

FISHER INFORMATION AND CONTINUITY ESTIMATES FOR NONLINEAR BUT 1-HOMOGENEOUS DIFFUSIVE PDES (VIA THE JKO SCHEME)

THIBAUT CAILLET AND FILIPPO SANTAMBROGIO

ABSTRACT. In this short paper we prove, using the JKO scheme, that quantities such as the Fisher information stay bounded or decrease across time for a family of 1-homogeneous diffusive PDEs. As a corollary, we prove that moduli of continuity are conserved across time for the solutions of those PDEs.

1. INTRODUCTION

The now classical JKO scheme, introduced in the seminal paper [18], and generalized later by many authors, is an approximation scheme used to construct solutions to nonlinear diffusion equations of the type

$$\partial_t \varrho = \nabla \cdot (\varrho \nabla c^*(\nabla f'(\varrho))),$$

where $c : \mathbb{R}^d \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ and $f : \mathbb{R}_+ \rightarrow \mathbb{R} \cup \{+\infty\}$ are convex functions. Among examples of PDEs which have this form, we can obtain equations of the form $\partial_t \varrho = \Delta_p(\phi(\varrho))$ (which are sometimes written in the form $\partial_t(\beta(u)) = \Delta_p u$, using $u = \phi(\varrho)$ and $\beta = \phi^{-1}$). These equations are usually called *doubly nonlinear*: they are a combination of porous-medium type equations (of the form $\partial_t \rho = \Delta(\phi(\rho))$, see [24]) and parabolic p -Laplacian equations (of the form $\partial_t u = \Delta_p u$, see [19]).

The (generalized) JKO scheme consists in building a time discrete approximation of such equations, which is of the form

$$\varrho_{k+1}^h = \operatorname{argmin}_\rho W_{c_h}(\rho, \varrho_k^h) + \mathcal{F}(\rho),$$

where $h > 0$ is the discrete time step, W_{c_h} is the transport cost for the cost $c_h = hc(\frac{\cdot}{h})$ (when $c(z) = \frac{1}{p}|z|^p$ we obtain $W_{c_h} = \frac{1}{ph^{p-1}}W_p^p$) and $\mathcal{F}(\varrho) = \int f(\varrho(x)) dx$. To prove the convergence of this scheme, it is often useful to derive estimates on the discrete solutions (ϱ_k^h) to gain some compactness that allows to pass to the limit, as it is done for example in [2, 21, 17] and many other papers. Some other estimates, derived on the discrete scheme, may be used to study the behaviour of the limit continuous PDE, see for instance [14, 10, 6]. In [23], the authors use discrete estimates

Date: Submitted July 13, 2024; accepted March 7, 2025.

2020 Mathematics Subject Classification. 35K40, 35K92, 49Q22.

Key words and phrases. Gradient flows, doubly nonlinear parabolic equations, Fisher information, relativistic heat equation.

obtained on the scheme to prove a stronger convergence in the case of the Fokker-Planck equation than what was originally proven. This proof used, for instance, a bound on the Hölder norms, obtained from a non-optimal $W^{1,p}$ assumption on the initial condition.

We restrict our analysis to the case where $\mathcal{F}(\varrho) = \mathcal{E}(\varrho) = \int \varrho \log(\varrho) dx$, meaning $\nabla f'(\varrho) = \nabla \log(\varrho)$. This equation is in some sense the most linear one, for fixed cost c , among the doubly non-linear equations that we presented at the beginning. Its almost linear behavior lies in the fact that with this logarithmic choice the equation becomes 1-homogeneous: if ρ is a solution, then $\lambda\rho$ as well for any $\lambda > 0$. This class of equations has already been widely studied, and we cite for instance [16] for an existence result (the relevant case in such a paper is case II), [8] for its connections with functional inequalities, and [1, 3] for the study of the trend to equilibrium (the 1-homogeneous case is called in [3] “Generalized heat equation”). As an example of this class of equations, we can think at the following one: $\partial_t \rho = \Delta_p(\rho^{q-1})$, where p and q are conjugate exponents (so that $(p-1)(q-1) = 1$).

For this very class of PDEs we can adapt results from [10] which readily generalize in our context thanks to the recent extension of the so called “five gradients inequality” (see [5, 9]) to other transport costs than the quadratic one. We prove that the Fisher information, as well as many generalizations of this quantity, decreases along the scheme, and therefore also decreases in time for the limit PDE. We underline that a direct computation of the time-derivative of the Fisher information on the limit equation does not show in a straightforward way that these quantities decrease in time. This is reminiscent of similar results (on very different equations) recently presented in [13, 15]: in these papers the authors show that the Fisher information decreases but their analysis is based on clever tricks as its time derivative was not easily seen to be negative.

As a limit case, we also obtain that the Lipschitz constant of $\log \rho$ is preserved across the iterations. This in turn implies that any modulus of continuity of $\log \rho$ is also preserved along iterations of the scheme, and therefore pass to the limit in the time continuous PDE. This analysis is done first for more classical cost functions c which are finite everywhere and then extended to the so-called “relativistic” costs used for example in the construction of solutions to the relativistic heat equation (see [20])

$$\partial_t \varrho = \nabla \cdot \left(\varrho \frac{\nabla \varrho}{\sqrt{\varrho^2 + |\nabla \varrho|^2}} \right).$$

In the whole paper, we will assume that Ω is the closure of an open bounded convex subset of \mathbb{R}^d . We take $c : \mathbb{R}^d \rightarrow \mathbb{R}_+$ to be C^1 , strictly convex, radially symmetric and such that $c(0) = 0$. Radial symmetry can be removed from the assumptions, as it will be explained later, and the finiteness of the cost will be removed in Section 3.2.

2. PRELIMINARIES

We refer to [4, 22, 25] for the general theory of optimal transport and its application to gradient flows, and we compile below a selection of helpful facts we will use in the sequel.

Theorem 2.1. *Let $\rho, g \in \mathcal{M}_+(\Omega)$ be two non-negative measures on Ω with the same mass. The following statements are classical :*

(1) *The problem*

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

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

(2) *We have*

$$W_c(\mu, \nu) = \max \left\{ \int_{\Omega} \varphi d\mu + \int_{\Omega} \psi d\nu \quad ; \quad \varphi(x) + \psi(y) \leq c(x-y) \quad \forall x, y \in \Omega \right\}.$$

The optimal φ and ψ are called Kantorovich potentials, and they can be taken c -concave, satisfying

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

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

(3) *Since c is strictly convex, $c^* \in C^1$ and the optimal transport map T is given by*

$$T(x) = x - \nabla c^*(\nabla \varphi(x)),$$

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

Definition 2.2. We define $\mathcal{E} : \mathcal{M}_+(\Omega) \rightarrow \mathbb{R} \cup \{+\infty\}$ by

$$\mathcal{E}(\mu) := \begin{cases} \int_{\Omega} \rho \log(\rho) dx & \text{if } \rho \ll dx \text{ and } \rho = \frac{d\mu}{dx}, \\ +\infty & \text{otherwise.} \end{cases}$$

Since we will deal only with absolutely continuous measures, we will use the usual abuse of notation of denoting μ by its density ρ .

Theorem 2.3. *Let $g \in \mathcal{M}_+(\Omega)$ be a non-negative measure, and set $c_h(z) := hc(z/h)$. The problem*

$$\rho \in \operatorname{argmin}_{\rho} W_{c_h}(\rho, g) + \mathcal{E}(\rho) \tag{2.1}$$

admits a unique solution which we denote $\Pi(g)$. Given g_1, g_2 two non-negative measures, if $g_1 \leq g_2$, then $\Pi(g_1) \leq \Pi(g_2)$, and for $\lambda > 0$, $\Pi(\lambda g) = \lambda \Pi(g)$.

Proof. Existence and uniqueness follow from the direct method in the calculus of variations, since both functionals in the minimization are lower semi-continuous and convex, and \mathcal{E} is strictly convex (see for example [2], section 2.1.). The second claim is a direct application of the L^1 contraction principle proved in Theorem 1.3 from [17]. For the last claim, we have

$$\begin{aligned} \inf_{\varrho \in \mathcal{M}_+(\Omega), \int \varrho = \int \lambda g} W_{c_h}(\varrho, \lambda g) + \mathcal{E}(\varrho) &= \inf_{\varrho \in \mathcal{M}_+(\Omega), \int \varrho = \int g} W_{c_h}(\lambda \varrho, \lambda g) + \mathcal{E}(\lambda \varrho) \\ &= \lambda \left(\inf_{\varrho \in \mathcal{M}_+(\Omega), \int \varrho = \int g} W_{c_h}(\varrho, g) + \mathcal{E}(\varrho) \right) + \lambda \log(\lambda) \int_{\Omega} g \, dx, \end{aligned}$$

so that the solution is $\lambda \Pi(g)$. \square

The following lemma, proved in [5] (see also [9]) as a generalization of the results of [7], is useful to prove that suitable quantities decrease along the steps of the JKO scheme.

Lemma 2.4. *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 and $H \in C^1(\mathbb{R}^d \setminus \{0\})$ be a radially symmetric convex function, then the following inequality holds*

$$\int_{\Omega} (\nabla \varphi \cdot \nabla H(\nabla \varphi) + \nabla g \cdot \nabla H(\nabla \psi)) \, dx \geq 0, \quad (2.2)$$

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 , with the convention that $\nabla H(0) = 0$.

Remark 2.5. The convention $\nabla H(0) = 0$ stems from the fact that the result is proved using an approximation of H by smooth radially symmetric functions, which must all therefore satisfy $\nabla H(0) = 0$.

Using Lemma 2.4, one can prove the following :

Lemma 2.6. *Let H be a radially symmetric convex function, $g \in W^{1,1}(\Omega)$ be non negative and denote by ϱ the solution of eq. (2.1). Then we have*

$$\int_{\Omega} H(\nabla \log g) \, dg \geq \int_{\Omega} H(\nabla \log \varrho) \, d\varrho.$$

Proof. We follow the proof of Proposition 5.1 from [10]. We start from the case where $H \in C^1(\mathbb{R}^d \setminus \{0\})$. We have from the fact that H is convex,

$$\int_{\Omega} H(\nabla \log g) \, dg \geq \int_{\Omega} H(\nabla \psi) \, dg + \int_{\Omega} \nabla H(\nabla \psi) \cdot (\nabla \log g - \nabla \psi) \, dg.$$

Using $g = T_{\#}\varrho$ and $\nabla \psi \circ T = -\nabla \varphi$, together with the optimality condition of eq. (2.1) $\nabla \varphi = -\nabla \log \varrho$, we have

$$\int_{\Omega} H(\nabla \psi) \, dg = \int_{\Omega} H(\nabla \psi \circ T) \, d\varrho = \int_{\Omega} H(-\nabla \varphi) \, d\varrho = \int_{\Omega} H(\nabla \log \varrho) \, d\varrho,$$

and

$$\int_{\Omega} \nabla H(\nabla \psi) \cdot \nabla \psi \, dg = - \int_{\Omega} \nabla H(-\nabla \varphi) \cdot \nabla \varphi \, d\varrho = \int_{\Omega} \nabla H(-\nabla \varphi) \cdot \nabla \varrho \, dx.$$

Using Lemma 2.4, we have

$$\begin{aligned} \int_{\Omega} \nabla H(\nabla \psi) \cdot (\nabla \log g) \, dg + \int_{\Omega} \nabla H(\nabla \varphi) \cdot \nabla \varrho \, dx = \\ \int_{\Omega} \nabla H(\nabla \psi) \cdot \nabla g \, dx + \int_{\Omega} \nabla H(\nabla \varphi) \cdot \nabla \varrho \, dx \geq 0, \end{aligned}$$

so that we finally conclude

$$\int_{\Omega} H(\nabla \log g) \, dg \geq \int_{\Omega} H(\nabla \log \varrho) \, d\varrho.$$

By approximation we can get rid of the assumption $H \in C^1(\mathbb{R}^d \setminus \{0\})$ by passing the above inequality to the limit and obtain the result for any convex and radially symmetric function H . \square

Remark 2.7. We observe that the assumption that H is radially symmetric can be weakened. H was supposed to be radial in [5] in order to handle the boundary term in the integration by parts which is needed in the proof of Lemma 2.4, but it is observed that we do not need neither c nor H to be radial, but only ∇c^* to be parallel to ∇H . Then we observe that in our proof (and in that of [5]), we freely moved the minus sign inside and outside of ∇H , assuming at least H to be even. This assumption can also be removed, but in this case the five-gradients inequality becomes

$$\int_{\Omega} (\nabla \varrho \cdot \nabla H(\nabla \varphi) - \nabla g \cdot \nabla H(-\nabla \psi)) \, dx \geq 0.$$

With this modification, the result of Lemma 2.6 still holds.

3. ESTIMATES ON STEPS OF THE JKO SCHEME

3.1. Finite costs.

Definition 3.1. For $h > 0$ and $\varrho_0 \in L^1(\Omega)$ such that $\mathcal{E}(\varrho_0) < +\infty$, we recursively define the sequence $(\varrho_k^h)_k$ by

$$\varrho_{k+1}^h = \operatorname{argmin}_{\rho} W_{c_h}(\rho, \varrho_k^h) + \mathcal{E}(\rho),$$

which is well defined because of Theorem 2.3

The following theorem is taken from [10] and can directly be generalized to our context thanks to Lemmas 2.4 and 2.6 with the same proof.

Theorem 3.2. *Let $\varrho_0 \in L^1(\Omega)$. We have*

$$\int_{\Omega} H(\nabla \log \varrho_0) d\varrho_0 \geq \int_{\Omega} H(\nabla \log \varrho_k^h) d\varrho_k^h$$

for all h, k , where H is a radially symmetric and convex function. In particular, the following estimates are uniform in h and k :

- (1) if $\log(\varrho_0)$ is L -Lipschitz, then $\log(\varrho_k^h)$ is L -Lipschitz;
- (2) for $p > 1$, if $\varrho_0^{1/p} \in W^{1,p}(\Omega)$, then $(\varrho_k^h)^{1/p} \in W^{1,p}(\Omega)$ and $\|(\varrho_k^h)^{1/p}\|_{W^{1,p}} \leq \|\varrho_0^{1/p}\|_{W^{1,p}}$;
- (3) if $\varrho_0 \in BV(\Omega)$, then $\varrho_k^h \in BV(\Omega)$ and $\|\varrho_k^h\|_{BV} \leq \|\varrho_0\|_{BV}$.
- (4) if $\varrho_0 \in W^{1,1}(\Omega)$, all ϱ_k^h belong to a weakly compact subset of $W^{1,1}(\Omega)$.

Proof. We only give the main ideas, as the details can be found in [10]. The first statement is a direct consequence of Lemma 2.6, iterated along the steps of the scheme. For 1., we pick H to be the indicator of the centered ball of radius L (by indicator we mean here the function which is 0 on the ball and $+\infty$ outside it). For 2. we take $H(z) = |z|^p$ with $p > 1$, and 3. is covered by the case $p = 1$. For 4., one can use the Dunford-Pettis theorem so find a superlinear convex function H such that $H(\nabla \log(\varrho_0)) \in L^1(\varrho_0)$, which decreases at each step, thus providing the weak compactness. \square

Remark 3.3. Many authors use bounds on the generalized Fisher information to help prove the convergence of the JKO scheme to the limit PDE, see for example [2, 20]. However, most often these bounds are integral in time, i.e. only bounds on $\int_0^T \int H(\nabla \varrho_t) d\varrho_t dt$ are used, and hence proven.

The above theorem includes the fact that the Lipschitz constant of the logarithm is preserved along the iterations of this JKO scheme. We want now to extend this to other moduli of continuity. We will use the following abstract fact. This statement, which we did not find stated in this very version elsewhere in the literature, is interesting in itself and is remindful of other results in the theory of LaxOleinik semigroup (see for instance [12, 11]).

Theorem 3.4. *Let $\pi : L^\infty(\Omega) \rightarrow L^\infty(\Omega)$ be an operator such that*

- (1) For all $u, v \in L^\infty(\Omega)$, if $u \geq v$, then $\pi(u) \geq \pi(v)$.
- (2) For all $\lambda \in \mathbb{R}$ and $u \in L^\infty(\Omega)$, $\pi(\lambda + u) = \lambda + \pi(u)$.
- (3) For all $u \in L^\infty(\Omega)$, if u is k -Lipschitz, then $\pi(u)$ is also k -Lipschitz.

Then, if u admits a concave modulus of continuity ω , then $\pi(u)$ admits the same modulus of continuity.

Proof. If u admits ω as a modulus of continuity, we start by approximating it with L -Lipschitz functions : for $x \in \Omega$, set

$$u_L(x) = \inf_{y \in \Omega} L|x - y| + u(y),$$

so that u_L is L -Lipschitz. Of course u_L satisfies $u_L \leq u$ (by taking $y = x$). Furthermore, from the inequality

$$u(x) - u(y) \leq \omega(|x - y|),$$

we deduce that we have

$$L|x - y| + u(y) \geq L|x - y| + u(x) - \omega(|x - y|),$$

so that, passing to the inf in y , we obtain

$$u_L(x) \geq u(x) + \alpha(L),$$

where $\alpha(L) = \inf_{\text{diam}(\Omega) > r > 0} Lr - \omega(r) \leq 0$. We can therefore conclude that

$$u \geq u_L \geq u + \alpha(L).$$

Applying π and using its properties 1 and 2 we obtain

$$\pi(u) \geq \pi(u_L) \geq \pi(u) + \alpha(L).$$

For $x, y \in \Omega$, since $\pi(u_L)$ is also L -Lipschitz, we can write

$$\pi(u)(x) - \pi(u)(y) \leq \pi(u_L)(x) - \pi(u_L)(y) - \alpha(L) \leq L|x - y| - \alpha(L).$$

Since ω is concave, we define $\partial^+ \omega$ its super-differential. Let us choose L such that $L \in \partial^+ \omega(|x - y|)$ (note that ω is also non-decreasing, so we can take $L \geq 0$; also note that the superdifferential is non-empty when $r > 0$ because ω is finite on \mathbb{R}_+). This implies that $r \mapsto Lr - \omega(r)$ (which is a convex function) is minimized at $r = |x - y|$ because $0 \in \partial(r \mapsto Lr - \omega(r))$. Then we obtain for this L the equality $\alpha(L) = L|x - y| - \omega(|x - y|)$, and thus

$$\pi(u)(x) - \pi(u)(y) \leq \omega(|x - y|),$$

so that $\pi(u)$ admits ω as a modulus of continuity. \square

Using the above theorem, we can then prove the following:

Theorem 3.5. *Let $\varrho_0 \in L^1(\Omega)$ be such that $\log(\varrho_0)$ admits ω as a modulus of continuity, then for all k and $h > 0$, $\log(\varrho_k^h)$ also admits ω as a modulus of continuity.*

Proof. One only has to apply Theorem 3.4 with $\pi = \log \circ \Pi \circ \exp$ to the L^∞ function $\log(\varrho_0)$, where Π is defined in theorem 2.3. Assumptions 1 and 2 are satisfied because of Theorem 2.3 since the \log function is increasing and satisfies $\log(ab) = \log(a) + \log(b)$ for $a, b > 0$. The last assumption is satisfied because of the first item of Theorem 3.2. We therefore obtain that $\log(\varrho_1^h)$ admits ω as a modulus of continuity, and iterating the above at each step of the JKO scheme gives the result. \square

We recall the following theorem (see [2], Theorem 3.11) which proves the convergence of the JKO scheme, under the additional conditions on the cost function c :

$$(1) \lim_{|z| \rightarrow +\infty} \frac{c(z)}{|z|} = +\infty,$$

(2) there exists $\alpha, \beta > 0$ such that

$$\alpha|z|^p \leq c(z) \leq \beta(1 + |z|^p) \quad \text{for some } p > 1. \quad (3.1)$$

Theorem 3.6. *Suppose $\log(\varrho_0) \in L^\infty(\Omega)$, and denote by ϱ^h the piecewise constant (in time) interpolation of $(\varrho_k^h)_k$. Then there exists $\varrho \in L^1(\Omega \times [0, T])$ such that $\log(\varrho(t)) \in L^\infty(\Omega)$ for all $t \in [0, T]$, and such that ϱ_h strongly converges to ϱ in $L^1(\Omega \times [0, T])$. Moreover, ϱ is the unique weak solution of*

$$\begin{cases} \partial_t \varrho = \nabla \cdot (\varrho \nabla c^*(\nabla \log(\varrho))), \\ \varrho(0) = \varrho_0. \end{cases} \quad (3.2)$$

Another proof of convergence, in the case where the cost c is exactly a power, is provided in ([6], Theorem 1.1). In this reference the bound from below and above on the initial density ϱ_0 is removed, but only the existence is proven, not the uniqueness.

Remark 3.7. Ω is assumed to be smooth in [2], but one can check that its convexity (which implies Lipschitz regularity) is enough in the paper, as it is done in [6].

Remark 3.8. In the above theorem, as one can see in Lemma 3.1 from [2], or in section 3.2 of [6], condition (3.1) implies that a suitable interpolation of (ϱ^h) is equicontinuous in $C^0([0, T], \mathbb{W}_p(\Omega))$, where $\mathbb{W}_p(\Omega)$ denotes the space $\mathcal{P}(\Omega)$ endowed with the p -Wasserstein distance.

From Theorem 3.5 and Theorem 3.6 we deduce the following :

Theorem 3.9. *Let $\varrho_0 \in L^1(\Omega)$ be non negative. If $\log(\varrho_0)$ admits ω as a modulus of continuity, then for all $t > 0$, $\log(\varrho(t))$ admits ω as a modulus of continuity, where ϱ is the solution of eq. (3.2).*

Proof. Since $\log(\varrho_0)$ is continuous, we can construct the unique solution of eq. (3.2) using Theorem 3.6. Applying Theorem 3.5, we have for all k that $\log(\varrho_k^h)$ admits ω as a modulus of continuity, and therefore for all $t \geq 0$, we have

$$|\log(\varrho^h(t, x)) - \log(\varrho^h(t, y))| \leq \omega(|x - y|).$$

Because one can show that (ϱ^h) is uniformly bounded from above and below (see [2] Proposition 2.3), this provides an equicontinuity estimate in space for the sequence (ϱ^h) , which is uniform in time. Recalling remark 3.8, we can conclude that (ϱ^h) converges uniformly in $\Omega \times [0, T]$ to ϱ , because we have equicontinuity both in time and space. Passing to limit $\varrho^h(t) \rightarrow \varrho(t)$ as $h \rightarrow 0$ in the above equation allows us to conclude that $\varrho(t)$ admits ω as a modulus of continuity \square

When we have uniqueness to eq. (3.2), which is the case for example is we look at bounded solutions, we can show that the generalized Fisher information decreases along the flow :

Theorem 3.10. *Let $\varrho_0 \in L^1(\Omega)$ be non negative and such that $\log(\varrho_0) \in L^\infty(\Omega)$. Denote by ϱ the solution of eq. (3.2), then the generalized Fisher information*

$$\int_{\Omega} H(\nabla \log \varrho) \, d\varrho$$

is non increasing in time.

Proof. From Theorems 3.2 and 3.6, we know that the discrete solution $(\tilde{\varrho}_k^h)_k$ constructed from the JKO scheme started from $\varrho(t)$ for any $t \geq 0$ satisfies

$$\int_{\Omega} H(\nabla \log \varrho(t)) \, d\varrho(t) \geq \int_{\Omega} H(\nabla \log \tilde{\varrho}_k^h) \, d\tilde{\varrho}_k^h.$$

By uniqueness of the solution, and the strong convergence of the discrete solutions provided by Theorem 3.6, we have $\tilde{\varrho}^h(s) \rightarrow \varrho(t+s)$ and the lower semi continuity of the right hand side (because H is convex) gives

$$\int_{\Omega} H(\nabla \log \varrho(t)) \, d\varrho(t) \geq \int_{\Omega} H(\nabla \log \varrho(t+s)) \, d\varrho(t+s)$$

for any $s \geq 0$. □

Remark 3.11. Doing this computation directly at the continuous level by differentiating the generalized Fisher Information and trying to show that its derivative is non-positive, seems to be, in our opinion, rather complicated. In particular, after some simplification we obtain the following computations (subject to caution and ignoring boundary terms), where we denote $u = \log \varrho$. If ϱ solves eq. (3.2) then the equation solved by u is

$$\partial_t u = K_{c^*} + D_{c^*},$$

where

$$D_{c^*} := \nabla \cdot (\nabla c^*(\nabla u)) = \sum_{j,k} c_{jk}^*(\nabla u) u_{jk}, \quad K_{c^*} := \nabla u \cdot \nabla c^*(\nabla u) = \sum_i u_i c_i^*(\nabla u).$$

We also use $\partial_t \varrho = \varrho \partial_t u$ and with an integration by parts we obtain

$$\begin{aligned} \frac{d}{dt} \int_{\Omega} \varrho H(\nabla u) &= \int_{\Omega} \partial_t \varrho H(\nabla u) + \varrho H_i(\nabla u) \partial_t u_i = \int_{\Omega} (\varrho H(\nabla u) - (\varrho H_i(\nabla u))_i) \partial_t u \\ &= \int_{\Omega} \varrho (K_{c^*} + D_{c^*}) (H(\nabla u) - K_H - D_H), \end{aligned}$$

where K_H and D_H are defined as for K_{c^*} and D_{c^*} , i.e. $D_H := \sum_{j,k} H_{jk}(\nabla u) u_{jk}$, $K_H := \sum_i u_i H_i(\nabla u)$. We do not find it easy to see that the r.h.s. is negative (even in the case $H = c^*$). On the other hand, the Lipschitz regularity result contained in Theorem 3.2 can indeed be recovered by a maximum principle using the equation satisfied by $u = \log \varrho$: differentiating it provides

$$\partial_t u_h = c_{jkl}^*(\nabla u) u_{lh} u_{jk} + c_{jk}^*(\nabla u) u_{jkh} + u_{ih} H_i(\nabla u) + u_i H_{il}(\nabla u) u_{lh}.$$

Multiplying by u_h and looking at $\max |\nabla u|^2$ allows to obtain

$$\frac{d}{dt} \max |\nabla u|^2 \leq 0.$$

Yet, of course, this formal computation would require to be suitably justified.

3.2. Extension to relativistic costs. In [20], the authors generalize the construction from [2] to “relativistic” cost functions, meaning cost functions of the form

$$c(z) = \begin{cases} \tilde{c}(z) & \text{if } |z| \leq 1, \\ +\infty & \text{if } |z| > 1, \end{cases}$$

where $\tilde{c} \in C([0, 1]) \cap C^2([0, 1))$ is a strictly convex function, with $|\nabla \tilde{c}(z)| \rightarrow +\infty$ when $|z| \rightarrow 1$. The main equation of interest entering in this framework is the so called Relativistic Heat Equation

$$\partial_t \varrho = \nabla \cdot \left(\varrho \frac{\nabla \varrho}{\sqrt{\varrho^2 + |\nabla \varrho|^2}} \right),$$

where $\tilde{c}(z) = 1 - \sqrt{1 - |z|^2}$. Their construction relies on an approximation argument of the irregular cost function c , by smoother cost functions c_ε by inf-convolution, which allows us to apply our previous results before passing to the limit.

Lemma 3.12. *Define*

$$c_\varepsilon(z) = \inf_{w \in \mathbb{R}^d} \left(c(z - w) + \frac{|w|^2}{2\varepsilon} \right),$$

it is a strictly convex and radially symmetric $C^2(\mathbb{R}^d)$ function. If we denote by ϱ_ε^h the solution of eq. (2.1) with the cost c_ε , then as $\varepsilon \rightarrow 0$, ϱ_ε^h weakly converges to the solution ϱ^h of eq. (2.1) with the cost c .

Proof. See Section 3.2 and Lemma 3.2 from [20]. □

As in the previous section, we recursively define the sequence $(\varrho_k^h)_k$. And we have the same results as before :

Theorem 3.13. *Let $\varrho_0 \in L^1(\Omega)$. The following estimates are uniform in h and k :*

- (1) *if $\log(\varrho_0)$ is L -Lipschitz, then $\log(\varrho_k^h)$ is L -Lipschitz;*
- (2) *for $p > 1$, if $\varrho_0^{1/p} \in W^{1,p}(\Omega)$, then $(\varrho_k^h)^{1/p} \in W^{1,p}(\Omega)$ and $\|(\varrho_k^h)^{1/p}\|_{W^{1,p}} \leq \|\varrho_0^{1/p}\|_{W^{1,p}}$;*
- (3) *if $\varrho_0 \in BV(\Omega)$, then $\varrho_k^h \in BV(\Omega)$ and $\|\varrho_k^h\|_{BV} \leq \|\varrho_0\|_{BV}$.*
- (4) *if $\varrho_0 \in W^{1,1}(\Omega)$, all ϱ_k^h belong to a weakly compact subset of $W^{1,1}(\Omega)$.*
- (5) *if $\log(\varrho_0)$ admits ω as a modulus of continuity, then all ϱ_k^h also admit ω as a modulus of continuity.*

Proof. We apply Theorems 3.2 and 3.5 from the previous section to the approximated problem with c_ε , and then pass to the limit $\varepsilon \rightarrow 0$. All estimates pass to the limit by lower semi continuity with respect to the weak convergence. □

Using the convergence result from ([20], Theorem 1.8), we have the following :

Theorem 3.14. *Let $\varrho_0 \in L^1(\Omega)$ be such that $\log(\varrho_0)$ admits ω as a modulus of continuity. The interpolated sequence (ϱ^h) strongly converges in $L^1(\Omega \times [0, T])$ to ϱ , a solution in the sense of distributions of*

$$\partial_t \varrho = \nabla \cdot (\varrho \nabla c^*(\nabla \log(\varrho))), \quad (3.3)$$

which is such that $\log(\varrho(t))$ admits ω as a modulus of continuity for all t .

Proof. We use Theorem 1.8 from [20] to have the convergence of the discrete sequence, and use the last point Theorem 3.13 to prove that the modulus of continuity of $\log(\varrho_0)$ is propagated. Indeed, for the constant in time interpolation ϱ^h , we have for all $t \geq 0$

$$|\varrho^h(t, x) - \varrho(t, y)| \leq \omega(|x - y|),$$

and we can pass to the limit $h \rightarrow 0$. □

Acknowledgments. The authors acknowledge the support of European union via the ERC AdG 101054420 EYAWKAJKOS.

REFERENCES

- [1] Martial Agueh. Asymptotic behavior for doubly degenerate parabolic equations. *Comptes Rendus Mathématique*, 337(5):331–336, 2003.
- [2] Martial Agueh. Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory. *Adv. Differ. Equ.*, 10(3):309–360, 2005.
- [3] Martial Agueh. Rates of decay to equilibria for p -laplacian type equations. *Nonlinear Analysis: Theory, Methods & Applications*, 68(7):1909–1927, 2008.
- [4] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the spaces of probability measures*. Lectures in Mathematics. Birkhäuser, 2008.
- [5] Thibault Caillet. The five gradients inequality for non quadratic costs. *Comptes Rendus. Mathématique*, 361:715–721, 2023.
- [6] Thibault Caillet and Filippo Santambrogio. Doubly nonlinear diffusive PDEs: new existence results via generalized Wasserstein gradient flows. *SIAM J. Math. Anal.*, 56(6):7043–7073, 2024.
- [7] Guido De Philippis, Alpár Richárd Mészáros, Filippo Santambrogio, and Bozhidar Velichkov. BV estimates in Optimal Transportation and Applications. *Arch. Ration. Mech. Anal.*, 212(2):829–860, 2020.
- [8] Manuel Del Pino, Jean Dolbeault, and Ivan Gentil. Nonlinear diffusions, hypercontractivity and the optimal L^p -Euclidean logarithmic Sobolev inequality. *J. Math. Anal. Appl.*, 293(2):375–388, 2004.
- [9] Simone Di Marino, Simone Murro, and Emanuela Radici. The five gradients inequality on differentiable manifolds. *J. Math. Pures Appl. (9)*, 187:294–328, 2024.
- [10] Simone Di Marino and Filippo Santambrogio. JKO estimates in linear and non-linear Fokker-Planck equations, and Keller-Segel: L^p and Sobolev bounds. *Ann. Inst. Henri Poincaré, Anal. Non Linéaire*, 39(6):1485–1517, 2022.
- [11] Lawrence C. Evans. Weak KAM theory and partial differential equations. In *Calculus of variations and nonlinear partial differential equations. Lectures given at the C.I.M.E. summer school, Cetraro, Italy, June 27–July 2, 2005. With a historical overview by Elvira Mascolo*, pages 123–154. Berlin: Springer, 2008.
- [12] Albert Fathi. A weak KAM theorem and Mather’s theory of Lagrangian systems. *C. R. Acad. Sci., Paris, Sér. I*, 324(9):1043–1046, 1997.

- [13] Nestor Guillen and Luis Silvestre. The Landau equation does not blow up. Preprint, arXiv:2311.09420 [math.AP] (2023), 2023.
- [14] Mikaela Iacobelli, Francesco Saverio Patacchini, and Filippo Santambrogio. Weighted ultrafast diffusion equations: from well-posedness to long-time behaviour. *Arch. Rati. Mech. An.*, 232(3):1165–1206, 2019.
- [15] Cyril Imbert, Luis Silvestre, and Cédric Villani. On the monotonicity of the Fisher information for the Boltzmann equation. Preprint, arXiv:2409.01183 [math.AP] (2024), 2024.
- [16] Kazuhiro Ishige. On the existence of solutions of the Cauchy problem for a doubly nonlinear parabolic equation. *SIAM Journal on Mathematical Analysis*, 27(5):1235–1260, 1996.
- [17] Matt Jacobs, Inwon C. Kim, and Jiajun Tong. The L^1 -contraction principle in optimal transport. *Annali Scuola Normale Superiore - Classe di Scienze*, 23(4):1871–1919, 2022.
- [18] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the fokker–planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.
- [19] Peter Lindqvist. *Notes on the p -Laplace equation*, volume 102 of *Rep., Univ. Jyväskylä, Dep. Math. Stat.* Jyväskylä: University of Jyväskylä, Department of Mathematics and Statistics, 2006.
- [20] Robert J. Mc Cann and Marjolaine Puel. Constructing a relativistic heat flow by transport time steps. *Annales de l’Institut Henri Poincaré C, Analyse non linéaire*, 26(6):2539–2580, 2009.
- [21] Felix Otto. Doubly Degenerate Diffusion Equations as Steepest Descent. unpublished.
- [22] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser, 2015.
- [23] Filippo Santambrogio and Gayrat Toshpulatov. Strong L^2H^2 convergence of the JKO scheme for the Fokker-Planck equation. *Arch. Ration. Mech. Anal.*, 248(6):29, 2024. Id/No 99.
- [24] J.L. Vazquez. *The Porous Medium Equation: Mathematical Theory*. Oxford Mathematical Monographs. Clarendon Press, 2007.
- [25] Cédric Villani. *Optimal Transport, Old and New*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer, 2009.

UNIVERSITÉ CLAUDE BERNARD LYON 1, ICJ UMR5208, CNRS, ÉCOLE CENTRALE DE LYON, INSA LYON,
UNIVERSITÉ JEAN MONNET, 69622 VILLEURBANNE, FRANCE.

Email address: caillet@math.univ-lyon1.fr

Email address: santambrogio@math.univ-lyon1.fr