists some superlinear convex function $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, which we can assume to be 0 in a neighborhood of 0 such that

$$\int_\Omega \varphi(\varrho_0) dx < \infty$$

We take Φ to be the one constructed in Lemma 2.6 from φ . It is smooth and satisfies Mc-Cann's condition **(G1)** so that we can use it in Lemma 2.8 : at the k -th step of the JKO scheme, we have, for the approximated problem,

$$\int_\Omega \Phi(\varrho_k^\tau) dx \geq \int_\Omega \Phi(\varrho_\varepsilon) dx,$$

and taking $\varepsilon \rightarrow 0$, we have $\varrho_\varepsilon \rightarrow \varrho_{k+1}^\tau$. By lower semi continuity of the functional, we get

$$\int_{\Omega} \Phi(\varrho_k^\tau) dx \geq \int_{\Omega} \Phi(\varrho_{k+1}^\tau) dx,$$

so that by induction we have, for all k ,

$$\int_{\Omega} \Phi(\varrho_k^\tau) dx \leq \int_{\Omega} \Phi(\varrho_0) dx \leq C \left(\int_{\Omega} \varphi(\varrho_0) + 1 dx \right) < \infty.$$

Passing to the limit $\tau \rightarrow 0$, and again using the lower semi continuity of the functional, we can deduce that for all $t \geq 0$

$$\int_{\Omega} \Phi(\varrho_t) dx \leq \int_{\Omega} \Phi(\varrho_0) dx,$$

and

$$\sup_{t \in [0, T]} \int_{\Omega} \Phi(\varrho_t) dx \leq \int_{\Omega} \Phi(\varrho_0) dx < \infty$$

□

Lemma 3.7. *If we have, for $\alpha = 2 - q(1 + \frac{1}{d}) + \theta(q - 1)$:*

$$\begin{cases} \int_{\Omega} \varrho_0 \log \varrho_0 dx < +\infty & \text{if } \alpha = 1, \\ \int_{\Omega} \varrho_0^\alpha dx < +\infty & \text{if } \alpha > 1, \end{cases} \quad (3.7)$$

then

$$\int_{\Omega_T} |\nabla h(\varrho_t)|^q dx dt < \infty,$$

where h is any function satisfying $h'(z) = z^{\frac{\alpha-1}{q}} f''(z)^{\frac{1}{p}}$

Proof. We shall use the flow interchange technique, which allows us to gain some first order regularity from the integrability assumption of the initial condition. First we will look at the case where $\alpha \in [1 - \frac{1}{d}, \infty) \setminus \{1\}$. Starting from the k -th step of our JKO scheme, we look at the ε -approximated problem and apply Lemma 2.8 with the function $z \mapsto \frac{1}{\alpha-1} z^\alpha$ and get

$$\frac{1}{\alpha-1} \int_{\Omega} (\varrho_k^\tau)^\alpha - \frac{1}{\alpha-1} \int_{\Omega} \varrho_\varepsilon^\alpha \geq \tau \int_{\Omega} \varrho_\varepsilon^{\alpha-1} f''(\varrho_\varepsilon)^{q-1} |\nabla \varrho_\varepsilon|^q dx \geq 0,$$

where ϱ_ε is the solution of 2.2 with $g = \varrho_k$. We can write $\varrho_\varepsilon^{\alpha-1} f''(\varrho_\varepsilon)^{q-1} |\nabla \varrho_\varepsilon|^q = |\nabla h(\varrho_\varepsilon)|^q$, and h being strictly increasing, when taking the limit $\varepsilon \rightarrow 0$ (and hence $\varrho_\varepsilon \rightarrow \varrho_{k+1}^\tau$), we can argue as in the proof of Lemma 3.4 to find

$$\frac{1}{\alpha-1} \int_{\Omega} (\varrho_k^\tau)^\alpha - \frac{1}{\alpha-1} \int_{\Omega} (\varrho_{k+1}^\tau)^\alpha \geq \tau \int_{\Omega} |\nabla h(\varrho_{k+1}^\tau)|^q dx.$$

Summing on each step of the JKO scheme, and using Fatou's Lemma when passing to the limit $\tau \rightarrow 0$, we can obtain

$$\frac{1}{\alpha - 1} \int_{\Omega} \varrho_0^\alpha \geq \int_{\Omega_T} |\nabla h(\varrho)|^q dx dt + \frac{1}{\alpha - 1} \int_{\Omega} \varrho_T^\alpha dx.$$

For the case $\alpha = 1$, we use the same techniques as above, but applying lemma 2.8 to the function $z \mapsto z \log z$ which satisfies Mc-Cann's condition and, in terms of second derivatives, behaves in a similar manner to power functions, so that all computations are the same. Finally, if $\alpha < 1 - \frac{1}{d}$, one can use the result for $\alpha = 1 - \frac{1}{d}$; indeed α is an increasing function of θ , and the condition $f''(r) \geq Cr^{-\theta}$ when r is large becomes weaker as θ increases. \square

4 Derivative of \mathcal{F} along the flow

In this section, we will endeavor to compute the derivative of \mathcal{F} along the curve given by (ϱ, v) . Formally, using the continuity equation, we can write

$$\frac{d}{dt} \mathcal{F}(\varrho_t) = \int_{\Omega} \frac{d}{dt} f(\varrho_t) dx = \int_{\Omega} f'(\varrho_t) \partial_t \varrho_t dt = \int_{\Omega} \nabla f'(\varrho_t) \cdot v_t d\varrho_t,$$

which gives

$$\mathcal{F}(\varrho_T) - \mathcal{F}(\varrho_0) = \int_0^T \int_{\Omega} \nabla f'(\varrho_t) \cdot v_t d\varrho_t dt.$$

Now, using $\nabla L_f(\varrho_t) = \nabla f(\varrho_t) \varrho_t$ and the EDI formulation 3.6,

$$0 \geq \int_0^T \int_{\Omega} \nabla f'(\varrho_t) \cdot v_t d\varrho_t dt + \frac{1}{q} \int_0^T \int_{\Omega} |\nabla f'(\varrho_t)|^q d\varrho_t dt + \frac{1}{p} \int_0^T \int_{\Omega} |v_t|^p d\varrho_t dt,$$

which, with Young's inequality gives

$$\begin{cases} \partial_t \varrho_t + \nabla \cdot (\varrho_t v_t) = 0 \\ v_t = -(\nabla f'(\varrho_t))^{q-1} \end{cases}$$

In the rest of this section we shall give a rigorous proof of the above result. First we can notice that for our purposes, it is enough to prove that

$$\mathcal{F}(\varrho_T) - \mathcal{F}(\varrho_0) \geq \int_0^T \int_{\Omega} \frac{\nabla L_f(\varrho_t)}{\varrho_t} \cdot v_t d\varrho_t dt.$$

Remark 4.1. The above statement could be proved in the case where $\delta \leq \varrho \leq \frac{1}{\delta}$ for some $\delta > 0$ (which is the case for example, when $\delta \leq \varrho_0 \leq \frac{1}{\delta}$ (see [1] or [16])) by regularizing the couple $(\varrho, \varrho v)$ via convolution. The L^∞ bound corresponds to the worst case $f'' > 0$ in our theorem, and we have no need for lower bounds in the following proofs.

Theorem 4.2 (Chain Rule). *Let $(\varrho_t)_{t \in [0, T]}$ be an absolutely continuous curve in $\mathbb{W}_p(\Omega)$ with velocity vector field $(v_t)_{t \in [0, T]}$, so that it solves the continuity equation*

$$\partial_t \varrho_t + \nabla \cdot (\varrho_t v_t) = 0.$$

If there exists a non-negative smooth superlinear strictly convex function Φ such that

$$\int_{\Omega_T} \Phi(\varrho_t) \, dx \, dt < \infty,$$

and we have

$$\int_{\Omega_T} \left| \frac{\nabla L_f(\varrho_t)}{\varrho_t} \right|^q \, d\varrho_t \, dt < \infty,$$

*and, if f does not satisfy Mc-Cann's condition **(G1)**,*

$$\int_{\Omega_T} |\nabla h(\varrho)|^q \, dx \, dt < +\infty,$$

where h is any function satisfying $h'(z) = z^{\frac{\alpha-1}{q}} f''(z)^{\frac{1}{p}}$, and $\alpha = 2 - q(1 + \frac{1}{d}) + \theta(q - 1)$, then

$$\mathcal{F}(\varrho_T) - \mathcal{F}(\varrho_0) = \int_{\Omega_T} \frac{\nabla L_f(\varrho)}{\varrho} v \, d\varrho \, dt.$$

Proof. Step 1 : Truncation and decomposition

In this first step, we write f as the difference of two functions f_1, f_2 which will satisfy Mc-Cann's condition, and for which computations are allowed. To be able to do this we first need to truncate f linearly near 0 and $+\infty$, and we will pass to the limit at the end to recover the initial function.

First we linearly truncate f near 0 and infinity and for ease of computations subtract a positive constant such that the truncated function is zero at zero.

$$\tilde{f}(z) = \begin{cases} a_0 z & \text{for } z \leq z_0, \\ f(z) + b_0 & \text{for } z_0 \leq z \leq z_1, \\ a_1 z + b_1 & \text{for } z_1 \leq z, \end{cases}$$

with $a_0 = f'(z_0) \leq a_1 = f'(z_1)$, $b_0 = L_f(z_0)$, $b_1 = L_f(z_0) - L_f(z_1) \leq 0$. Its associated Mc-Cann function is

$$\mathcal{M}\tilde{f}(s) = \begin{cases} a_1 + s^d b_1 & \text{for } s \leq z_1^{-\frac{1}{d}}, \\ \mathcal{M}f(s) + b_0 s^d & \text{for } z_1^{-\frac{1}{d}} \leq s \leq z_0^{-\frac{1}{d}}, \\ a_0 & \text{for } z_0^{-\frac{1}{d}} \leq s. \end{cases}$$

We take the second derivative which is

$$(\mathcal{M}\tilde{f})''(s) = \begin{cases} b_1(d-1)ds^{d-2} & \text{for } s \leq z_1^{-\frac{1}{d}}, \\ (\mathcal{M}f)''(s) + b_0(d-1)ds^{d-2} & \text{for } z_1^{-\frac{1}{d}} \leq s \leq z_0^{-\frac{1}{d}}, \\ 0 & \text{for } z_0^{-\frac{1}{d}} \leq s. \end{cases}$$

We now take $\tilde{g}_2''(s) = \left((\mathcal{M}\tilde{f})''(s) \right)_-$ and integrate twice so that, choosing the constants c_0 and c_1 we have

$$\tilde{g}_2(s) = \begin{cases} -b_1s^d + c_1s & \text{for } s \leq z_1^{-\frac{1}{d}}, \\ \Theta(s) & \text{for } z_1^{-\frac{1}{d}} \leq s \leq z_0^{-\frac{1}{d}}, \\ c_0 & \text{for } z_0^{-\frac{1}{d}} \leq s, \end{cases}$$

where Θ is the double primitive of $\tilde{g}_2(s)$ with suitable integration constants so that \tilde{g}_2 is C^2 (except at $z_1^{-\frac{1}{d}}$ and $z_0^{-\frac{1}{d}}$). \tilde{g}_2 is convex by construction, and also decreasing since it is constant at infinity. And finally we can define

$$\tilde{f}_2(z) = \mathcal{M}^{-1}\tilde{g}_2(z) = \begin{cases} c_0z & \text{for } z \leq z_0, \\ \mathcal{M}^{-1}\Theta(z) & \text{for } z_0 \leq z \leq z_1, \\ c_1z^{1-\frac{1}{d}} - b_1 & \text{for } z_1 \leq z, \end{cases}$$

\tilde{f}_2 is a C^2 (except at z_0 and z_1) convex function that is linear near the origin and that satisfies the Mc-Cann condition and

$$\tilde{f}_2''(z) \leq Cz^{-(1+\frac{1}{d})} \text{ when } z \geq z_0.$$

Now we set

$$\tilde{f}_1 = \tilde{f} + \tilde{f}_2,$$

which, by construction also satisfies Mc-Cann's condition ($\mathcal{M}\tilde{f}_1$ is convex and constant for large enough values).

Step 2 : Getting regularity for the slope of f_2

In this second step, we use the integrability assumption on $\nabla h(\varrho)$ to gain the necessary integrability on $\nabla L_{\tilde{f}_2}(\varrho)$.

To obtain our needed regularity we will use a finite differences trick, indeed we can write, for $z \geq z_0$,

$$\left(L_{\tilde{f}_2} \right)'(z) = z\tilde{f}_2''(z) \leq Cz^{-\frac{1}{d}} \leq Cz^{\frac{q-1}{q}} h'(z) z^{-\frac{1}{d} + \frac{q-1}{q}\theta - \frac{\alpha-1}{q} - \frac{q-1}{q}} = Cz^{\frac{q-1}{q}} h'(z),$$

where we used the hypothesis $f''(r) \geq Cr^{-\theta}$ and our condition $\alpha = 2 - q(1 + \frac{1}{d}) + \theta(q-1)$ so that $-\frac{1}{d} + \frac{q-1}{q}\theta - \frac{\alpha-1}{q} - \frac{q-1}{q} = 0$. Since $\tilde{f}_2''(z) = 0$ when $z < z_0$, the above inequality is also trivially true in this case. We can therefore write, for $s > t$,

$$L_{\tilde{f}_2}(s) - L_{\tilde{f}_2}(t) = \int_t^s \left(L_{\tilde{f}_2} \right)'(z) dz \leq C \int_t^s z^{\frac{q-1}{q}} h'(z) dz$$

and

$$\left| L_{\tilde{f}_2}(s) - L_{\tilde{f}_2}(t) \right| \leq C \max(s, t)^{\frac{q-1}{q}} |h(s) - h(t)|$$

Taking a test function $\varphi \in C_c^\infty(\Omega)$ and $\delta \in \mathbb{R}^d$, we have

$$\begin{aligned} \left| \int_{\Omega} L_{\tilde{f}_2}(\varrho(x)) \nabla \varphi(x) \cdot \delta \, dx \right| &= \left| \lim_{t \rightarrow 0} \int_{\Omega} L_{\tilde{f}_2}(\varrho(x)) \frac{\varphi(x - t\delta) - \varphi(x)}{t} \, dx \right| \\ &= \left| \lim_{t \rightarrow 0} \int_{\Omega} \frac{L_{\tilde{f}_2}(\varrho(x + t\delta)) - L_{\tilde{f}_2}(\varrho(x))}{t} \varphi(x) \, dx \right| \\ &\leq \limsup_{t \rightarrow 0} \int_{\Omega} \max(\varrho(x + t\delta), \varrho(x))^{\frac{q-1}{q}} \frac{|h(\varrho(x + t\delta)) - h(\varrho(x))|}{t} |\varphi(x)| \, dx \\ &= \int_{\Omega} \varrho(x)^{\frac{q-1}{q}} |\nabla h(\varrho)(x) \cdot \delta| |\varphi(x)| \, dx, \end{aligned}$$

where we used the strong convergence $\varrho(\cdot + t\delta) \rightarrow \varrho$ in L^1 and $\frac{|h(\varrho(\cdot + t\delta)) - h(\varrho)|}{t} \rightarrow |\nabla h(\varrho) \cdot \delta|$ in L^q . This proves that $L_{\tilde{f}_2}(\varrho)$ is a $W^{1,1}$ function such that $|\nabla L_{\tilde{f}_2}(\varrho)| \leq \varrho^{\frac{q-1}{q}} |\nabla h(\varrho)|$, so that

$$\int_{\Omega_T} \frac{|\nabla L_{\tilde{f}_2}(\varrho)|^q}{\varrho^{q-1}} \, dx \, dt \leq \int_{\Omega_T} |\nabla h(\varrho)|^q \, dx \, dt < \infty$$

Step 3 : Proving the chain rule

In this step, we will prove the chain rule for the truncated function \tilde{f} . The idea is to discretize the curve $(\varrho_t)_t$ and approximate it with its geodesic interpolation, for which we can use the geodesic convexity of \tilde{f}_1, \tilde{f}_2 and the previous regularity results to obtain the discrete chain rule. We then pass to the limit and recover the chain rule for the original curve.

Lemma 4.3. *Let $(\varrho_t)_{t \in [0, T]}$ be an absolutely continuous curve in $\mathbb{W}_p(\Omega)$ with velocity vector field $(v_t)_{t \in [0, T]}$, so that it solves the continuity equation*

$$\partial_t \varrho_t + \nabla \cdot (\varrho_t v_t) = 0,$$

and there exists a non-negative smooth superlinear strictly convex function Φ such that

$$\int_{\Omega_T} \Phi(\varrho_t) \, dx \, dt < \infty.$$

Let $\ell : \mathbb{R}_+ \rightarrow \mathbb{R}$ be a C^1 function satisfying **(H1-G1)**, such that $L_\ell(\varrho) \in L^1_{t,x}$, and

$$\int_{\Omega_T} \frac{|\nabla L_h(\varrho)|^q}{\varrho^{q-1}} \, dx \, dt < +\infty,$$

then

$$\mathcal{L}(\varrho_T) - \mathcal{L}(\varrho_0) = \int_{\Omega_T} \frac{\nabla L_\ell(\varrho)}{\varrho} \cdot v \, d\varrho \, dt.$$

Proof. We use Theorem A.1 to find a sequence of partitions $0 = t_0^N < t_1^N < \dots < t_{N+1}^N = T$ of $[0, T]$ with $\sup_i t_{i+1}^N - t_i^N \xrightarrow{N \rightarrow \infty} 0$, such that

$$\int_{\Omega} L_{\ell}(\varrho_{t_i^N}) dx < \infty,$$

and

$$\sum_{k=1}^N (t_{k+1}^N - t_k^N) A(t_k^N) \rightarrow \int_0^T A(t) dt,$$

where $A(t) = \int_{\Omega} \left| \frac{\nabla L_{\ell}(\varrho_t)}{\varrho_t} \right|^q d\varrho_t + \int_{\Omega} \Phi(\varrho_t) dx$. Like before, we denote by $\bar{\varrho}^N$ the piecewise constant interpolation of $(\varrho_{t_i^N})_i$, defined by $\bar{\varrho}_t^N = \varrho_{t_i^N}$ if $t \in [t_i^N, t_{i+1}^N[$ for $i = 1, \dots, N$, and $\bar{\varrho}_t^N = 0$ for $t \in [0, t_1^N[$ so that we have

$$\int_{\Omega_T} \left| \frac{\nabla L_{\ell}(\bar{\varrho}_t^N)}{\bar{\varrho}_t^N} \right|^q d\bar{\varrho}_t^N dt + \int_{\Omega_T} \Phi(\bar{\varrho}_t^N) dx dt \xrightarrow{N \rightarrow \infty} \int_{\Omega_T} \left| \frac{\nabla L_{\ell}(\varrho_t)}{\varrho_t} \right|^q d\varrho_t dt + \int_{\Omega_T} \Phi(\varrho_t) dx dt < \infty. \quad (4.1)$$

Using the same argument as in the proof of Lemma 2.8, the geodesic convexity of \mathcal{H} , and the fact that $L_h(\varrho)$ is $W^{1,1}$ allows us to write

$$\begin{aligned} \mathcal{L}(\varrho_{t_{i+1}^N}) - \mathcal{L}(\varrho_{t_i^N}) &\geq \int_{\Omega} \frac{\nabla L_{\ell}(\varrho_{t_i^N})}{\varrho_{t_i^N}} (T_i - \text{id}) d\varrho_{t_i^N} \\ &= \int_{t_i^N}^{t_{i+1}^N} \int_{\Omega} \frac{\nabla L_{\ell}(\bar{\varrho}_t^N)}{\bar{\varrho}_t^N} \bar{v}_t^N d\bar{\varrho}_t^N dt, \end{aligned}$$

where T_i is the optimal transport map from $\varrho_{t_i^N}$ to $\varrho_{t_{i+1}^N}$ and $\bar{v}_t^N = \frac{T_i - \text{id}}{t_{i+1}^N - t_i^N}$. Summing everything gives

$$\mathcal{L}(\varrho_T) - \mathcal{L}(\varrho_{t_1}) \geq \int_{\Omega_T} \frac{\nabla L_{\ell}(\bar{\varrho}_t^N)}{\bar{\varrho}_t^N} \bar{v}_t^N d\bar{\varrho}_t^N dt. \quad (4.2)$$

We will now prove that this inequality passes to the limit. Since $\varrho \in C^0([0, T], \mathbb{W}_p)$, we have that $\bar{\varrho}_t^N \rightarrow \varrho_t$ weakly in L^1 for any fixed $t \in]0, T]$. From Lemmas 2.4 and 3.4, and using Fatou's Lemma, we therefore obtain that

$$\begin{aligned} \liminf \int_{\Omega_T} \left| \frac{\nabla L_{\ell}(\bar{\varrho}_t^N)}{\bar{\varrho}_t^N} \right|^q d\bar{\varrho}_t^N dt &\geq \int_{\Omega_T} \left| \frac{\nabla L_{\ell}(\varrho)}{\varrho} \right|^q d\varrho dt \\ \liminf \int_{\Omega_T} \Phi(\bar{\varrho}_t^N) dx dt &\geq \int_{\Omega_T} \Phi(\varrho_t) dx dt. \end{aligned}$$

Together with (4.1), this implies that

$$\begin{aligned} \int_{\Omega_T} \left| \frac{\nabla L_\ell(\bar{\varrho}_t^N)}{\bar{\varrho}_t^N} \right|^q d\bar{\varrho}_t^N dt &\rightarrow \int_{\Omega_T} \left| \frac{\nabla L_\ell(\varrho_t)}{\varrho_t} \right|^q d\varrho_t dt \\ \int_{\Omega_T} \Phi(\bar{\varrho}_t^N) dx dt &\rightarrow \int_{\Omega_T} \Phi(\varrho_t) dx dt. \end{aligned}$$

We will now use a classical trick of convex functions to obtain the almost everywhere convergence of $\bar{\varrho}_t$ (see [12]). We define ω to be the convexity gap of Φ , which is to say

$$\omega(z, x) = \Phi(z) - \Phi(x) - \Phi'(x)(z - x) \geq 0.$$

By the strict convexity of Φ , $\omega(z, x) = 0$ if and only if $z = x$, and if $(z_n)_n$ is a sequence such that $\omega(z_n, x) \rightarrow 0$, then $z_n \rightarrow x$. Let $A_M = \{\varrho \leq M\}$, we have:

$$\int_{\Omega_T} \Phi(\bar{\varrho}_t^N) dx dt \geq \int_{A_M} \Phi(\varrho_t) dx dt + \int_{A_M} \Phi'(\varrho_t)(\bar{\varrho}_t^N - \varrho_t) dx dt + \int_{A_M} \omega(\bar{\varrho}_t^N, \varrho_t) dx dt.$$

We know that $\bar{\varrho}^N$ weakly converges in $L_{t,x}^1$ to ϱ , and on A_M , $\Phi'(\varrho)$ is bounded, so that the second integral vanishes in the limit. Taking the lim sup we therefore get

$$\int_{A_M^c} \Phi(\varrho_t) dx dt \geq \limsup \int_{A_M} \omega(\bar{\varrho}_t^N, \varrho_t) dx dt.$$

Notice that the left hand side is decreasing in M , while the right hand side is increasing in M , therefore, taking the limit $M \rightarrow +\infty$, we have

$$0 = \lim \int_{\Omega_T} \omega(\bar{\varrho}_t^N, \varrho_t) dx dt$$

Up to extraction, we can therefore assume that $\omega(\bar{\varrho}, \varrho) \rightarrow 0$ almost everywhere, which, as stated above entails that $\bar{\varrho} \rightarrow \varrho$ almost everywhere. Again 4.1 implies that $\int_{\Omega_T} |\nabla L_\ell(\bar{\varrho}_t^N)| dx dt$ is bounded, so that up to extraction we can assume $(\nabla L_\ell(\bar{\varrho}^N))$ converges to some measure. Using the strong convergence of $\bar{\varrho}$, one can identify this limit as $\nabla L_\ell(\varrho)$. We also know that $\int_{\Omega_T} \left| \frac{\nabla L_\ell(\bar{\varrho}_t^N)}{\bar{\varrho}_t^N} \right|^q d\bar{\varrho}_t^N dt$ is bounded, so that up to extraction, $(\nabla L_\ell(\bar{\varrho}^N)(\bar{\varrho}^N)^{-\frac{1}{p}})$ weakly converges in $L_{t,x}^q$ to some ξ . Multiplying by $(\bar{\varrho}^N)^{\frac{1}{p}}$, which strongly converges to $\varrho^{\frac{1}{p}}$ in $L_{t,x}^p$, we obtain $\nabla L_\ell(\bar{\varrho}^N) \rightharpoonup \xi \varrho^{\frac{1}{p}}$, so that by uniqueness of the limit, $\xi = \nabla L_\ell(\varrho) \varrho^{-\frac{1}{p}}$. Since we also have the convergence of the L^q norm, we actually have strong convergence of the slope term.

We now prove the convergence of the discrete velocity field \bar{v}^N . Just like in section 3.2, by looking at the geodesic interpolation we can prove the following inequality :

$$\int_{\Omega_T} |\bar{v}_t^N|^p d\bar{\varrho}_t^N dt \leq \int_{\Omega_T} |v_t|^p d\varrho dt,$$

which implies that the sequence $(\bar{v}^N(\bar{\varrho}^N)^{\frac{1}{p}})$ is bounded in $L^p_{t,x}$, so that up to extraction it weakly converges to some $\xi \in L^p_{t,x}$. Again like before we also have

$$\int_{\Omega_T} |\bar{v}_t^N \bar{\varrho}_t^N| dx dt \leq T^{\frac{1}{q}} \left(\int_0^T \|\bar{v}_t^N\|_{L^p(\bar{\varrho}_t^N)}^p dt \right)^{\frac{1}{p}} \leq C,$$

so that the sequence $(\bar{v}^N \bar{\varrho}^N)$ weakly converges as a measure to some $E \in \mathcal{M}(\Omega)^d$. Using the convergence $\bar{\varrho}^N \rightarrow \varrho$ and Lemma 2.3, we deduce that there exists some $\hat{v} \in L^p(\varrho)$ such that $E = \hat{v}\varrho$ and

$$\int_{\Omega_T} |v|^p d\varrho dt \geq \liminf_{N \rightarrow \infty} \int_{\Omega_T} |\bar{v}_t^N|^p d\bar{\varrho} dt \geq \int_{\Omega_T} |\hat{v}|^p d\varrho dt.$$

Using arguments as in [18] section 9.3, we can find that (ϱ, E) satisfy the continuity equation, so that since we chose v to be the velocity field with minimum norm, we have $\hat{v} = v$ in $L^p(\varrho_t)$ for almost every t and

$$\int_{\Omega_T} |v|^p d\varrho dt = \int_{\Omega_T} |\hat{v}|^p d\varrho dt = \lim_{N \rightarrow \infty} \int_{\Omega_T} |\bar{v}_t^N|^p d\bar{\varrho}_t^N dt.$$

We will now use the same convex trick to prove that $\xi = v\varrho^{\frac{1}{p}}$. We write $\omega(z, x) = \frac{|z|^p}{p} - \frac{|x|^p}{p} - x^{p-1}(z-x) \geq 0$, which enjoys the same properties as before, and $A_M = \{|v| \leq M\}$, and we have

$$\int_{\Omega_T} \bar{\varrho}_t^N \frac{|\bar{v}_t^N|^p}{p} dx dt \geq \int_{A_M} \bar{\varrho}_t^N \frac{|v|^p}{p} dx dt + \int_{A_M} \bar{\varrho}_t^N |v|^{p-1} (\bar{v}_t^N - v) dx dt + \int_{A_M} \bar{\varrho}_t^N \omega(\bar{v}_t^N, v) dx dt. \quad (4.3)$$

We know that $(\bar{\varrho}_t^N \bar{v}_t^N)$ is uniformly integrable, indeed, if $B \subset \Omega \times [0, T]$ is measurable, then

$$\int_B \bar{\varrho}_t^N \bar{v}_t^N dx dt \leq \left(\int_B \bar{\varrho}_t^N |\bar{v}_t^N|^p dx dt \right)^{\frac{1}{p}} \left(\int_B \bar{\varrho}_t^N dx dt \right)^{\frac{1}{q}} \leq C \left(\int_B \bar{\varrho}_t^N dx dt \right)^{\frac{1}{q}},$$

so that the uniform integrability of $\bar{\varrho}^N$ implies the uniform integrability of $(\bar{\varrho}^N \bar{v}^N)$. Therefore, we conclude that $\bar{\varrho}^N \bar{v}^N \rightharpoonup \varrho v$ weakly in $L^1_{t,x}$ and

$$\int_{A_M} \bar{\varrho}_t^N \bar{v}_t^N |v|^{p-1} dx dt \rightarrow \int_{A_M} \varrho |v|^p dx dt$$

Taking the limit in 4.3, we therefore get

$$\int_{A_M^c} \varrho \frac{|v|^p}{p} dx dt \geq \limsup_{N \rightarrow \infty} \int_{A_M} \bar{\varrho}_t^N \omega(\bar{v}_t^N, v) dx dt.$$

Again, the left hand side is decreasing in M while the right hand side is increasing in M , so that we have

$$0 = \limsup_{N \rightarrow \infty} \int_{\Omega_T} \bar{\varrho}_t^N \omega(\bar{v}_t^N, v) dx dt,$$

and, up to extracting again, we find that $\bar{v} \rightarrow v$ almost everywhere on $\{\varrho > 0\}$. From this convergence and $\bar{v}^N(\bar{\varrho}^N)^{\frac{1}{p}} \rightarrow \xi$, we deduce that $\xi = v\varrho^{\frac{1}{p}}$ on $\{\varrho > 0\}$. Putting everything together, we can write

$$\begin{aligned} \int_{\Omega_T} |\xi|^p dx dt &\leq \liminf_{N \rightarrow \infty} \int_{\Omega_T} |\bar{v}_t^N|^p \bar{\varrho}_t^N dx dt = \int_{\Omega_T} |v|^p \varrho dx dt \\ &= \int_{\{\varrho > 0\}} |v|^p \varrho dx dt \\ &= \int_{\{\varrho > 0\}} |\xi|^p dx dt \leq \int_{\Omega_T} |\xi|^p dx dt, \end{aligned}$$

so that $\xi = v\varrho^{\frac{1}{p}}$ almost everywhere. In conclusion, we have proved that $\bar{v}^N(\bar{\varrho}^N)^{\frac{1}{p}} \rightarrow v\varrho^{\frac{1}{p}}$ in $L^p_{t,x}$, because we also have convergence of the norm. We can now finally pass to the limit in 4.2, and we obtain

$$\mathcal{L}(\varrho_T) - \mathcal{L}(\varrho_0) \geq \int_{\Omega_T} \frac{\nabla L_\ell(\varrho)}{\varrho} \cdot v d\varrho dt.$$

Applying this result to the same curve while inverting the time variable, i.e. to the curve $\rho_t = \varrho_{T-t}$ (whose velocity vector is $-v_{T-t}$) we obtain

$$\mathcal{L}(\varrho_T) - \mathcal{L}(\varrho_0) = \int_{\Omega_T} \frac{\nabla L_\ell(\varrho)}{\varrho} v d\varrho dt.$$

□

We use Lemma 4.3 with the functions $h = \tilde{f}_1$, and $h = \tilde{f}_2$ to obtain

Corollary 4.4. With our previous notations, we have the following :

$$\begin{aligned} \tilde{\mathcal{F}}_1(\varrho_T) - \tilde{\mathcal{F}}_1(\varrho_0) &= \int_{\Omega_T} \frac{\nabla L_{\tilde{f}_1}(\varrho)}{\varrho} \cdot v d\varrho dt, \\ \tilde{\mathcal{F}}_2(\varrho_T) - \tilde{\mathcal{F}}_2(\varrho_0) &= \int_{\Omega_T} \frac{\nabla L_{\tilde{f}_2}(\varrho)}{\varrho} \cdot v d\varrho dt, \end{aligned}$$

so that

$$\tilde{\mathcal{F}}(\varrho_T) - \tilde{\mathcal{F}}(\varrho_0) = \int_{\Omega_T} \frac{\nabla L_{\tilde{f}}(\varrho)}{\varrho} \cdot v d\varrho dt$$

Step 4 : Passing the truncation to the limit

We will now prove that we can pass our truncation process to the limit and still keep our inequality. As a reminder, we defined our linear truncation \tilde{f}_n as

$$\tilde{f}_n(z) = c^n + \begin{cases} a_0^n z & \text{for } z \leq z_0^n, \\ f(z) + b_0^n & \text{for } z_0^n \leq z \leq z_1^n, \\ a_1^n z + b_1^n & \text{for } z_1^n \leq z, \end{cases}$$

with $a_0^n = f'(z_0^n) \leq a_1^n = f'(z_1^n)$, $b_0^n = L_f(z_0^n)$, $b_1^n = L_f(z_0^n) - Lf(z_1^n) \leq 0$ and $c^n = f(z_0^n) - a_0^n z_0^n$. This formula just amounts to setting $\tilde{f}_n''(z) = 0$ when $z \leq z_0^n$ and $z \geq z_1^n$. Since f is convex, we have $\tilde{f}_n \leq f$, so that by the monotone convergence theorem, we directly have

$$\tilde{\mathcal{F}}_n(\varrho) \rightarrow \mathcal{F}(\varrho)$$

for any density $\varrho \in L^1(\Omega)$.

For the slope term, since $L_h'(z) = zh''(z)$, we have $L_{\tilde{f}_n}(\varrho) = (\alpha^n \wedge (\beta^n \vee L_f(\varrho)))$ for some constants α^n and β^n , so that $L_{\tilde{f}_n}(\varrho)$ is a Lipschitz function of $L_f(\varrho)$ and we can write

$$\nabla L_{\tilde{f}_n}(\varrho) = \nabla L_f(\varrho) \mathbb{1}_{\{\alpha^n \leq \varrho \leq \beta^n\}}$$

Using dominated convergence, we can therefore conclude that

$$\mathcal{F}(\varrho_T) - \mathcal{F}(\varrho_0) = \int_{\Omega_T} \frac{\nabla L_f(\varrho)}{\varrho} \cdot v \, d\varrho \, dt \quad (4.4)$$

In the case when f already satisfies Mc-Cann's condition (**G1**), one can skip steps 1,2 and 4 and directly use step 3 with the untruncated function f itself. \square

We are now ready to give the proof of Theorem 1.1.

Proof. We will prove that the curve $(\varrho_t)_{t \in [0, T]}$ we constructed in Section 3 is the solution of the PDE (1.2). We know that there exists a velocity field $(v_t)_{t \in [0, T]}$ such that we have

$$\partial_t \varrho_t + \nabla \cdot (\varrho_t v_t) = 0,$$

and the EDI condition (3.6)

$$\mathcal{F}(\varrho_0) \geq \mathcal{F}(\varrho_T) + \frac{1}{q} \int_0^T \int_{\Omega} \left| \frac{\nabla L_f(\varrho_t)}{\varrho_t} \right|^q \, d\varrho_t \, dt + \frac{1}{p} \int_0^T \int_{\Omega} |v_t|^p \, d\varrho_t \, dt.$$

Using Lemmas 3.6 and 3.7, we can apply Theorem 4.2 to obtain

$$\mathcal{F}(\varrho_T) - \mathcal{F}(\varrho_0) = \int_{\Omega_T} \frac{\nabla L_f(\varrho)}{\varrho} v \, d\varrho \, dt,$$

so that we have

$$0 \geq \int_{\Omega_T} \frac{\nabla L_f(\varrho)}{\varrho} \cdot v + \frac{1}{q} \left| \frac{\nabla L_f(\varrho)}{\varrho} \right|^q + \frac{1}{p} |v|^p \, d\varrho \, dt.$$

According to Young's inequality, this implies that

$$v = - \left(\frac{\nabla L_f(\varrho)}{\varrho} \right)^{q-1} \quad \varrho \text{ a.e. for a.e. } t.$$

\square

5 BV Estimates for the solution

In this section we give some BV estimates on the limit curve obtained by our p -JKO scheme, both our results are obtained using the flow interchange technique, with suitable assumptions on the initial condition and the function f . The following lemma, proved in [6] (see also [10]) as a generalization of the results of [9], is useful to prove that BV bounds are conserved along out p -JKO scheme.

Lemma 5.1. *Let $\Omega \subset \mathbb{R}^d$ be bounded and convex with non-empty interior, $\varrho, g \in W^{1,1}(\Omega)$ be two probability densities, $h \in C^1(\mathbb{R}^d)$ a radially symmetric strictly convex function and $H \in C^1(\mathbb{R}^d \setminus \{0\})$ be a radially symmetric convex function, then the following inequality holds*

$$\int_{\Omega} (\nabla \varrho \cdot \nabla H(\nabla \varphi) + \nabla g \cdot \nabla H(\nabla \psi)) \, dx \geq 0, \quad (5.1)$$

where (φ, ψ) is a choice of Kantorovich potentials for the optimal transport problem between ϱ and g for the transport cost given by $c(x, y) = h(x - y)$, with the convention that $\nabla H(0) = 0$.

Lemma 5.2. *Let $g \in \mathcal{P}(\Omega) \cap BV(\Omega)$, and denote by ϱ the solution of problem 2.1. Then $\varrho \in BV(\Omega)$ and $\|\varrho\|_{BV(\Omega)} \leq \|g\|_{BV(\Omega)}$*

Proof. For the case $p = 2$, the proof is done in [11] using the five gradients inequality for $p = 2$. For the sake of completeness, we give below the same proof adapted to the general case, using the generalized five gradients inequality 5.1. As usual, we start by looking at the approximated problem 2.2, and assume that $g \in W^{1,1}(\Omega)$. Applying Lemma 5.1, with the function $H(z) = |z|$, we obtain

$$\int_{\Omega} \left(\nabla \varrho_{\varepsilon} \cdot \frac{\nabla \varphi}{|\nabla \varphi|} + \nabla g \cdot \frac{\nabla \psi}{|\nabla \psi|} \right) \, dx \geq 0$$

Using the optimality condition $f''_{\varepsilon}(\varrho_{\varepsilon}) \nabla \varrho_{\varepsilon} = -\frac{\nabla \varphi}{r^{p-1}}$, we know that $\nabla \varrho_{\varepsilon}$ and $\nabla \varphi$ point in opposite directions, so that we get

$$\int_{\Omega} |\nabla \varrho_{\varepsilon}| \, dx \leq \int_{\Omega} \nabla g \cdot \frac{\nabla \psi}{|\nabla \psi|} \, dx \leq \int_{\Omega} |\nabla g| \, dx.$$

Passing to the limit $\varepsilon \rightarrow 0$, we obtain $\|\varrho\|_{BV} \leq \|g\|_{W^{1,1}}$, and we pass to $g \in BV(\Omega)$ by approximating g by $g_n \in W^{1,1}$ with converging BV norm. This yields uniform convergence of the functional $\varrho \mapsto W_p(\varrho, g_n)$ to $\varrho \mapsto W_p(\varrho, g)$ and therefore Γ -convergence of the functional in problem (2.1), so that the corresponding minimizers ϱ_n converge to the minimizer of the limit problem ϱ . Using the lower semi continuity of the BV norm, we obtain the claimed inequality $\|\varrho\|_{BV(\Omega)} \leq \|g\|_{BV(\Omega)}$ \square

A direct application of the above result immediately gives the following corollaries :

Corollary 5.3. The recursive sequence (ϱ_k^T) obtained from the p -JKO scheme (3.1) satisfies $\|\varrho_{k+1}\|_{BV}^T \leq \|\varrho_k^T\|_{BV}$.

Corollary 5.4. If $\varrho_0 \in BV(\Omega)$, then $\varrho_t \in BV(\Omega)$ for all $t \geq 0$ and $\|\varrho_t\|_{BV} \leq \|\varrho_0\|_{BV}$.

Remark 5.5. If we know that the solution to (1.2) is unique, when given some $t \geq 0$, we can use Corollary 5.4 to prove that for all $s \geq 0$, we have $\|\varrho_{t+s}\|_{BV} \leq \|\varrho_t\|_{BV}$. Indeed, we can look at the same p -JKO scheme with ϱ_t as an initial condition, and obtain a solution $(\bar{\varrho}_s)_s$ such that $\|\bar{\varrho}_s\|_{BV} \leq \|\varrho_t\|_{BV}$. Uniqueness gives $\bar{\varrho}_s = \varrho_{t+s}$ and the result.

We now prove an instantaneous BV regularization property for the limit curve constructed by our p -JKO scheme. Notice that in the following theorem, we assume that $f''(z) \geq Cz^{-\theta}$ not only for large z , but for z near 0 as well.

Theorem 5.6. *Assume $f''(z) \geq Cz^{-\theta}$ for all $z \in \mathbb{R}_+$, $\theta \geq -\frac{p-1}{d}$, and let $\beta = \max\left(1 - \frac{1}{d}; \frac{\theta}{p} + \frac{1}{q}\right)$. Assume moreover that $\varrho_0 \in L^\beta$ whenever $\beta > 1$, or $\int_\Omega \varrho_0 \log \varrho_0 \, dx < +\infty$ if $\beta = 1$, then there exists some $\tilde{C} > 0$ such that for all $t > 0$, we have $\|\varrho_t\|_{BV} \leq \tilde{C}t^{-\frac{1}{q}}$*

Proof. We will once again use the flow interchange technique, starting from the k -th step of our p -JKO scheme, we look at the ε approximated problem and apply Lemma 2.8 with the function $z \mapsto \frac{1}{\beta-1}z^\beta$ (or $z \mapsto z \log(z)$ in the case $\beta = 1$ with similar computations) to get

$$\begin{aligned} \frac{1}{\beta-1} \int_\Omega (\varrho_k^\tau)^\beta \, dx - \frac{1}{\beta-1} \int_\Omega \varrho_\varepsilon^\beta \, dx &\geq \tau\beta \int_\Omega \varrho_\varepsilon^{\beta-1} f''(\varrho_\varepsilon)^{q-1} |\nabla \varrho_\varepsilon|^q \, dx \\ &\geq C\beta\tau \int_\Omega \varrho_\varepsilon^{\beta-1-\theta(q-1)} |\nabla \varrho_\varepsilon|^q \, dx. \end{aligned}$$

Now, using Hölder's inequality we obtain

$$\int_\Omega |\nabla \varrho_\varepsilon| \, dx = \int_\Omega \varrho_\varepsilon^{-\gamma} \varrho_\varepsilon^\gamma |\nabla \varrho_\varepsilon| \, dx \leq \left(\int_\Omega \varrho_\varepsilon^{\gamma p} \, dx \right)^{\frac{1}{p}} \left(\int_\Omega \varrho_\varepsilon^{-\gamma q} |\nabla \varrho_\varepsilon|^q \, dx \right)^{\frac{1}{q}},$$

where we chose γ to satisfy $-\gamma q = \beta - 1 - \theta(q-1)$. In order for $\int_\Omega \varrho_\varepsilon^{\gamma p} \, dx$ to be bounded, we want γ to be such that $0 \leq \gamma p \leq \beta$, which is equivalent to $\frac{\theta}{p} + \frac{1}{q} \leq \beta \leq \left(\frac{\theta}{p} + \frac{1}{q}\right)q$. To apply the flow interchange technique, we also need $\beta \geq 1 - \frac{1}{d}$, so that the condition $\theta \geq -\frac{p-1}{d}$, which implies $q\left(\frac{\theta}{p} + \frac{1}{q}\right) \geq 1 - \frac{1}{d}$, allows us to take $\beta = \max\left(1 - \frac{1}{d}; \frac{\theta}{p} + \frac{1}{q}\right)$. We can therefore write

$$\tau \left(\int_\Omega |\nabla \varrho_\varepsilon| \, dx \right)^q \leq \left(\int_\Omega (\varrho_k^\tau)^{\gamma p} \, dx \right)^{\frac{q}{p}} \left(\frac{1}{\beta-1} \int_\Omega (\varrho_k^\tau)^\beta \, dx - \frac{1}{\beta-1} \int_\Omega \varrho_\varepsilon^\beta \, dx \right),$$

and taking the limit $\varepsilon \rightarrow 0$ and summing over steps of the scheme gives

$$\sum_k \tau \left(\int_\Omega |\nabla \varrho_k^\tau| \, dx \right)^q \leq \tilde{C}.$$

Using Corollary 5.3 we can obtain

$$\int_\Omega |\nabla \bar{\varrho}_t^\tau| \leq \tilde{C}t^{-\frac{1}{q}},$$

and passing to the limit $\tau \rightarrow 0$ we have $\|\varrho_t\|_{BV(\Omega)} \leq \tilde{C}t^{-\frac{1}{q}}$ for all $t > 0$. \square

A Discrete integral convergence lemma

Theorem A.1. *Let $f : [0, T] \rightarrow \mathbb{R}_+$ be a L^1 function. For all $\varepsilon > 0$, there exists $0 = t_0 < t_1 < t_2 < \dots < t_{n+1} = T$ such that for all $i \in \{0, \dots, n\}$, $t_{i+1} - t_i \leq \varepsilon$ and*

$$\left| \sum_{i=1}^n f(t_i)(t_{i+1} - t_i) - \int_0^T f(x) dx \right| \leq \varepsilon.$$

The $(t_i)_{i=1, \dots, n}$ can also be taken from a set $[0; T] \setminus A$ where A has zero Lebesgue measure.

Proof. For the sake of clarity we will pick arbitrary $\varepsilon, w > 0$ corresponding respectively to the maximum tolerated error in approximating the integral and the maximum mesh size, then one can set $w = \varepsilon$ at the end of the proof.

By absolute continuity of the Lebesgue integral, there exists some $v > 0$ such that for all measurable subsets E of $[0, T]$ with $|E| \leq v$ we have

$$\int_E f(x) dx \leq \frac{\varepsilon}{8} \tag{A.1}$$

By the dominated convergence theorem, there exists some $M > 1$ such that

$$|\{x \in [0, T]; f(x) > M\}| < \min\left(\frac{v}{2}, \frac{w}{2}\right).$$

Let $H = \{f(x) \leq M\} \setminus A$. By the Lebesgue differentiation theorem, for a.e. $x \in H$, there exists $\delta(x)$ such that for all intervals I containing x of length less than $\delta(x)$, we have

$$\left| \frac{1}{|I|} \int_I f(y) dy - f(x) \right| \leq \frac{\varepsilon}{8(T+1)} \tag{A.2}$$

let G be the (full measure) subset of H such that the above holds.

For every $x \in G$, consider the closed intervals $[x, x+h]$, where h ranges across all positive values such that $f(x+h) - f(x) \leq \frac{\varepsilon}{8}$, and $h \leq \min(\frac{\varepsilon}{8M}, \frac{w}{2}, \delta(x))$.

The above collection, ranging across x forms a Vitali cover of the set G (because otherwise the Lebesgue differentiation theorem would be contradicted). By Vitali's covering lemma, there exists a countable subcollection I_i such that $G \setminus \cup_i I_i$ has measure 0, and the I_i are pairwise disjoint. By the Lebesgue dominated convergence theorem, there exists some $N > 0$ such that $|G \setminus \cup_{i=1}^N I_i| \leq \min(\frac{\varepsilon}{8M}, \frac{v}{2}, \frac{w}{2})$. Now we write each I_i as (a_i, b_i) , and relabel them so that they are in increasing order. We now take the partition given by the points t_j to be $a_1, b_1, \dots, a_N, b_N$. We denote by A_i, B_i the partition cells (a_i, b_i) and (b_i, a_{i+1}) , respectively, with the convention that $a_{N+1} = T$; and define $S(A_i) = f(a_i)(b_i - a_i)$; $S(B_i) = f(b_i)(a_{i+1} - b_i)$. Finally, we call $\mathcal{R}(f)$ the left Riemann sum of f over this partition. We have:

$$\left| \mathcal{R}(f) - \int_0^T f dx \right| \leq \sum_{i=1}^N \left| \int_{A_i} f dx - S(A_i) \right| + \sum_{j=1}^N \left| \int_{B_j} f dx - S(B_j) \right| + \left| \int_{[0, a_1]} f dx \right|$$

Using [A.2](#), we have:

$$\left| \int_{A_i} f \, dx - S(A_i) \right| \leq |A_i| \left(\frac{\varepsilon}{8(T+1)} \right)$$

Using [A.1](#) and noticing that $\cup_{j=1}^N B_j \subset [0, T] \setminus \cup_{i=1}^N I_i$ whose measure is less than v we have

$$\begin{aligned} \sum_{j=1}^N \left| \int_{B_j} f \, dx - S(B_j) \right| &\leq \left| \int_{\cup_{j=1}^{N-1} B_j} f \, dx \right| + \sum_{j=1}^{N-1} |S(B_j)| \\ &\leq \frac{\varepsilon}{8} + \sum_{j=1}^{N-1} |S(B_j)| \\ &\leq \frac{\varepsilon}{8} + \sum_{j=1}^N |B_j| f(b_j) \\ &\leq \frac{\varepsilon}{8} + \sum_{j=1}^N |B_j| (f(a_j) + \frac{\varepsilon}{8}) \\ &\leq \frac{\varepsilon}{8} + \left(\sum_{j=1}^N |B_j| \right) (M + \frac{\varepsilon}{8}) \\ &= \frac{\varepsilon}{8} + \frac{\varepsilon}{8M} (M + \frac{\varepsilon}{8}) \\ &\leq \frac{\varepsilon}{8} + \frac{\varepsilon}{8} + \frac{\varepsilon^2}{16M} \\ &= \frac{3\varepsilon}{8} \end{aligned}$$

Now for the last term, the domain of integration has measure less than v by construction, and therefore

$$\left| \int_{[0, a_1]} f \, dx \right| \leq \frac{\varepsilon}{8}.$$

Thus putting all the above together, we have

$$\begin{aligned} \left| \mathcal{R}(f) - \int_0^T f \, dx \right| &\leq \frac{\varepsilon}{8(T+1)} \sum_{i=1}^N |A_i| + \frac{3\varepsilon}{8} + \frac{\varepsilon}{8} \\ &\leq \frac{\varepsilon}{8(T+1)} T + \frac{3\varepsilon}{8} + \frac{\varepsilon}{8} \\ &\leq \varepsilon \end{aligned}$$

Finally, using that $\left| [0, T] \setminus \cup_{i=1}^N I_i \right| \leq |H^c| + \left| G \setminus \cup_{i=1}^N I_i \right| \leq w$ and the fact that the intervals are disjoint, we get that the mesh is of size at most w , and the first point of the partition is at most w away from 0. \square

References

- [1] Martial Agueh. “Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory”. In: *Adv. Differ. Equ.* 10.3 (2005), pp. 309–360.
- [2] Hans W. Alt and Stephan Luckhaus. “Quasilinear Elliptic-Parabolic Differential Equations.” In: *Mathematische Zeitschrift* 183 (1983), pp. 311–342.
- [3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the spaces of probability measures*. Lectures in Mathematics. Birkhäuser, 2008.
- [4] Adrien Blanchet and Guillaume Carlier. “Optimal Transport and Cournot-Nash Equilibria”. In: *Mathematics of Operations research* (2015), pp. 125–145.
- [5] Andrea Braides. *Gamma-convergence for Beginners*. Oxford lecture series in mathematics and its applications. Oxford University Press. ISBN: 9780198507840.
- [6] Thibault Caillet. “The five gradients inequality for non quadratic costs”. In: *Comptes Rendus. Mathématique* 361 (2023), pp. 715–721.
- [7] Clément Cancès et al. “Finite elements for Wasserstein W_p gradient flows”. Preprint. 2022.
- [8] Ennio De Giorgi. “New problems on minimizing movements”. In: *Boundary Value Problems for PDE and Applications, C. Baiocchi and J. L. Lions, eds., Masson* (1993), pp. 81–98.
- [9] Guido De Philippis et al. “BV estimates in Optimal Transportation and Applications”. In: *Arch. Ration. Mech. Anal.* 212.2 (2020), pp. 829–860.
- [10] Simone Di Marino, Simone Murro, and Emanuela Radici. “The five gradients inequality on differentiable manifolds”. arxiv.org/abs/2307.11451. 2023.
- [11] Simone Di Marino and Filippo Santambrogio. “JKO estimates in linear and non-linear Fokker-Planck equations, and Keller-Segel: L^p and Sobolev bounds”. to appear in *Ann. Inst. H. Poincaré C, Anal. Non Linéaire*. 2019. eprint: [1911.10999](https://arxiv.org/abs/1911.10999).
- [12] Lawrence Craig Evans. *Weak Convergence Methods for Nonlinear Partial Differential Equations*. CBMS Regional Conference Ser. in Mathematics Series. American Mathematical Soc., 1990. ISBN: 9780821807248.
- [13] Richard Jordan, David Kinderlehrer, and Felix Otto. “The Variational Formulation of the Fokker-Planck Equation”. In: *SIAM Journal on Mathematical Analysis* 29.1 (1998), pp. 1–17.
- [14] Robert J. Mc Cann and Marjolaine Puel. “Constructing a relativistic heat flow by transport time steps”. In: *Annales de l’Institut Henri Poincaré C, Analyse non linéaire* 26.6 (2009), pp. 2539–2580. ISSN: 0294-1449.
- [15] Robert J. McCann. “A Convexity Principle for Interacting Gases”. In: *Advances in Mathematics* 128 (1997), pp. 153–179.
- [16] Felix Otto. “Doubly Degenerate Diffusion Equations as Steepest Descent”. unpublished.
- [17] Filippo Santambrogio. “{Euclidean, metric, and Wasserstein} gradient flows: an overview”. In: *Bulletin of Mathematical Sciences* 7.1 (2017), pp. 87–154.

- [18] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*. Vol. 87. Progress in Non-linear Differential Equations and their Applications. Birkhäuser, 2015.
- [19] Cédric Villani. *Optimal Transport, Old and New*. Vol. 338. Grundlehren der Mathematischen Wissenschaften. Springer, 2009.
- [20] Cédric Villani. *Topics in Optimal Transportation*. Vol. 58. Graduate Studies in Mathematics. American Mathematical Society, 2003.