NONLINEAR DIFFUSION: GEODESIC CONVEXITY IS EQUIVALENT TO WASSERSTEIN CONTRACTION

FRANÇOIS BOLLEY AND JOSÉ A. CARRILLO

Abstract. It is well known that nonlinear diffusion equations can be interpreted as a gradient flow in the space of probability measures equipped with the Euclidean Wasserstein distance. Under suitable convexity conditions on the nonlinearity, due to R. J. McCann [12], the associated entropy is geodesically convex, which implies a contraction type property between all solutions with respect to this distance. In this note, we give a simple straightforward proof of the equivalence between this contraction type property and this convexity condition, without even resorting to the entropy and the gradient flow structure.

We consider the nonlinear diffusion equation

$$\frac{\partial u_t}{\partial t} = \Delta f(u_t), \qquad t > 0,\, x \in \mathbb{R}^d \tag{1}$$

where $f(r)$ is an increasing continuous function on $r \in [0, +\infty)$ and C^2 smooth for $r > 0$ such that $f(0) = 0$. The well-posedness theory in $L^1(\mathbb{R}^d)$ for this equation is a classical matter in the nonlinear parabolic PDEs theory developed in the last 40 years, see [16] and the references therein. Following [16, Chap. 9], by a solution we mean a map $u = (u_t)_{t\geq 0} \in C([0,\infty), L^1(\mathbb{R}^d))$, with $u_t \geq 0$ and mass $\int u_t(x)dx = M$ for all t, such that

i) for all $T > 0$, the function $\nabla(f \circ u_t) \in L^2((0,T) \times \mathbb{R}^d)$,

ii) u weakly satisfies equation (1), i.e., it satisfies the identity

$$\int_0^\infty \int_{\mathbb{R}^d} \left\{ \nabla(f \circ u_t) \cdot \nabla \chi - u_t \frac{\partial \chi}{\partial t} \right\} dx\, dt = \int_{\mathbb{R}^d} u_0(x)\, \chi(0,x)\, dx\,,$$

for all compactly supported functions $\chi \in C^1([0,\infty) \times \mathbb{R}^d)$.

For all nonnegative $u_0 \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ with $\int u_0(x)dx = M$, there exists a unique such solution to (1) with initial datum u_0, see [16, Sect. 9.8]. Let us normalize the mass M to unity in the rest of the introduction for convenience.

Equation (1) admits the map

$$\mathcal{U}(u) = \int_{\mathbb{R}^d} U(u(x))\, dx$$

as a Liapunov functional. Here the map $U \in C([0,+\infty)) \cap C^3((0,+\infty))$ is defined in a unique way by the relations $f(r) = rU'(r) - U(r)$ on $(0,+\infty)$ and $U(0) = U'(1) = 0$. The map f being increasing is equivalent to U being strictly convex, since $f'(r) = rU''(r)$, and f being positive on $(0,+\infty)$ is equivalent to $\psi : r \mapsto r^d U(r^{-d})$ being decreasing, since $\psi'(r) = -dr^{d-1} f(r^{-d})$. The map $\mathcal{U}$ can be extended to the set $\mathcal{P}_2(\mathbb{R}^d)$ of Borel probability measures on $\mathbb{R}^d$ with

Date : February 5th, 2014. Mathematics Subject Classification : 28A33, 35B35, 35K10, 37L15, 76S05.
Keywords : nonlinear diffusions, optimal transport, entropy, convexity.

finite second moment, by setting $+\infty$ for non absolutely continuous measures with respect to Lebesgue measure. The seminal work of R. J. McCann [12] shows that $\mathcal{U}$ is (geodesically) displacement convex on the space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ if and only if ψ is moreover convex on $(0, +\infty)$, that is, if and only if

$$(d-1)f(r) \leq d\, r\, f'(r), \qquad r > 0, \tag{2}$$

or equivalently,

$$r \mapsto r^{-1+1/d} f(r) \quad \text{is} \quad \text{nondecreasing} \quad \text{on} \quad (0, +\infty)\,. \tag{3}$$

We also refer to [1, Chap. 9], [8, p. 26], [14, Th. 1.3] or [18, Chap. 17] for this classical notion. In particular, for the porous medium and fast diffusion equations when $f(r) = r^m$, the condition (3) writes $m \geq 1 - 1/d$, which is now classical. Here, W_2 is the Wasserstein distance defined for μ, ν in $\mathcal{P}_2(\mathbb{R}^d)$ by

$$W_2(\mu,\nu)^2 = \inf_{\pi} \iint_{\mathbb{R}^{2d}} |y-x|^2 \, d\pi(x,y)$$

where π runs over the set of measures on $\mathbb{R}^{2d}$ with marginals μ and ν, see [1, 18].

F. Otto [13] has interpreted (1) as the gradient flow of $\mathcal{U}$ in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ and, for an f satisfying (3), deduced the contraction type property

$$W_2(u_t, v_t) \leq W_2(u_0, v_0), \qquad t \geq 0\,, \tag{4}$$

for positive smooth solutions of the Neumann problem for (1) on a smooth open bounded set of $\mathbb{R}^d$, see [13, Eq. (133) and Prop. 1]. This point of view has been extended by L. Ambrosio, N. Gigli and G. Savaré [1], through the deep and very general theory of gradient flows of geodesically convex functionals in metric spaces: in particular, for any u_0 in $\mathcal{P}_2(\mathbb{R}^d)$ there exists a unique so-called gradient flow solution $(S_t u_0)_{t\geq 0}$ to (1)-(3); these solutions satisfy (4), a consequence of the displacement convexity of $\mathcal{U}$ and the key Evolution Variational Inequality. Moreover, $S_t u_0 = u_t$ if $u_0 \in P_2(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ due to the uniqueness of the Cauchy problem for weak solutions in [16, Chap. 9] and gradient flows in [2, Sect. 6.4].

K.-T. Sturm and M. von Renesse [15] have further proved that for the *heat equation* on a Riemannian manifold, the geodesic convexity of the Boltzmann entropy is *equivalent* to (4), and to the Ricci curvature being nonnegative. The equivalence of geodesic convexity of the internal energy to curvature-dimension conditions was extended in [14] to nonlinear diffusions. As shown in [14], they imply the contraction property (4) for nonlinear diffusions. Similar ideas were early used for gradient flows in Aleksandrov spaces, see [11] and the references therein. In this work, we give the proof of the reverse implication, contraction (4) for nonlinear diffusions implies (3), for the first time to our knowledge.

Let us finally remark that the above normalization of the mass is for convenience. Actually, the contraction property (4) holds for all weak solutions with nonnegative initial data in $L^1 \cap L^\infty(\mathbb{R}^d)$ with finite second moment and equal initial mass. For this purpose in the rest we extend the W_2 distance to nonnegative finite measures with equal mass.

In this note, we give a simple straightforward proof of the contraction property (4) for such f, resorting neither to the gradient flow structure, nor even to the functional $\mathcal{U}$, as in [1] or [13]. We first give the *equivalence* between condition (3) on the nonlinearity and the Wasserstein contraction type property (4) for positive smooth solutions on the ball $B_R = \{x \in \mathbb{R}^d, |x| < R\}$:

Theorem 1. *Let f be an increasing continuous map on $[0, +\infty)$, which is C^2 on $(0, +\infty)$ and satisfies $f(0) = 0$ and let $R > 0$ be fixed. Then McCann's displacement convexity condition* (3) *holds if and only if* (4) *holds for all solutions u and v to the Neumann problem*

$$\frac{\partial u_t}{\partial t} = \Delta f(u_t), \quad t > 0, \, x \in B_R, \qquad \frac{\partial u_t}{\partial \eta} = 0, \quad t > 0, \, |x| = R \tag{5}$$

with C^2 positive initial data u_0 and v_0 on $\{|x| \leq R\}$ respectively, with equal mass.

Here, $\frac{\partial u_t}{\partial \eta}$ denotes the normal derivative of u_t on the sphere $|x| = R$. We notice that the well-posedness theory for the Neumann problem for nonlinear diffusions is a classical question, see [16, Chap. 11]. Moreover, the solutions corresponding to positive initial data remain positive and bounded below for all times by a straightforward use of the maximum principle [16, Th. 11.2], and thus the solutions we deal with in Theorem 1 are classical and C^∞ smooth for all times.

As a direct consequence of the approximation procedures developed in [13, Sect. 5.5] and [8, Rem. 18, Prop. 13], one can derive the W_2 contraction property (4) for solutions of the Cauchy problem:

Corollary 2. *Let f be as in Theorem 1 and satisfy McCann's displacement convexity condition* (3)*. Then the contraction property* (4) *holds for all solutions u and v to* (1) *with initial data u_0 and v_0 in $P_2(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ respectively.*

Of course, proving equivalence directly with the contraction property for gradient flows in $\mathbb{R}^d$ would be more natural. The arguments below formally work on $\mathbb{R}^d$, however, we lack enough regularity of the solutions to make them rigorous. The rest of this paper is devoted to the proof of the main result in Theorem 1.

Proof of Theorem 1.
We first prove the *sufficient* part by adapting the strategy of [3]. Let u_0, v_0 be two C^2 positive initial data on $\bar{B}_R = \{|x| \leq R\}$ with equal mass. Then so are the solutions u and v of (5) at all times, by the comparison principle [16, Th. 11.2]. For all $t \geq 0$, we let $\xi_t[u] = -\nabla(f \circ u_t)/u_t$ be the velocity field governing the evolution of u, written as the continuity equation

$$\frac{\partial u_t}{\partial t} + \nabla \cdot (u_t \, \xi_t[u]) = 0, \qquad t > 0, \, x \in B_R.$$

We also let φ_t be a convex map on $\bar{B}_R$ given by the Brenier Theorem such that $\nabla\varphi_t$ pushes forward u_t onto v_t, denoted by $v_t = \nabla\varphi_t \# u_t$, and $\pi_t = (I_{\mathbb{R}^d} \times \nabla\varphi_t)\# u_t$ is optimal in the definition of $W_2(u_t, v_t)$ (see [5], [1, Sect. 6.2.3], or [18, Th. 10.28] for instance). Here, $I_{\mathbb{R}^d}$ denotes the identity map. We observe that $u_t = \nabla\varphi_t^* \# v_t$ and $\pi_t = (\nabla\varphi_t^* \times I_{\mathbb{R}^d})\# v_t$ for the Legendre transform φ_t^* of φ_t.

Then, for all $t \geq 0$,

$$\frac{1}{2}\frac{d}{dt} W_2^2(u_t, v_t) = \iint_{B_R \times B_R} (\xi_t[v](y) - \xi_t[u](x)) \cdot (y - x) \, d\pi_t(x, y). \tag{6}$$

This follows from steps 2 and 3 in the proof of [18, Th. 23.9] since for all $t \geq 0$ the solutions u_t and v_t are C^2 and positive on the compact set $\bar{B}_R$, so that $\xi_t[u]$ and $\xi_t[v]$ are Lipschitz on $\bar{B}_R$. By the definition of the velocity field ξ_t and the image measure property, this is equal to

$$-\iint_{B_R \times B_R} \Big(\frac{1}{v_t(y)} \nabla (f \circ v_t)(y) - \frac{1}{u_t(x)} \nabla (f \circ u_t)(x) \Big) \cdot (y-x) \, d\pi_t(x,y)$$
$$= - \int_{B_R} \nabla (f \circ v_t)(y) \cdot (y - \nabla \varphi_t^*(y)) \, dy + \int_{B_R} \nabla (f \circ u_t)(x) \cdot (\nabla \varphi_t(x) - x) \, dx. \quad (7)$$

Now, the probability densities u_t and v_t are $C^{0,\alpha}$ and bounded from below and above by positive constants on the ball $\bar{B}_R$, so the maps φ_t and φ_t^* are $C^{2,\alpha}$ on $\bar{B}_R$, see [7, Th. 5.1] and [18, Th. 12.50]. Hence, by integration by parts, the term with u in (7) is

$$- \int_{B_R} f \circ u_t(x) \, (\Delta \varphi_t(x) - d) \, dx + \int_{|x|=R} f \circ u(x) \, (\nabla \varphi_t(x) - x) \cdot \frac{x}{|x|} \, dx \leq - \int_{B_R} f \circ u_t \, (\Delta \varphi_t - d).$$

Here we use the fact that $\nabla \varphi_t(x) \in \bar{B}_R$, so that $\nabla \varphi_t(x) - x$ points inwards $\bar{B}_R$ if $|x| = R$, and then $(\nabla \varphi_t(x) - x) \cdot x \leq 0$. Likewise, the term with v in (7) is

$$\begin{aligned} \leq - \int_{B_R} f \circ v_t \, (\Delta \varphi_t^* - d) \quad &= \quad - \int_{B_R} \frac{f\big(v_t(\nabla \varphi_t(x))\big)}{v_t(\nabla \varphi_t(x))} \, (\Delta \varphi^*(\nabla \varphi_t(x)) - d) \, u_t(x) \, dx \\ &= \quad - \int_{B_R} \det \nabla^2 \varphi_t(x) \, f\Big(\frac{u_t(x)}{\det \nabla^2 \varphi_t(x)} \Big) \, \big(\Delta \varphi_t^*(\nabla \varphi_t(x)) - d \big) \, dx \end{aligned}$$

by the push forward property and the Monge-Ampère equation that holds in the classical sense, see [7] and [18, Ex. 11.2] for instance. Hence, we deduce

$$\begin{aligned} &\frac{1}{2d} \frac{d}{dt} W_2^2(u_t, v_t) \\ &\leq - \int_{B_R} \Big[\det \nabla^2 \varphi_t(x) \, f\Big(\frac{u_t(x)}{\det \nabla^2 \varphi_t(x)} \Big) \, \big(\frac{\Delta \varphi_t^*(\nabla \varphi_t(x))}{d} - 1 \big) + f(u_t(x)) \, \big(\frac{\Delta \varphi_t(x)}{d} - 1 \big) \Big] \, dx. \end{aligned}$$

Notice that for fixed x and t the bracket in the integral is

$$p^d \, f(r p^{-d})(S-1) + f(r)(s-1)$$

where $r = u_t(x)$, $p = (\det \nabla^2 \varphi_t(x))^{1/d}$, $s = \frac{1}{d} \Delta \varphi_t(x)$ and $S = \frac{1}{d} \Delta \varphi_t^*(\nabla \varphi_t(x))$. Observe now that $s \geq p$ by the arithmetic-geometric inequality on the positive eigenvalues of the symmetric matrix $\nabla^2 \varphi_t(x)$. Moreover $\nabla \varphi_t^*(\nabla \varphi_t(x)) = x$ so $\nabla^2 \varphi_t^*(\nabla \varphi_t(x)) \, \nabla^2 \varphi_t(x) = \mathrm{Id}$ by differentiation in x, that is, $\nabla^2 \varphi_t^*(\nabla \varphi_t(x)) = (\nabla^2 \varphi_t(x))^{-1}$; hence also $S \geq p^{-1}$ for this matrix. Finally $f \geq 0$, so the expression above is

$$\geq p^d f(r p^{-d})(p^{-1} - 1) + f(r)(p-1) = (p-1)(f(r) - p^{d-1} f(r p^{-d})) \geq 0$$

by condition (3). This concludes the proof of the sufficient part.

We now turn to the proof of the *necessary* part. The strategy is to let u_0 and v_0 be uniform distributions on concentric balls with similar radii and estimate the dissipation rate of the $W_2^2(u_0, v_0)$ distance as the difference between the radii tends to zero. More precisely, let us define u_0 to be a suitably normalized characteristic function of a ball with radius a, and v_0 be a small variation of u_0, namely the uniform distribution with the same mass on a centered ball with radius $(1+\delta)a$. Then we will apply the contraction property to these two specific initial data and we will let $\delta \to 0$, recovering the condition on f as the first order term in δ. Similar heuristics were used to show the converse for the characterization of the displacement convexity of the entropy, see [17, Remark 5.18, Exercise 5.22].

We now define all in detail: let $r > 0$ and $0 < a < R$ be fixed, and define u_0 by $u_0(x) = r$ on the centered ball of radius a and $u_0(x) = 0$ outside. On the other hand, let $0 < \delta < (R-a)/(2a)$ be fixed and let v_0 be the image measure $\nabla\varphi \# u_0$ of u_0 by the map $\nabla\varphi$ where

$$\varphi(x) = \int_0^{|x|} \psi(z)\, dz, \qquad |x| \le R\,,$$

and ψ is the continuous increasing map defined on $[0, R]$ by

$$\psi(z) = \begin{cases} (1+\delta)z & \text{if} \quad 0 \le z \le a \\ (z + (1+2\delta)a)/2 & \text{if} \quad a \le z \le (1+2\delta)a \\ z & \text{if} \quad (1+2\delta)a \le z \le R. \end{cases}$$

Our aim is to write the contraction type property for these initial data u_0 and v_0 and t close to 0, and let δ goes to 0. However, since u_0 is neither positive nor smooth we need to proceed by approximation. The rest of this proof is devoted to show the technical details to make this density argument possible.

Step 1: Regularizing.- To take advantage of the exact expression of the dissipation of the Wasserstein distance as given in (6) for smooth positive densities we mollify u_0 and φ (whence v_0) in the following way.

For given $0 < \varepsilon < \varepsilon_0$, where ε_0 will depend only on a and δ, we let $\tilde{u}_0^\varepsilon$ be a C^2 radial function, nonincreasing on each ray, such that $\tilde{u}_0^\varepsilon(x) = r$ if $|x| \le a$ and $= \varepsilon$ if $|x| \ge a + \varepsilon$; then we let $u_0^\varepsilon = M\tilde{u}_0^\varepsilon / \int \tilde{u}_0^\varepsilon$, where $\int \tilde{u}_0^\varepsilon$ is greater than M but tends to M as ε goes to 0. Here $M = c_d a^d r$ is the mass of u_0, with c_d being the volume of the unit ball in $\mathbb{R}^d$.

We also mollify ψ into a C^3 increasing map ψ^ε on $[0, R]$ which is equal to ψ on $[0, a]$, $[(1+\varepsilon)a, (1+2\delta-\varepsilon)a]$, and $[(1+2\delta)a, R]$, satisfies

$$\|\psi^\varepsilon - \psi\|_{L^\infty[0,R]} \le A\varepsilon$$

for some $A = A(a, \delta)$ and $(\psi^\varepsilon)'(z) \in [\Lambda^{-1}, \Lambda]$ for some $\Lambda \ge 1 + \delta > 0$, uniformly in $z \in [0, R]$ and $\varepsilon \in (0, \varepsilon_0)$. Then the map

$$\varphi^\varepsilon(x) = \int_0^{|x|} \psi^\varepsilon(z)\, dz, \qquad |x| \le R$$

is radial, C^4 and strictly convex on $\bar{B}_R$ by composition (in particular at 0 since there it is given by $\varphi^\varepsilon(x) = (1+\delta)|x|^2/2$). Moreover

$$\nabla\varphi^\varepsilon(x) = \psi^\varepsilon(|x|)\frac{x}{|x|}$$

and

$$\|\nabla\varphi^\varepsilon - \nabla\varphi\|_{L^\infty(B_R)} = \|\psi^\varepsilon - \psi\|_{L^\infty[0,R]} \le A\varepsilon.$$

It is also classical that for all $x \ne 0$ the eigenvalues of the matrix $\nabla^2\varphi^\varepsilon(x)$ are $(\psi^\varepsilon)'(|x|)$ (with multiplicity 1 and eigenvector x) and $\psi^\varepsilon(|x|)/|x|$ (with multiplicity $d-1$ and eigenvectors the vectors orthogonal to x), as can be seen by writing the matrix $\nabla^2\varphi^\varepsilon(x)$. But $(\psi^\varepsilon)'(|x|) \in [\Lambda^{-1}, \Lambda]$ and $\psi^\varepsilon(|x|)/|x| \in [1, 1+\delta]$ by construction, so all eigenvalues of all matrices $\nabla^2\varphi^\varepsilon(x)$ are in $[\Lambda^{-1}, \Lambda]$, where the constant Λ is independent of ε.

We finally let $v_0^\varepsilon = \nabla\varphi^\varepsilon \# u_0^\varepsilon$.

Step 2: Use of the Dissipation of W_2.- The initial datum u_0^ε is C^2 and positive on the ball $\bar{B}_R$ and the matrices $\nabla^2\varphi^\varepsilon(x)$ have positive eigenvalues. In particular, the map $\nabla\varphi^\varepsilon$ is a C^1 diffeomorphism, and v_0^ε is also C^2 and positive on $\bar{B}_R$ by the change of variables

$$v_0^\varepsilon(y) = \Big(\frac{u_0^\varepsilon}{\det \nabla^2\varphi^\varepsilon}\Big)\big((\nabla\varphi^\varepsilon)^{-1}(y)\big). \tag{8}$$

Hence, (6)-(7) hold, in particular at $t = 0$, and the contraction property (4) for the solutions with respective initial data u_0^ε and v_0^ε, and t close to 0 ensures that

$$0 \geq \int_{B_R} \nabla(f \circ v_0^\varepsilon)(y) \cdot (\nabla\varphi^{\varepsilon,*}(y) - y)dy + \int_{B_R} \nabla(f \circ u_0^\varepsilon)(x) \cdot (\nabla\varphi^\varepsilon(x) - x)dx := I_1^\varepsilon + I_2^\varepsilon \tag{9}$$

for fixed a and δ, and $\varepsilon > 0$ small enough. Here we have let $\varphi^{\varepsilon,*} := (\varphi^\varepsilon)^*$.

Step 3: Passing to the limit $\varepsilon \to 0$.- We need to find the limit of I_1^ε and I_2^ε as $\varepsilon \to 0$. We will use weak-strong duality to pass to the limit. Both terms are treated in the same manner, then we present our arguments for I_1^ε. We divide it into two substeps:

Step 3.1.- First of all, we claim that the first term in the scalar product in I_1^ε, i.e. $\nabla(f\circ v_0^\varepsilon)$, weakly converges to the measure

$$-f(r(1+\delta)^{-d})\,\frac{y}{|y|}\,\mu_{(1+\delta)a}(dy)$$

with μ_α denoting the uniform measure (with density 1) on the $d-1$-dimensional sphere of radius $\alpha > 0$. The idea here is that if F^ε is a C^1 map on $\mathbb{R}^d$ which is equal to A on the ball $\{|x| \leq C\}$, equal to ε on $\{|x| \geq C(1+\varepsilon)\}$, and say radially decreasing in between, then ∇F weakly converges to the measure $-A\,|y|^{-1}\,y\,\mu_C(dy)$ as ε goes to 0.

Let indeed ζ be a vector test function on $\bar{B}_R$. Observing that by construction v_0^ε is constant outside the annulus $(1+\delta)a \leq |y| \leq (1+2\delta)a$, we obtain

$$\int_{B_R} \nabla(f\circ v_0^\varepsilon)\cdot\zeta = \int_{(1+\delta)a\leq|y|\leq(1+2\delta)a} \nabla(f\circ v_0^\varepsilon)\cdot\zeta = \int_{|y|=(1+2\delta)a} f\circ v_0^\varepsilon(y)\,\zeta(y)\cdot\frac{y}{|y|}$$
$$-\int_{|y|=(1+\delta)a} f\circ v_0^\varepsilon(y)\,\zeta(y)\cdot\frac{y}{|y|} - \int_{(1+\delta)a\leq|y|\leq(1+\delta)a+\Lambda\varepsilon} f\circ v_0^\varepsilon\,\nabla\cdot\zeta - \int_{(1+\delta)a+\Lambda\varepsilon\leq|y|\leq(1+2\delta)a} f\circ v_0^\varepsilon\,\nabla\cdot\zeta$$

by integration by parts.

The first integral is bounded by

$$f(\varepsilon)\int_{|y|=(1+2\delta)a} |\zeta(y)|$$

and $f(\varepsilon)$ tends to 0 as $\varepsilon \to 0$. Hence it tends to 0. The second integral is equal to

$$f\left(\frac{Mr}{(1+\delta)^d\int\tilde{u}_0^\varepsilon}\right)\int_{|y|=(1+\delta)a}\zeta(y)\cdot\frac{y}{|y|} \quad \text{converging to} \quad f\left(\frac{r}{(1+\delta)^d}\right)\int_{|y|=(1+\delta)a}\zeta(y)\cdot\frac{y}{|y|},$$

as $\varepsilon \to 0$.

To deal with the third integral, recall that $u_0^\varepsilon \leq r$ and that all eigenvalues of all matrices $\nabla^2\varphi^\varepsilon(x)$ are in $[\Lambda^{-1},\Lambda]$. Then again by (8) the regularized target densities verify $v_0^\varepsilon \leq r\Lambda^d$. Therefore, the third integral tends to 0 since v_0^ε and $\nabla\cdot\zeta$ are bounded uniformly in ε, and the volume of the domain of integration tends to 0 as $\varepsilon \to 0$.

The fourth integral also tends to 0 since the domain of integration is bounded and the integrand is bounded by a multiple of $f(\varepsilon\Lambda^d)$: Let indeed

$$|y| \geq (1+\delta)a + \Lambda\varepsilon = |\nabla\varphi^\varepsilon(a)| + \Lambda\varepsilon\,.$$

Since all eigenvalues of all matrices $\nabla^2\varphi^\varepsilon(x)$ are in $[\Lambda^{-1},\Lambda]$, then $\nabla\varphi^\varepsilon$ is Λ-Lipschitz. Thus, $|y| \geq |\nabla\varphi^\varepsilon(a+\varepsilon)|$ or equivalently $|(\nabla\varphi^\varepsilon)^{-1}(y)| \geq a+\varepsilon$ since $\nabla\varphi^\varepsilon$ is moreover radial. Therefore, we conclude from (8) and the definition of ψ^ε that

$$v_0^\varepsilon(y) = \frac{u_0^\varepsilon\big((\nabla\varphi^\varepsilon)^{-1}(y)\big)}{\det\nabla^2\varphi^\varepsilon\left((\nabla\varphi^\varepsilon)^{-1}(y)\right)} \leq \varepsilon\,\Lambda^d\,,$$

and the claim is proved.

Step 3.2.- We will show now that the second term in the scalar product in I_1^ε converges strongly. Using that $\nabla\varphi^\varepsilon$ is a diffeomorphism and that $\nabla\varphi^{\varepsilon,*}$ is also Λ-Lipschitz, we obtain

$$\begin{aligned}\|\nabla\varphi^{\varepsilon,*}(y) - \nabla\varphi^*(y)\|_{L^\infty(B_R)} &= \|x - \nabla\varphi^*\big(\nabla\varphi^\varepsilon(x)\big)\|_{L^\infty(B_R)}\\ &= \|\nabla\varphi^*\big(\nabla\varphi(x)\big) - \nabla\varphi^*\big(\nabla\varphi^\varepsilon(x)\big)\|_{L^\infty(B_R)} \leq \Lambda\,\|\nabla\varphi - \nabla\varphi^\varepsilon\|_{L^\infty(B_R)} \leq \Lambda\,A\,\varepsilon.\end{aligned}$$

Hence $\nabla\varphi^{\varepsilon,*}$ strongly converges to $\nabla\varphi^*$ in $L^\infty(B_R)$.

Step 4: Conclusion.- By Steps 3.1 and 3.2 and the weak-strong duality, it follows that I_1^ϵ in (9) converges to

$$-f\left(\frac{r}{(1+\delta)^d}\right)\int_{|y|=(1+\delta)a}\big(\nabla\varphi^*(y) - y\big)\cdot\frac{y}{|y|} = f\left(\frac{r}{(1+\delta)^d}\right)\,\delta\,a^d(1+\delta)^{d-1}c_d$$

since $\nabla\varphi^*(y) = (1+\delta)^{-1}y$ if $|y| = (1+\delta)a$. Analogously, I_2^ε in (9) converges to $-f(r)\delta\,a^d\,c_d$. Therefore, by collecting terms we conclude from passing to the limit in (9): $0 \geq I_1^\varepsilon + I_2^\varepsilon$, as $\varepsilon \to 0$ that

$$0 \geq \delta\,a^d\,c_d\left[(1+\delta)^{d-1}\,f\left(\frac{r}{(1+\delta)^d}\right) - f(r)\right]$$

for any fixed r and $\delta > 0$. Letting δ go to 0 leads to (2) at point r by simple Taylor expansion, hence to the equivalent condition (3). This concludes the proof of Theorem 1. □

Remark 3. R. J. McCann's proof of the displacement convexity of the entropy $\mathcal{U}$ under condition (3) in [12] is also based on the arithmetic-geometric inequality, though in a much less trivial manner than here.

Remark 4. Such contraction properties, or alternatively rates of convergence to equilibrium for the mean field equation

$$\frac{\partial u_t}{\partial t} = \Delta f(u_t) + \nabla\cdot(u_t\,\nabla W * u_t), \qquad t > 0,\ x \in \mathbb{R}^d \tag{10}$$

have been derived in [9, 10, 6]. For W strictly convex, and uniformly at infinity only, as in the physically motivated case of $W(z) = |z|^3$ on $\mathbb{R}$, only polynomial rates, or exponential but depending on the initial datum, have been obtained by entropy dissipation techniques. Then, bounding from above the (squared) Wasserstein distance between a solution and the steady state by its dissipation along the evolution, universal exponential rates have been derived in [4] for $f(u) = u$ (linear diffusion).

The same strategy can be pursued for a nonlinear diffusion. Assume for instance that $d = 1$, $f(r) = r^m$ with $1 < m < 2$ and W is a C^2 map on $\mathbb{R}$ for which for all $R > 0$ there exists $k(R) > 0$ such that $W''(x) \geq k(R)$ for all $|x| \geq R$. Then one can prove that for $x_0 \in \mathbb{R}$, equation (10) admits a unique steady state u_∞ in $\mathcal{P}_2(\mathbb{R})$ with center of mass x_0, and a constant $c > 0$ such that

$$W_2(u_t, u_\infty) \leq e^{-ct} W_2(u_0, u_\infty), \qquad t \geq 0$$

for all solutions $(u_t)_{t \geq 0}$ to (10) with initial datum u_0 in $\mathcal{P}_2(\mathbb{R})$ with center of mass x_0.

As in [4], the proof is based on considering the dissipation of the Wasserstein distance. Two key ingredients consist on one hand in bounding from below the contribution of the (now nonlinear) diffusion term in the dissipation and on the other hand on a good knowledge of the support of the steady state, and of the behaviour of this steady state near the boundary of its support.

Acknowledgments. The authors wish to thank Marco Di Francesco and Robert J. McCann for discussion on this question, and the referees for relevant remarks and suggestions. This note was partly written while they were visiting ACMAC in Heraklion and finished while the second author was a visiting professor at MSRI in Berkeley. It is a pleasure for them to thank these institutions for their kind hospitality. The first author acknowledges support from the French ANR-12-BS01-0019 STAB project and the second author acknowledges support from projects MTM2011-27739-C04-02 (Spain), 2009-SGR-345 from AGAUR-Generalitat de Catalunya, the Royal Society through a Wolfson Research Merit Award, and by the Engineering and Physical Sciences Research Council grant number EP/K008404/1.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, second edition, 2008.

[2] L. Ambrosio and G. Savaré. Gradient flows of probability measures. In *Handbook of differential equations: evolutionary equations. Vol. III*, pages 1–136. Elsevier/North-Holland, Amsterdam, 2007.

[3] F. Bolley, I. Gentil, and A. Guillin. Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations. *J. Funct. Anal.*, 263(8):2430–2457, 2012.

[4] F. Bolley, I. Gentil, and A. Guillin. Uniform convergence to equilibrium for granular media. *Arch. Ration. Mech. Anal.*, 208(2):429–445, 2013.

[5] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[6] M. Burger and M. Di Francesco. Large time behavior of nonlocal aggregation models with nonlinear diffusion. *Netw. Heterog. Media*, 3(4):749–785, 2008.

[7] L. A. Caffarelli. Boundary regularity of maps with convex potentials. II. *Ann. Math. (2)*, 144(3):453–496, 1996.

[8] J. A. Carrillo, A. Jüngel, P. A. Markowich, G. Toscani, and A. Unterreiter. Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities. *Monatsh. Math.*, 133(1):1–82, 2001.

[9] J. A. Carrillo, R. J. McCann, and C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoam.*, 19(3):971–1018, 2003.

[10] J. A. Carrillo, R. J. McCann, and C. Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Ration. Mech. Anal.*, 179(2):217–263, 2006.

[11] A. Lytchak. Differentiation in metric spaces. *Algebra i Analiz*, 16(6):128–161, 2004.

[12] R. J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.

[13] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Part. Diff. Eq.*, 26(1-2):101–174, 2001.
[14] K.-T. Sturm. Convex functionals of probability measures and nonlinear diffusions on manifolds. *J. Math. Pures Appl.*, 84:149–168, 2005.
[15] K.-T. Sturm and M.-K. von Renesse. Transport inequalities, gradient estimates, entropy and Ricci curvature. *Comm. Pure Appl. Math.*, 68:923–940, 2005.
[16] J. L. Vázquez. *The porous medium equation. Mathematical theory.* Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, Oxford, 2007.
[17] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2003.
[18] C. Villani. *Optimal transport. Old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer, Berlin, 2009.

Ceremade, Umr Cnrs 7534, Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75775 Paris cedex 16, France.
E-mail address: bolley@ceremade.dauphine.fr

Department of Mathematics, Imperial College London, SW7 2AZ - London, United Kingdom.
E-mail address: carrillo@imperial.ac.uk